

Departamento de Psicología Biológica y de la Salud
Programa de Doctorado de Psicología Clínica y de la Salud



EVALUACIÓN EXPERIMENTAL DE UNA INTERVENCIÓN
CONDUCTUAL MULTICOMPONENTE MEDIADA POR RED SOCIAL
EN EDUCACIÓN TERCIARIA

A MULTICOMPONENT BEHAVIORAL INTERVENTION IN TERTIARY
EDUCATION DELIVERED THROUGH A SOCIAL NETWORK: AN
EXPERIMENTAL EVALUATION

AIDA TARIFA RODRÍGUEZ

DIRECTORES

Dra. Ana Calero Elvira
Dr. Javier Virués Ortega

TESIS DOCTORAL
MENCIÓN INTERNACIONAL
Madrid 2022

The teacher who uses natural contingencies really abandons his role as a teacher.

He has only to expose the student to an environment;
the environment will do the teaching.

El educador que usa contingencias naturales de reforzamiento
ciertamente abandona su rol.

Solo tiene que exponer al alumno a un ambiente adecuado;
el ambiente será el que enseñará.

B. F. Skinner, *Technology of Teaching* (1968)

AGRADECIMIENTOS

A mi madre,
que, desde el cielo, me ha enviado fuerzas para seguir y ha hecho posible
que a día de hoy sea quien soy.

A mi hermano y mi padre, gracias por apoyarme
y recordarme que soy capaz de hacer todo lo que me proponga
a pesar de las adversidades.

A mis directores, Ana Calero y Javier Virués, por confiar en mí
y en mi trabajo diario, este logro es compartido.

A mi pareja, un faro de luz en la oscuridad,
que con paciencia ama cada una de mis versiones.

A Anita y Olga, dos tesoros de amigas que me acompañan desde
pequeña otorgándome muestras de cariño y ánimo.

Y a mí,
Por recorrer el camino
hasta el final.

ABREVIATURAS DE USO COMÚN

Abreviaturas	Español	Inglés
<i>BE-Social</i>	Paquete de intervención conductual mediante redes sociales	Behavioral Education and Social Media
<i>CABAS</i>	Aplicación integral del análisis de conducta a la enseñanza escolar	Comprehensive Application of Behavior Analysis to Schooling
<i>CAPSI</i>	Sistema Personalizado de Instrucción Asistido por Ordenador	Computer-aided Personalized System of Instruction
<i>DECU</i>	Diseño experimental de caso único	Single-case experimental designs (SCED)
<i>ECA</i>	Ensayo controlado aleatorizado	Randomized controlled trials (RCT)
<i>EHC</i>	Entrenamiento en Habilidades Conductuales	Behavior skills training
<i>HLM</i>	Modelo lineal jerárquico	Hierarchical linear models (HLM)
<i>PLATO</i>	Lógica Programada para la Operación Automática de la Enseñanza	Programmed Logic for Automatic Teaching Operations
<i>RAE</i>	Respuesta activa del estudiante	Active Student Response (ASR)
<i>RCT/ECA</i>	Ensayo controlado aleatorizado	Randomized controlled trial
<i>SPI</i>	Sistema personalizado individualizado	Personalized system of instruction
<i>TEA</i>	Trastorno del espectro autista	Autism spectrum disorder (ASD)

GLOSARIO SELECCIONADO

- **BE-Social:** Este acrónimo proviene de la intervención conductual propuesta en la tesis, denominada Behavioral Education and Social Media Program. El programa está compuesto por estrategias de autogestión con videomodelado, aprendizaje cooperativo y reforzamiento semi-inmediato.
- **Instructor o educador:** Ambos términos se emplean indistintamente a lo largo de la tesis para hacer referencia al profesional que aplica procedimientos analítico-conductuales en un aula, virtual o física, con estudiantes de educación terciaria.
- **Maestro:** En la presente tesis, se utiliza la terminología “maestro” cuando se trata de un profesional de la educación que trata con estudiantes de primaria y secundaria. Por ejemplo, al explicar las escuelas CABAS.
- **Retroalimentación/Feedback semi-inmediato:** En la presente tesis, este concepto se utiliza para hacer referencia a la entrega del reforzador en un periodo máximo de 24 horas.
- **Respuestas de interacción y participación:** A lo largo de la tesis este término se utiliza como traducción al término en inglés “engagement”. En concreto, implica cualquier respuesta que requiera interacción y cooperación en línea con otros usuarios que formen parte de la misma red social.

TABLAS

Número	Título	Pág.
Tabla 2.1	<i>Resumen sobre estudios conceptuales-metodológicos</i>	39
Tabla 2.2	<i>Resumen sobre estudios aplicados-experimentales</i>	39
Tabla 4.1	<i>Aplicación del sistema de economía de fichas integrada en una red social para mejorar la conducta de trabajo en grupo.</i>	125
Tabla 4.2	<i>Funcionalidades Comunes de Redes Sociales y su Valor Práctico en Modelos de Educación Conductual</i>	130
Table 5.1	<i>Discrimination Complexity of Various Multiple-Choice Item Types.</i>	143
Table 5.2	<i>Multiple-Choice Items by Type and Selected Underlying Propositions.</i>	144
Table 6.1	<i>Structured criteria for the visual analysis of a multielement design functional analysis (according to Hagopian et al., 1997, and Cox & Virues-Ortega, 2016).</i>	173
Table 6.2	<i>Structured visual analysis for Annette, Bob, Jed, and Max.</i>	174
Table 6.3	<i>Key characteristics of selected single-subject effect size metrics and analytic techniques.</i>	174
Table 6.4	<i>Standards of selected single-subject effect size metrics and analytic techniques.</i>	176
Table 6.5	<i>Design matrix for one participant.</i>	177
Table 7.1	<i>Linear Mixed Effects Model for Accuracy Proportion (n = 10).</i>	205
Table 8.1	<i>Characteristics of Included Studies (n = 75)</i>	231
Table 8.2	<i>Objective Measures of Social Media Engagement in the Tertiary Education Literature</i>	239
Table 8.3	<i>A Proposed Taxonomy of Quantitative Metrics of Social Media Engagement</i>	240
Table 10.1	<i>Sociodemographic Characteristics of Study Participants</i>	277
Table 10.2	<i>Two-way Analysis of Variance during baseline (Time 1), intervention (Time 2), and Intervention without Semi-Immediate Feedback (Time 3)</i>	277

Número	Título	Pág.
Table 10.3	<i>Visual Analysis Summary</i>	278
Table 10.4	<i>Multiple Baseline Design Effect Size Analysis of Participants in the Intervention Group (n = 28)</i>	278
Table 11.1	<i>Sociodemographic Characteristics of Participants (n = 141)</i>	309
Table 11.2	<i>Pairwise t Test Comparisons for Academic Achievement across Groups</i>	310
Table 11.3	<i>Linear Mixed Effects Model for Combined Engagement</i>	311

FIGURAS

Número	Título	Pág.
Figure 5.1	A decision tree for the combination a randomized controlled trial (RCT) and a single-case experimental design (SCED).	141
Figure 6.1	Some graphical examples of common visual analysis scenarios.	178
Figure 6.2	Examples of local trend (first B phase), and non-reversibility (second A phase, lower graph).	178
Figure 6.3	Cases of attention- (Bob and Annette) and escape-maintained (Max and Grey) self-injury.	179
Figure 6.4	R code for a two-level hierarchical linear model with lag1 autocorrelation and heterogeneity of variances	180
Figure 6.5	R output for a two-level hierarchical linear model with lag1 autocorrelation and heterogeneity of variances.	181
Figure 7.1	Observation time across phases and observation methods.	206
Figure 7.2	Accuracy across phases and observation methods.	206
Figure 7.3	Observation Accuracy of The Observer XT® vs. Paper and Pencil (top panel) and of Big Eye Observer® vs. Paper and Pencil (Lower Panel) across Participants over Sequential Sessions with the Observation Method	207
Figure 7.4	Accuracy by Number of Recordable Events across All Observation Sessions.	207
Figure 7.5	Distribution of Interobserver Agreement Values in the Literature.	208
Figure 8.1	Study Selection Flowchart	241
Figure 10.1	Temporal distribution of the baseline (light grey bars) and intervention (black bars) phases across the arms of the randomized controlled trial.	279
Figure 10.2	Single-Subject Analysis of Participants 1 through 18	280
Figure 10.3	Single-Subject Analysis of Participants 19 through 36	281
Figure 10.4	Single-Subject Analysis of Participants 37 through 46	282

Número	Título	Pág.
Figure 11.1	<i>Participant Attrition Flowchart</i>	312
Figure 11.2	<i>Changes in Academic Achievement during Pre-Intervention, Intervention, and Post-Intervention</i>	313
Figure 11.3	<i>Academic Performance and Social Media Engagement across Intervention Groups</i>	314

Tabla de Contenidos

RESUMEN	21
ABSTRACT	23
PRIMERA PARTE: PRESENTACIÓN Y OBJETIVOS	25
CAPÍTULO 1. INTRODUCCIÓN.....	27
CAPÍTULO 2. OBJETIVOS E HIPÓTESIS	31
1. ESTUDIOS CONCEPTUALES-METODOLÓGICOS.....	31
2. ESTUDIOS EXPERIMENTALES-APLICADOS	33
SEGUNDA PARTE: MARCO TEÓRICO.....	41
CAPÍTULO 3. ANÁLISIS APLICADO DE CONDUCTA EN LA EDUCACIÓN	43
1.EL ANÁLISIS APLICADO DE CONDUCTA.....	43
1.1 Historia previa al análisis de conducta	43
2.ANÁLISIS APLICADO DE CONDUCTA EN LA EDUCACIÓN	46
2.1 Las máquinas de enseñanza.....	47
2.2 Instrucción programada	49
2.3 Sistema personalizado de instrucción	51
2.3.1 Énfasis en la comunicación escrita	52
2.3.2 El efecto del ritmo	52
2.3.3 El dominio del contenido.....	53
2.3.4 El apoyo de los supervisores	53
2.3.5 Conferencias motivacionales.....	53
2.4 Instrucción directa.....	54
3.EDUCACIÓN CONDUCTUAL FUNDAMENTADA EN LA EVIDENCIA:	
APLICACIONES EDUCATIVAS FOCALIZADAS.....	56
3.1 Estrategias con base en la evidencia	57
3.2 Gestión del aula	58
3.2.1 Opciones de enseñanza	59
3.2.2 Tasa de oportunidades de respuesta.....	59
3.3 Aprendizaje cooperativo	60
3.4 Tutoría entre iguales.....	61
3.4.1 Respuesta activa del estudiante.....	62
3.4.2 Oportunidades de respuesta.....	62
3.4.3 Unidad de aprendizaje.....	63
3.5 Economía de fichas.....	63
3.6 Entrenamiento de habilidades conductuales.....	64
3.6.1 Instrucciones.....	65
3.6.2 Modelado	65
3.6.3 Ensayo conductual.....	66
3.6.4 Retroalimentación	66
3.6.5 Aplicaciones.....	66
3.7 Videomodelado.....	68
3.8 Autogestión.....	69
3.8.1 Automonitorización	70

3.8.2 Autoevaluación	71
3.8.3 Identificación de objetivos.....	72
3.8.4 Autorreforzamiento	73
3.9 Instrucción sobre equivalencia.....	74
3.10 Retroalimentación	75
4. EDUCACIÓN CONDUCTUAL FUNDADA EN LA EVIDENCIA:	
APLICACIONES EDUCATIVAS INTEGRALES	78
4.1 Enseñanza de Precisión	78
4.1.1 El estudiante siempre tiene la razón.....	78
4.1.2 Uso de medidas directas.....	78
4.1.3 Registrar la fluidez	79
4.1.4 Uso del gráfico de aceleración estándar	80
4.1.5 Tutorías.....	81
4.1.6 Aplicaciones prácticas	81
4.2 CABAS.....	82
4.2.1 Componentes de CABAS	83
4.2.2 Formación y supervisión	84
4.2.3 Modelos de medición continua	85
4.3 Instrucción Directa	87
4.3.1 Lenguaje.....	87
4.3.2 Lectura	88
4.3.3 Matemáticas.....	88
4.3.4 Escritura	89
4.3.5 Investigación.....	89
4.4 Sistema personalizado de instrucción	91
4.4.1 Componentes esenciales según la literatura reciente.....	91
4.4.2 Áreas de investigación.....	92
4.5 Interteaching.....	93
5. CONSIDERACIONES FINALES.....	97
CAPÍTULO 4. APLICACIONES ANALÍTICO-CONDUCTUALES EN EL	
ÁMBITO EDUCATIVO MEDIANTE TICS.....	99
1. APLICACIÓN DE SOFTWARES EDUCATIVOS EN EDUCACIÓN TERCIARIA	99
1.1 Enseñanza asistida por ordenador o software educativos.....	99
1.1.1 Software basado en ejercicios y práctica.....	100
1.1.2 Software basado en tutoriales	100
1.1.3 Software basado en simulacros.....	101
1.2 Componentes eficaces para elaborar una enseñanza asistida por ordenador	101
1.2.1 Estrategias de enseñanza en base a los antecedentes.....	101
1.2.2. Estrategia de enseñanza en base a la conducta deseada.....	102
1.2.3 Estrategias de enseñanza en base a las consecuencias.....	103
1.3 Sistemas asistidos por ordenador en análisis aplicado de conducta.....	103
2.ENSEÑANZA ASISTIDA POR ORDENADOR Y TELESALUD.....	105
2.1 Telesalud	105
2.2 Telesalud en el análisis aplicado de conducta	107
3.ENSEÑANZA ASISTIDA EN LA EDUCACIÓN SUPERIOR CON SOPORTE DE REDES SOCIALES.....	108
3.1 Las redes sociales	108
3.1.1 Definición.....	108
3.1.2 Características.....	109
3.1.3 Facebook como modelo de red social con aplicación educativa.....	110
3.2 Educación superior y redes sociales.....	111
4. COMPATIBILIDAD DE LAS INTERVENCIONES FOCALIZADAS E INTEGRALES EN LAS REDES SOCIALES...	113
4.1 Intervenciones educativas integrales.....	115

4.1.1 Enseñanza de precisión.....	116
4.1.2 Instrucción Directa.....	117
4.1.3 Sistema personalizado de instrucción.....	118
4.1.4 CABAS.....	118
4.1.5 Interteaching.....	119
4.2 Intervenciones educativas focalizadas.....	120
4.2.1 Gestión del aula.....	121
4.2.2 Aprendizaje cooperativo.....	122
4.2.3 Tutoría con iguales.....	122
4.2.4 Economía de fichas.....	123
4.2.5 Videmodelado.....	125
4.2.6 Entrenamiento en habilidades conductuales.....	126
4.2.7 Estrategias de autogestión.....	127
4.2.8 Instrucción sobre equivalencia.....	128
4.2.9 Retroalimentación.....	128
5. CONSIDERACIONES FINALES.....	129
TERCERA PARTE: BASES METODOLÓGICAS.....	133
CHAPTER 5. METHODOLOGICAL BASIS OF BEHAVIOR-ANALYTIC INTERVENTIONS IN EDUCATION.....	135
1. CHOOSING A MEANINGFUL BEHAVIORAL OUTCOME.....	135
2. INTEROBSERVER AGREEMENT.....	137
3. PROCEDURAL INTEGRITY.....	138
4. INTENSITY OF THE INSTRUCTIONAL ACTIVITY.....	139
5. COMBINATION OF GROUP-BASED AND SINGLE-SUBJECT APPROACHES.....	140
6. STATISTICAL APPROACHES TO ANALYZING COMBINED GROUP-BASED AND SINGLE-SUBJECT DESIGNS.....	142
7. METHODOLOGICAL COROLLARIES AND THE DESIGN OF STUDIES WITH A METHODOLOGICAL AND AN APPLIED FOCUS.....	142
CHAPTER 6. QUANTIFYING OUTCOMES IN APPLIED BEHAVIOR ANALYSIS THROUGH VISUAL AND STATISTICAL ANALYSES: A SYNTHESIS.....	145
ABSTRACT.....	145
INTRODUCTION.....	145
VISUAL ANALYSIS.....	148
SYNERGY BETWEEN STATISTICAL AND VISUAL ANALYSES.....	156
STATISTICAL TECHNIQUES.....	157
CONCLUSION.....	164
REFERENCES.....	165
CHAPTER 7. ACCURACY OF PAPER-AND-PENCIL SYSTEMATIC OBSERVATION VERSUS COMPUTER-AIDED SYSTEMS.....	183
ABSTRACT.....	183
INTRODUCTION.....	184
MATERIAL AND METHODS.....	186
RESULTS.....	193
DISCUSSION.....	194
CONCLUSIONS.....	197
ACKNOWLEDGEMENTS.....	198
OPEN PRACTICE STATEMENT.....	198
REFERENCES.....	198

CHAPTER 8. QUANTITATIVE INDICES OF STUDENT SOCIAL MEDIA ENGAGEMENT IN TERTIARY EDUCATION: A SYSTEMATIC REVIEW AND A TAXONOMY	209
ABSTRACT	209
INTRODUCTION	210
METHODS.....	211
RESULTS.....	213
DISCUSSION.....	215
CONCLUSION.....	220
REFERENCES.....	220
CHAPTER 9. GENERAL METHODOLOGICAL APPROACH: CONCLUSIONS.....	243
1. SYNERGY OF SINGLE-SUBJECT AND GROUP-BASED ANALYSES.....	243
2. VALIDITY AND PRECISION OF COMPUTER-BASED SYSTEMS FOR VIDEO-BASED BEHAVIORAL OBSERVATION	244
3. QUANTITATIVE ENGAGEMENT OUTCOMES AND A CONCEPTUAL MODEL FOR EVIDENCE-BASED INTERVENTIONS.....	244
4. PROCESS-BASED INTERVENTIONS MAY REQUIRE COMPONENT ANALYSIS AND ULTIMATELY MULTI-ARM RCTS.....	245
CUARTA PARTE: ESTUDIOS EXPERIMENTALES-APLICADOS	247
CHAPTER 10. EFFECT OF THE BEHAVIORAL EDUCATION AND SOCIAL MEDIA (BE-SOCIAL) PROGRAM ON POSTGRADUATE ACADEMIC PERFORMANCE: A RANDOMIZED CONTROLLED TRIAL	249
ABSTRACT	249
INTRODUCTION	249
MATERIALS AND METHODS	254
RESULTS.....	260
DISCUSSION.....	263
CONCLUSIONS	268
REFERENCES.....	269
CAPÍTULO 11. ENHANCING ACADEMIC PERFORMANCE IN TERTIARY EDUCATION THROUGH SOCIAL MEDIA: A MULTI-ARM RCT.....	283
ABSTRACT	283
INTRODUCTION	283
METHODS.....	288
RESULTS.....	295
DISCUSSION.....	297
CONCLUSIONS	301
REFERENCES.....	302
QUINTA PARTE: DISCUSIÓN.....	315
CAPÍTULO 12. DISCUSIÓN, LIMITACIONES Y LÍNEAS FUTURAS	317
DISCUSIÓN.....	317
CONCLUSIONES GENERALES	333
GENERAL CONCLUSIONS.....	335
REFERENCES.....	337

ANEXOS

Anexo A Modelo de consentimiento informado

Anexo B Accuracy of paper-and-pencil Systematic Observation versus Computer-Aided systems.

Anexo C Quantitative Indices of Student Social Media Engagement in Tertiary Education: A Systematic Review and a Taxonomy

Anexo D Effect of the Behavioral Education and Social Media (BE-Social) Program on Postgraduate Academic Performance: A Randomized Controlled Trial

Anexo E Enhancing Academic Performance in Tertiary Education through Social Media: A Multiple-Arm RCT

Anexo F Publicaciones, participaciones en proyectos, ponencias internacionales.

RESUMEN

El uso generalizado de redes sociales tales como Facebook, YouTube, Twitter, WhatsApp, o Instagram entre estudiantes de educación terciaria, así como el uso creciente de dichos sistemas con fines educativos reclama un análisis del potencial de dichos ambientes digitales para la realización de procedimientos educativos basados en la evidencia. Pese al número de investigaciones en este ámbito, la literatura carece de análisis conductuales que hayan operativizado las conductas de interacción y participación y analizado experimentalmente el impacto del uso educativo de redes sociales en el rendimiento académico. Desde una perspectiva educativa, el uso de intervenciones analítico-conductuales por parte del profesorado puede proporcionar estrategias eficaces que incrementen la interacción positiva en redes sociales con fines educativos dando lugar a un incremento en rendimiento académico.

La presente tesis doctoral “Evaluación experimental de una intervención conductual multicomponente mediada por red social en educación terciaria” presenta una propuesta sistemática de intervención analítico-conductual dirigida a incrementar las respuestas de interacción y participación en redes sociales educativas, así como el rendimiento académico de estudiantes de bajo desempeño. A tal fin, la presente tesis doctoral aborda los siguientes objetivos principales. Primero, diseñar y evaluar un programa de intervención conductual multicomponente denominado BE-Social y determinar su impacto en la interacción y participación en estudiantes adultos, así como examinar los posibles cambios en el rendimiento académico. Segundo, y a fin de optimizar nuestra propuesta de intervención, realizaremos un ensayo controlado aleatorizado (ECA) multi-grupo para determinar la contribución relativa de los varios elementos del programa multicomponente BE-Social.

A fin de llevar a efecto los objetivos principales, y dada la escasa investigación conductual en este ámbito, se hacen esenciales un conjunto de estudios metodológicos: (a) revisión de los procedimientos de análisis visuales y estadísticos en diseños conductuales de series temporales (i.e., diseños experimentales de caso único), (b) validación de sistemas informatizados de observación conductual de video mediante aplicación móvil a fin de asegurar la integridad de los

procedimientos de intervención mediante red social, y (c) revisión sistemática de las métricas cuantitativas sobre respuestas de interacción y participación en red social de estudiantes de educación terciaria. Los resultados obtenidos en la presente colección de estudios son los siguientes: se valida la aplicación tecnológica propuesta para observar videos que emplean: (a) se identifican estrategias prácticas para la combinación de diseños ECA y de caso único, así como los procedimientos estadísticos adecuados a dichos escenarios, (b) se valida el app de observación de video Big Eye Observer, afianzándose su idoneidad para su uso en los análisis de integridad de procedimiento de los dos estudios ECA, y (c) la revisión sistemática arroja un conjunto de métricas cuantitativas que suplementamos con una taxonomía original acorde a la dimensión conductual del análisis aplicado de conducta (Baer et al., 1968).

Por otro lado, el primer estudio experimental demostró que la mayoría de participantes que recibió el programa BE-Social incrementaban sus respuestas de interacción y participación, además del rendimiento académico. Otro hallazgo importante a considerar es la compatibilidad de una red social como canal eficaz para administrar intervenciones conductuales a estudiantes de posgrado que asistan a una formación online. Finalmente, el último estudio concluye que el programa BE-Social con retroalimentación semi-inmediata, impacta positivamente al rendimiento académico.

Este conjunto de estudios arroja las conclusiones siguientes sobre el uso de intervención educativa conductual mediada por red social: (a) se validan numerosas estrategias metodológicas en este ámbito con gran potencial para su aplicación en futuras investigaciones (p.ej., combinación de diseños experimentales de grupo y caso único, utilización de métricas cuantitativas de interactividad), (b) se valida la viabilidad, efectividad y aceptabilidad de un paquete multicomponente de intervención, y (c) se establece preliminarmente la naturaleza aditiva de los procesos operantes, principalmente el uso de feedback semi-inmediato del instructor, de los elementos del programa multicomponente BE-Social. Los resultados obtenidos trazan nuevos horizontes hacia la formación online efectiva de estudiantes de posgrado empleando estrategias conductuales compatibles con las redes sociales.

ABSTRACT

The widespread use of social networks such as Facebook, YouTube, Twitter, WhatsApp, or Instagram among tertiary education students, as well as the increasing use of such systems for educational purposes calls for an examination and analysis of the potential of such digital environments for evidence-based educational procedures. Despite the number of investigations in this area, the literature lacks behavioral analyses that have operationalized interaction and participation behaviors and experimentally analyzed the impact of educational use of social networks on academic performance. From an educational perspective, the use of behavioral-analytic interventions by teachers can provide effective strategies to increase positive interaction in social networks for educational purposes, leading to an increase in academic performance.

The present doctoral dissertation "A multicomponent behavioral intervention in tertiary education delivered through a social network: an experimental evaluation" presents a systematic proposal of an analytical-behavioral intervention aimed at increasing the interaction and participation responses in educational social networks, as well as the academic performance of low-performing students. To this end, this doctoral dissertation addresses the following main objectives. First, to design and evaluate a multicomponent behavioral intervention program called BE-Social and to determine its impact on interaction and participation in adult students, examining possible changes in academic performance. Second, in order to optimize our intervention proposal, we will conduct a multi-group RCT to determine the relative contribution of the various elements of the BE-Social multicomponent program.

In order to carry out the main objectives, and given the scarcity of behavioral research in this area, a set of methodological studies are essential: (a) we identify practical strategies for the combination of RCT and single-case designs, as well as statistical procedures suitable for such scenarios, (b) we validate the Big Eye Observer video observation app, strengthening its suitability for use in the procedural integrity analyses of the two RCT studies, and (c) the systematic review yields a set of quantitative metrics that we supplement with an original taxonomy in line with the behavioral dimension of applied behavior analysis (Baer et al. , 1968).

On the other hand, the first experimental study showed that the majority of participants who received the BE-Social program increased their interaction and participation responses, in addition to academic performance. Another important finding to consider is the compatibility of a social network as an effective channel for administering behavioral interventions to graduate students attending online training. Finally, the last study concludes that the BE-Social program with semi-immediate feedback positively impacts academic performance.

This set of studies yields the following conclusions on the use of social network-mediated behavioral educational interventions: (a) a number of methodological strategies are validated in this area with great potential for application in future research, (b) the studies support the feasibility, effectiveness and acceptability of a multicomponent intervention package, and (c) the empirical studies are consistent with the additive nature of the elements of the BE-Social multicomponent program (mainly the use of semi-immediate instructor feedback). The results obtained set new methodological and practical references towards the effective use of online training in tertiary education with the aid of social media platforms.

PRIMERA PARTE
PRESENTACIÓN Y OBJETIVOS

CAPÍTULO 1. Introducción

Las investigaciones en el área del uso de redes sociales con fines educativos en educación terciaria que aplican estrategias basadas en la evidencia, es una cuestión que a día de hoy permanece sin explorar. La presente tesis se desarrolla desde el enfoque teórico y sistemático del aprendizaje. La investigación que predomina en el análisis aplicado de conducta se centra en los primeros años de aprendizaje, provocando un vacío en las etapas de adolescencia (instituto) y adultez (universidad y posgrado) (Sulzer-Azaroff & Gillat, 1990). Por lo general, existe una brecha entre la investigación (procedimientos de enseñanza empíricos) y la aplicación en contextos educativos (Kratochwill et al., 2004). El capítulo 3, “Análisis aplicado de la conducta en la educación”, expone las primeras investigaciones que contribuyeron a abordar las conductas socialmente significativas en el ámbito de la educación y desde un enfoque analítico-conductual. Además, proporciona información detallada sobre los mecanismos conductuales propuestos por B. F. Skinner como las contingencias de reforzamiento, el control de estímulo, los métodos de enseñanza basados en la instrucción efectiva y los procedimientos que operan en entornos de aprendizaje (por ejemplo, el videomodelado, el aprendizaje cooperativo, la retroalimentación, etc.).

Paralelamente, la educación sufre una continua transformación desde el desarrollo de las nuevas tecnologías de la información y la comunicación (TIC) impactando en la enseñanza desde la etapa preescolar hasta la educación terciaria. La incorporación de softwares educativos, de dispositivos electrónicos y la revolución de las redes sociales han transformado las contingencias de interacción entre los estudiantes y los docentes. Además de modificar las contingencias de interacción y participación en las aulas, físicas o virtuales, el formato y la cantidad de materiales de enseñanza, incluso las estrategias de instrucción, los docentes sufren, a día de hoy un periodo de adaptación en la nueva era de la educación basada en las TICs. A raíz de esta revolución tecnológica, la educación es más interactiva e individualizada, enfatizando las habilidades de autogestión y autoaprendizaje del propio estudiante. La omnipresencia de las redes sociales en la vida cotidiana de los estudiantes ha de contemplarse como un canal versátil de enseñanza. Además, las redes sociales contienen herramientas que fomentan la interacción y la colaboración entre sus miembros. A continuación, el capítulo 4, “Las redes sociales en la educación”, explora el papel

de las nuevas tecnologías en la educación terciaria y analiza qué recursos podrían adaptarse a la aplicación de intervenciones integrales y focalizadas analítico conductuales.

Por otro lado, la aplicabilidad de la metodología experimental del análisis aplicado de conducta en el contexto de educación superior es limitada. Es por este motivo que la presente investigación ha destinado una sección a los capítulos y estudios categorizados como “Bases conceptuales-metodológicas”. El capítulo 5 introduce la sección y justifica la necesidad de conocer los análisis básicos metodológicos y sus limitaciones en la aplicación de intervenciones educativas conductuales. El capítulo 6 revisa la compatibilidad de los análisis visuales y estadísticos utilizados en los diseños experimentales de caso único, comúnmente empleados en el análisis aplicado de conducta. El capítulo 7 valida un procedimiento de observación conductual sobre video utilizando un app de observación, estrategia que luego se integrará en la evaluación de la integridad de procedimiento de los estudios aplicados. El capítulo 8 presenta una revisión sistemática de la literatura actual sobre las repuestas de interacción y participación generadas en las redes sociales por estudiantes de educación terciaria. El propósito es conocer el estado de la literatura reciente y proponer una taxonomía con métricas cuantitativas a partir de la información recopilada. Se presenta una herramienta objetiva de análisis que incluye las dimensiones de la conducta (p.ej., latencia, tiempo entre respuesta, duración, etc.). Por último, el bloque “conceptual-metodológico” finaliza con el capítulo 9 que señala y recopila las conclusiones de los estudios presentados y esculpe las bases para las investigaciones empíricas de la siguiente sección.

El cuarto bloque de la presente tesis lo conforman dos estudios experimentales-aplicados; ambos relacionados entre sí. El primero presenta la creación de un paquete de estrategias de enseñanza conductuales (paquete que hemos denominado *BE-Social*) a un grupo de estudiantes de posgrado, con el fin de evaluar si la aplicación sistemática de la intervención provoca un efecto directo en la interacción y participación de los estudiantes y, además, observa si se producen cambios favorables en el rendimiento académico. Las limitaciones obtenidas en este primer estudio experimental tratan de minimizarse con un segundo estudio, es decir, con una replicación del primer estudio redefiniendo factores experimentales. El incremento de la muestra, conocer las características de los participantes, modificar los parámetros de las variables independientes y el uso de ensayos controlados aleatorizados de grupos múltiples permite mostrar qué conjunto de componentes de la intervención *BE-Social* son esenciales para provocar cambios en

el rendimiento académico, así como determinar si existe correlación entre el índice de respuestas de interacción y participación con los resultados del rendimiento académico.

La tesis concluye con la quinta y última sección, la discusión final y las conclusiones. Se exponen los resultados generales y las consideraciones conceptuales y metodológicas que aporta cada estudio. Además, se discuten las principales limitaciones y se examinan las líneas futuras y las recomendaciones prácticas a partir de los resultados obtenidos a lo largo de la tesis.

CAPÍTULO 2. Objetivos e Hipótesis

A lo largo de este capítulo se plantean las razones principales que justifican la necesidad de realizar cinco estudios, tres de ellos son estudios conceptuales-metodológicos y los dos restantes son estudios experimentales-aplicados en contextos de formación terciaria. La presente investigación pretende evaluar la hipótesis de que las respuestas de interacción y participación que generen las redes sociales de estudiantes de especialización profesional vinculados a cursos reglados pueden ser un elemento manipulable para mejorar la eficiencia del docente en la enseñanza online e impacte cuantificablemente en el rendimiento de los estudiantes en pruebas objetivas. Para ello, es primordial revisar sistemáticamente cómo se han definido y medido cuantitativamente las variables de interacción y participación en estudiantes de posgrado y así conocer qué análisis descriptivos o experimentales se han empleado en los últimos veinte años. Finalmente, esta tesis pretende ocupar un vacío en la literatura de la educación conductual contribuyendo con dos estudios experimentales con el fin de diseñar, evaluar y manipular sistemáticamente un paquete de intervención de estrategias analítico-conductuales para demostrar experimentalmente si existe relación entre la interactividad mediada a través de redes sociales y la mejora del rendimiento académico.

Para alcanzar este objetivo general se han llevado a cabo los siguientes estudios, con sus respectivos objetivos e hipótesis.

1. ESTUDIOS CONCEPTUALES-METODOLÓGICOS

Estudio 1: Quantifying Outcomes in Applied Behavior Analysis through Visual and Statistical Analyses: A Synthesis.

Referencia: Virues-Ortega, J., Moeyaert, M., Sivaraman, M., Tarifa-Rodriguez, & Fernandez Castilla, B. (2022). Quantifying Outcomes in Applied Behavior Analysis through Visual and Statistical Analyses: A Synthesis. In J. L. Matson, *Applied Behavior Analysis: A Comprehensive Handbook*. SpringerNature.

- **Objetivo:** Revisar las estrategias de análisis de datos en análisis aplicado de conducta y, en particular, la combinación de análisis visuales y estadísticos en diseños experimentales de caso único (DECU)¹. Con ello se pretende realizar decisiones informadas en la selección de las estrategias metodológicas a utilizar en los estudios de intervención descritos a continuación.

Estudio 2: Accuracy of paper-and-pencil systematic observation versus computer-aided systems.

Referencia: Virues-Ortega, J., Casas, C. D., Martín, N., Tarifa-Rodríguez, A., Hidalgo, A., Cox, A. D., & Navarro Guzmán, J. I. (2022). Accuracy of paper-and-pencil systematic observation versus computer-aided systems. *Behavior Research Methods*. <https://doi.org/10.3758/s13428-022-01861-0>

- **Objetivo 1:** Validar un procedimiento de observación conductual sobre video asistido por dispositivo móvil mediante la evaluación de varios sistemas (lápiz y papel, Observer XT y Big Eye Observer). Ello tiene el fin de utilizar los sistemas que hayan sido debidamente validados para el registro de conductas de interés sobre video como parte del protocolo de integridad de procedimiento de los estudios empíricos de intervención descritos a continuación.
- **Hipótesis 1a:** Se espera que la aplicación Big Eye Observer 2.0 tenga un nivel de precisión comparable al de Observer XT y registro de lápiz y papel en protocolos de observación conductual de intervalo y de registro de eventos realizados retrospectivamente sobre vídeos de intervenciones conductuales con múltiples códigos de observación.

Estudio 3: Quantitative Indices of Student Social Media Engagement in Tertiary Education: A Systematic Review and a Taxonomy

- **Objetivo 1:** Evaluar la metodología que se aplica en la ciencia de la educación para medir las respuestas de interacción y participación que generan estudiantes universitarios o de posgrado al utilizar las redes sociales en entornos educativos.

1 Single-case experimental designs (SCED), en la literatura original

- *Hipótesis 1a: Se espera que la mayoría de estudios utilicen evaluaciones cualitativas (cuestionarios ad hoc de tipo Likert o encuestas) y muy pocas investigaciones empleen diseños experimentales ni ensayos controlados aleatorizados (ECA), es decir, la presencia de una condición control que proporcione credibilidad al comparar las intervenciones aplicadas en los grupos experimentales.*
- **Objetivo 2:** *Elaborar una taxonomía utilizando métricas cuantitativas sobre las respuestas de interacción y participación que generan las plataformas de redes sociales que pueda utilizarse para el estudio de procesos operantes de enseñanza-aprendizaje en educación terciaria.*
 - *Hipótesis 2a: Se espera que la literatura existente ofrezca varias métricas cuantitativas de participación en red social y que estas puedan ser ampliadas con dimensiones conductuales de uso habitual en estudios aplicados y de operante humana (p.ej., frecuencia, latencia, tiempo entre respuestas, duración).*

2. ESTUDIOS EXPERIMENTALES-APLICADOS

Estudio 4: *Effect of the Behavioral Education and Social Media (BE-Social) Program on Postgraduate Academic Performance: A Randomized Controlled Trial*

- **Objetivo 1:** *Diseñar el impacto global de un paquete de procedimientos analítico-conductual (BE-Social) adaptable a un grupo cerrado de Facebook en cursos de formación online de estudiantes de posgrado.*
 - *Hipótesis 1a: Esperamos que la intervención multicomponente (BE-Social) se adapte a las funcionalidades de un grupo cerrado de Facebook y permita incrementar objetivamente las respuestas de interacción y participación de los estudiantes.*
- **Objetivo 2:** *Evaluar experimentalmente el paquete de intervención conductual diseñado (BE-Social) con el objetivo de incrementar objetivamente respuestas de interacción y participación con el profesorado y compañeros.*

- *Hipótesis 2a: Se espera que el paquete de intervención BE-Social promueva la interacción y la participación de los estudiantes y se observe un incremento cuantitativo.*
- **Objetivo 3:** Examinar la relación entre las respuestas objetivas de interacción y participación con el rendimiento académico obtenido a través de pruebas semanales de opción múltiple, en estudiantes de posgrado que utilizan las redes sociales en entornos educativos.
 - *Hipótesis 3a: Esperamos que los estudiantes que obtengan mayor tasa de respuestas de interacción y participación obtengan mayores calificaciones durante el curso académico.*

Estudio 5: *Enhancing Academic Performance in Tertiary Education through Social Media: A Multiple-Arm RCT*

- **Objetivo 1:** Identificar y verificar experimentalmente qué procedimiento o combinaciones de procedimientos del paquete de intervención BE-Social provocan mayor interacción y participación en estudiantes de posgrado en un grupo de formación online que tenga integrado un grupo cerrado de Facebook como canal de enseñanza-aprendizaje.
 - *Hipótesis 1a: En vista de la importancia de la inmediatez desde los inicios del estudio del reforzamiento relacionado con el aprendizaje (Liu, 2021), se espera que el incremento de respuestas futuras de interacción y participación se deban al procedimiento o la combinación de procedimientos que tengan presente este elemento.*
- **Objetivo 2:** Evaluar el impacto del componente de autogestión en comparación al resto de elementos del multipaquete conductual.
 - *Hipótesis 2a: Se espera que los estudiantes que reciban como intervención estrategias de autogestión aumenten ligeramente sus respuestas de interacción y participación en el grupo privado de Facebook.*
 - *Hipótesis 2b: Esperamos que los resultados objetivos de los exámenes semanales mejoren al implementar las estrategias de autogestión.*

- Objetivo 3: Evaluar el impacto del procedimiento de aprendizaje cooperativo en comparación al resto de estrategias aplicadas en la intervención BE-Social con retroalimentación semi-inmediata.
 - *Hipótesis 3a: Se prevé observar un aumento objetivo de la participación e interacción de los estudiantes durante la fase de intervención.*
 - *Hipótesis 3b: Esperamos que los alumnos mejoren objetivamente sus calificaciones semanales durante las semanas que reciben la intervención conductual.*

- Objetivo 4: Evaluar el impacto de la intervención multicomponente BE-Social sin retroalimentación semi-inmediata, sino proporcionando retroalimentación con una latencia de más de 72 horas.
 - *Hipótesis 4a: Se prevé observar un aumento objetivo de la participación e interacción de los estudiantes durante la fase de intervención.*
 - *Hipótesis 4b: Se espera que las calificaciones de los alumnos que reciben la intervención, mejoren objetivamente durante la fase de intervención.*

- Objetivo 5: Evaluar el impacto del componente de retroalimentación semi-inmediata proporcionado en el grupo experimental BE-Social, en comparación al resto de procedimientos.
 - *Hipótesis 5a: Se espera que el grupo de estudiantes que reciban el paquete de intervención con retroalimentación semi-inmediata incremente el número de respuestas de interacción y participación obteniendo mayor resultado que con el resto de procedimientos.*
 - *Hipótesis 5b: Se prevé que el incremento de interacción y participación se observe también en los resultados objetivos de las pruebas semanales.*

- Objetivo 6: Analizar la existencia de una correlación entre las respuestas objetivas de interacción y participación semanal generadas en los grupos cerrados de Facebook con las calificaciones objetivas semanales en estudiantes de posgrado que utilizan las redes sociales en entornos educativos.

- *Hipótesis 6a: Esperamos que los estudiantes que obtengan mayor tasa de respuestas de interacción y participación obtengan mayores calificaciones durante el curso académico, en función, de los resultados del estudio anterior.*

La consecución de los objetivos y el contraste de las hipótesis relacionadas se llevan a cabo en cinco estudios independientes, pero conectados entre ellos.

1. El primer capítulo examina la compatibilidad de utilizar análisis visuales y estadísticos en diseños experimentales de caso único (DECU). Los DECU son adecuados para los profesionales que se dedican al ámbito práctico porque permiten evaluar los efectos de una intervención a través del tiempo, tras obtener medidas repetidas de la conducta. Por este motivo, son comúnmente útiles en el análisis aplicado de conducta y además proporcionan la base para realizar un análisis visual y comprender detalladamente el patrón de datos que emerge durante las diferentes condiciones de la investigación. Por ejemplo, en el cuarto estudio se prevé emplear un diseño de línea base múltiple, un subtipo de DECU con el fin de analizar visualmente los datos de interacción y participación conjuntamente con los datos académicos. Uno de los modelos de análisis estadístico compatible con los DECU es el modelo lineal jerárquico o análisis multinivel². Se considera una extensión del análisis de regresión de un solo nivel teniendo en cuenta la medición de otros participantes, característica clave que nos permite determinar aspectos de la variabilidad a través del tiempo. Este modelo se espera emplearlo en el segundo y quinto estudio que se plantean en este capítulo. Este capítulo también hace repaso sobre el uso de g de Hedges en diseños de caso único. Su finalidad es conocer cada tamaño de efecto, siendo entre ellos diferentes y así poder calcular el tamaño de efecto global. Este análisis se prevé realizarse en los dos estudios experimentales-aplicados. Por lo tanto, este capítulo se espera que responda al objetivo planteado en la sección anterior.
2. El segundo estudio evalúa y compara la precisión de observación conductual utilizando dos métodos asistidos por ordenador (Observer XT y Big Eye Observer 2.0) con la observación de papel y lápiz. Doce estudiantes de posgrado se sometieron a un protocolo de entrenamiento de observación conductual para luego registrar 60 sesiones clínicas reales con múltiples conductas a observar aleatorizadas a uno de los

2 Hierarchical linear models (HLM), en la literatura original

tres métodos de observación. El objetivo de este estudio es responder al objetivo y las hipótesis correspondientes, planteadas en el apartado anterior. La validación de una aplicación de dispositivo móvil como método de observación de múltiples conductas permitirá su uso en los siguientes estudios planteados en la tesis como experimentales-aplicados.

3. El tercer estudio es una revisión sistemática que trata de examinar los índices cuantitativos de interacción y compromiso en línea en contextos educativos que integren plataformas de redes sociales. El objetivo de este estudio es cumplir con el primer objetivo e hipótesis para conocer el estado actual de la literatura. Así pues, trata de descubrir qué estudios (cualitativos o cuantitativos) y metodología (ECA o presencia de grupos control) predominan en el análisis de variables académicas como la participación y la interacción en redes sociales. Además, permite crear y elaborar una taxonomía con métricas cuantitativas; a raíz de la literatura existente para luego ser ampliada con otras dimensiones conductuales aplicadas en contextos de interacción y participación en las redes sociales integradas en áreas educativas.
4. El cuarto estudio es un estudio experimental, ocupando así, el vacío encontrado en la revisión sistemática tras la ausencia de estudios de control aleatorizados (ECA). Para ello, primero se elabora un paquete de intervención con procedimientos analítico-conductuales denominado, BE-Social. La intervención consta de: estrategias de autogestión, videomodelado, aprendizaje cooperativo y reforzamiento casi inmediato. Este estudio consiste en aplicar el paquete BE-Social a un total de 46 estudiantes de posgrado que realizan una formación online y utilizan un grupo cerrado de Facebook. El objetivo de este estudio es responder a los objetivos e hipótesis creadas en la sección anterior. Para ello, primero se utilizará un diseño de línea base múltiple donde cada participante sirve como control al ser expuesto de forma sucesiva al paquete de intervención BE-Social y así, registrar la frecuencia total de respuestas de interacción y participación. Respecto al segundo objetivo, la obtención total de respuestas de interacción y participación durante la fase de intervención y los resultados de los exámenes semanales permite verificar la hipótesis del incremento de rendimiento debido a una mayor interacción y participación.
5. El quinto estudio se centra en identificar qué procedimiento o combinaciones de estrategias del paquete de intervención BE-Social provocan un mayor incremento en

las respuestas de interacción y participación en grupos cerrados de Facebook. Para ello, se emplea un ensayo controlado aleatorizado formado por cinco grupos de participantes, de entre un total de 141 estudiantes. A cada conjunto de estudiantes se le asigna un grupo cerrado de Facebook. El grupo 1 recibe el paquete completo BE-Social, es decir, están expuestos a reforzamiento casi inmediato, el grupo 2 se expone al paquete BE-Social pero obteniendo reforzamiento semanal, el grupo 3 se acoge al procedimiento de aprendizaje cooperativo, el grupo 4 experimenta las estrategias de autogestión y videomodelado y el grupo 5 es el control, esto quiere decir que, recibe la misma cantidad de estímulos que los grupos experimentales pero sin utilizar estrategias analítico-conductuales. El objetivo de esta investigación es abordar los objetivos y poner a prueba las hipótesis planteadas en el apartado anterior. Para lograrlo, se obtienen datos diarios y semanales de las respuestas de interacción y participación generadas en los grupos cerrados de Facebook, además de calificaciones semanales tras realizar pruebas de respuestas múltiples del temario.

A continuación, se presenta una tabla con todas las investigaciones que se han llevado a cabo con el fin de responder los objetivos e hipótesis planteadas en este capítulo.

Tabla 2.1 Resumen sobre los estudios conceptuales-metodológicos

Título del estudio	Autores	Estado	Factor de impacto	Diseño	Análisis de datos
Quantifying Outcomes in Applied Behavior Analysis through Visual and Statistical Analyses: A Synthesis.	Virues-Ortega, J., Moeyaert, M., Sivaraman, M., Tarifa-Rodríguez, B., Fernández Castilla, B.	Springer Nature (en prensa)	-	-	-
Accuracy of paper-and-pencil systematic observation versus computer-aided systems.	Virues-Ortega, J., Casas, C. D., Martín, N., Tarifa-Rodríguez, A., Hidalgo, A., Cox, A. D., Navarro Guzmán,	Behavior Research Methods (publicado)	6.2 (Q1)	Diseño de reversión y	Análisis multinivel
Quantitative Indices of Student Social Media Engagement in Tertiary Education: A Systematic Review and a Taxonomy	Aida Tarifa-Rodríguez, Javier Virues-Ortega, Ana Calero-Elvira	Journal of Behavior Education (en revisión)	2.4 (Q2)	Revisión sistemática	-

Tabla 2.2 Resumen sobre los estudios aplicados-experimentales

Título del estudio	Autores	Revista (estado)	Factor de impacto	n	Metodología	Análisis de datos
Effect of the Behavioral Education and Social Media (BE-Social) Program on Postgraduate Academic Performance: A Randomized Controlled Trial	Aida Tarifa-Rodríguez, Javier Virues-Ortega, Ana Calero-Elvira	British Journal of Educational Technologies (en revisión)	4.9 (Q1)	46	Diseño de línea base múltiple concurrente Ensayo controlado aleatorizado	Between-case standardized mean difference ANOVA mixto
Enhancing Academic Performance in Tertiary Education through Social Media: A Multiple-Arm RCT	Aida Tarifa-Rodríguez, Javier Virues-Ortega, Ana Calero-Elvira	British Journal of Educational Technologies (en revisión)	4.9 (Q1)	141	Ensayo controlado aleatorizado múltiple.	Modelo lineal jerárquico

SEGUNDA PARTE
MARCO TEÓRICO

CAPÍTULO 3. Análisis aplicado de conducta en la educación

El propósito de este capítulo es destacar los componentes eficaces y los beneficios que incorporan los principios conductuales en la educación. El análisis de conducta es una ciencia relevante y viable que ofrece una tecnología como guía y base de la enseñanza eficaz. Como introducción al tema, este capítulo proporciona una breve historia del análisis de conducta desde sus inicios hasta lo que conocemos actualmente sobre la educación conductual en el siglo XXI. El estudio de la conducta de interacción y la participación en redes sociales y su posible relación con el rendimiento académico serán abordadas a lo largo de esta investigación desde la perspectiva del análisis aplicado de conducta.

1. EL ANÁLISIS APLICADO DE CONDUCTA

Esta introducción tiene como objetivo exponer la historia de la ciencia de conducta.

1.1 Historia previa al análisis de conducta

El conductismo nace a principios del siglo XX, desde este enfoque la psicología se centra en estudiar la conducta humana como una ciencia natural, predecible y controlable por factores ambientales. Esta aproximación busca comprender la causa de lo que la gente hace o dice (conductas observables) buscando los cambios producidos en el ambiente. Este enfoque presupone que el aprendizaje es adquirido y controlado por el entorno.

El precursor del conductismo fue Edward L. Thorndike (1874-1949), su principal contribución fue descubrir las bases del condicionamiento operante, que más tarde fue refinado por Skinner. Thorndike realizó un experimento con gatos y “cajas problema” o jaulas. Observó que los gatos aprendían a golpear una palanca que había en la “caja problema” para poder salir. Midió el tiempo y se percató que salían cada vez más rápido. Con estos y otros datos, en 1911 anunció la “ley del efecto”, es decir, cuando un estímulo va seguido de una respuesta y ésta de una recompensa, fortalecen las conexiones entre el estímulo y la respuesta. De forma similar, si va seguido de una consecuencia desagradable, las conexiones se debilitan.

Esta teoría fue refinada por B. F. Skinner (1904-1990) y simplificó la *ley del efecto* describiendo lo que conocemos actualmente como condicionamiento operante. La formulación esencial de este nuevo modelo es la contingencia (estímulo antecedente – respuesta- estímulo consecuente), otorgando importancia a los efectos selectivos de las consecuencias como controladores de la aparición futura de la conducta. El aprendizaje de nuevas conductas es el resultado de una conducta espontánea que interacciona con las variables ambientales y su frecuencia de aparición se ve alterada por las consecuencias. De esta manera, las conductas que han sido seleccionadas, mantenidas y moldeadas previamente por las consecuencias pasadas, Skinner las denominó conductas operantes. Por lo tanto, una conducta operante está definida por sus variables antecedentes y consecuentes. De esta noción proviene el término conocido como contingencia operante de tres términos, contribución que supuso una revolución conceptual en la historia del conductismo. En definitiva, Skinner fue un científico notable que impulsó el estudio de la conducta humana y sentó las bases de la metodología del análisis de la conducta.

El análisis de la conducta es la interrelación de tres áreas: (a) el conductismo; la filosofía de la ciencia de la conducta, (b) el análisis experimental de la conducta; la investigación básica y (c) el análisis aplicado de conducta; la investigación aplicada. El análisis experimental de la conducta es un método de análisis científico, fundado por Skinner. Las características de este método incluyen el estudio de la conducta de individuos, la precisa descripción de la conducta y de las variables independientes de las que la conducta es función, el registro automático de las conductas cuando es posible y el estudio de la conducta en entornos controlados. Es un método de tipo inductivo (por oposición al método hipotético-deductivo, común en psicología, pero no tanto en otras ciencias), que estudia personas individuales (no se hacen estudios de grupo ni se hacen promedios). El resultado es que se estudian variables que ejercen una influencia sobre la conducta.

La década de los años 50 y principios de los 60 se caracterizó por determinar si los hallazgos demostrados en el laboratorio podían ser replicados con humanos y así crear tecnologías que mejorasen la educación. El diccionario de la Real Academia Española define tecnología como “conjunto de teorías y de técnicas que permiten el aprovechamiento práctico del conocimiento científico”. Así pues, de los estudios del análisis de la conducta deriva una potente tecnología, el

análisis aplicado de conducta³. Esta tecnología ofrece descripciones detalladas de los procedimientos que se han utilizado para cambiar conductas significativas y socialmente válidas; y ser posteriormente, replicados (Baer, 1968). Fue en 1957, cuando Baer, Wolf y Risley publicaron un artículo en el que describieron las dimensiones del análisis aplicado de conducta. La describió como “aplicada”, dimensión que se refiere al estudio de conductas relevantes y útiles para la persona. “Conductual” porque se centra en la conducta como objetivo de cambio. A nivel empírico la dimensión “analítico” adquiere importancia porque el profesional busca identificar las relaciones funcionales entre la conducta y las variables ambientales por medio del estudio científico. Los procedimientos definidos al detalle y de forma completa, permiten ser replicados y garantizar la efectividad. Esta definición pertenece a la dimensión “tecnológico”.

Los procedimientos aplicados en análisis aplicado de conducta están derivados de los principios básicos de la conducta, por este motivo, los teóricos se refieren a procedimientos “conceptualmente sistemáticos”. La característica “general” hace referencia al uso de estrategias que favorecen la generalización y el mantenimiento del cambio en la conducta. Por último, relacionada con esta última, la dimensión “efectivo”, significa que los profesionales provocan cambios en la conducta que son suficientemente grandes para que tengan un impacto significativo en la vida de las personas. En definitiva, el análisis aplicado de conducta se centra en la conducta en sí misma como objetivo de cambio y utiliza los principios básicos de conducta para adquirir o incrementar conductas socialmente significativas a través de observaciones minuciosas y la experimentación sistemática. Esta característica hace que se diferencie del análisis experimental, dónde la conducta específica no importa, sino que adquiere relevancia el estudio de procesos o relaciones funcionales entre variables.

Es importante recalcar que el análisis aplicado de conducta pretende identificar las relaciones funcionales entre la conducta y los eventos del entorno por medio del estudio científico. Esto significa que busca las variables reales de las que depende la conducta. Se contrapone a la búsqueda de variables hipotéticas (por ejemplo, fundadas en supuestos procesos mentales, como se hace a menudo en la psicología). A lo largo de esta investigación, esta característica es

3 *Applied Behavior Analysis (ABA)*, en la literatura original.

importante porque analizaremos y modificaremos las variables del entorno que afectarán a los estudiantes durante el proceso de aprendizaje.

Desde entonces, el análisis de conducta ha crecido exponencialmente en la investigación básica y en la investigación aplicada. Esto ha supuesto muchos avances, fundamentalmente el descubrimiento de procesos básicos y complejos de aprendizaje, por medio del método científico del análisis de conducta.

2. ANÁLISIS APLICADO DE CONDUCTA EN LA EDUCACIÓN

Abordar cuestiones como la enseñanza y la adquisición de contenidos ha sido un tema de discusión durante muchos años. Skinner declaró que el análisis de conducta podía responder a conflictos que se enfrentaba la humanidad, considerando la educación como herramienta de resolución más importante. Afirmó en su libro, *la Tecnología de la enseñanza* que: “la educación es quizás la rama más importante de la tecnología científica” (Skinner 1968, pág. 19). En 1954, Skinner describió como la conducta se mantiene cuando se refuerza de forma intermitente, en función del tiempo o en función del número de respuestas. Por ejemplo: (1) la conducta se mantiene a una tasa estable cuando el reforzamiento es variable (el reforzador se administra a intervalos variables o en función de un número variable de respuestas); (2) la conducta se mantiene a una tasa más alta cuando depende solo del número de respuestas y no del tiempo. Este proceso recibió el nombre de programas de reforzamiento intermitentes vigentes actualmente en la adquisición de habilidades académicas.

En los años 60 una serie de acontecimientos contribuyeron al inicio formal del análisis aplicado de conducta contemporáneo. Donald Baer expandió los procesos básicos hacia el ámbito educativo y el aprendizaje de conductas nuevas. De entre sus hallazgos, se destacaron: (a) el procedimiento de imitación como un fenómeno de condicionamiento operante, (b) el estudio de la generalización de aprendizajes en nuevos contextos, con personas diferentes y hacia conductas operantes similares y novedosas, (c) numerosos estudios sobre el aprendizaje utilizando procesos discriminativos, y (d) el desarrollo de diseños experimentales de caso único para investigar conductas significativas.

En 1964, Susan Meyer Markle, alumna de Skinner, transfirió los conocimientos de laboratorio a la aplicación práctica educativa desarrollando la instrucción programada. Esta temprana

contribución culminó en 1968 con la publicación de la primera revista para tratar problemas aplicados desde este enfoque, reconocida como *Journal of Applied Behavior Analysis* (JABA). Durante esta década, se acumularon métodos educativos que apoyaban la eficacia de la tecnología de la enseñanza derivada de la ciencia de la conducta. A lo largo de la tesis se expondrán las intervenciones que representaron la ciencia de la conducta en la educación: la instrucción programada, de Skinner (1968), el sistema personalizado de instrucción, de Keller (1968), la enseñanza de precisión, de Lindsley (1960) y el sistema de instrucción directa, de Engelmann (1970).

2.1 Las máquinas de enseñanza

A principios de la década de 1920, Sidney L. Pressey (1888-1979), profesor de psicología educativa de la Universidad Estatal de Ohio, construyó una máquina para proporcionar preguntas de opción múltiple a los estudiantes que recibió el nombre de “máquina para pruebas de inteligencia”. Trató de incorporar las leyes de Thorndike (1912) sobre el aprendizaje: la ley de lo reciente, del efecto y del ejercicio. La ley del efecto afirmaba que una respuesta tenía más probabilidad de aparecer en un futuro tras la contigüidad de una consecuencia reforzante. La ley de lo reciente establece que la última respuesta reforzada tiene más probabilidad de emitirse de nuevo. Por último, la ley del ejercicio sostiene que las conexiones entre estímulos y respuesta se fortalecen con la repetición. Pressey exigía a sus estudiantes seleccionar la respuesta correcta. Tenían que pulsar una tecla con la opción que consideraba correcta, la máquina registraba la respuesta en un contador. Si la respuesta era correcta, pasaba a la siguiente pregunta, de lo contrario, permanecía en la misma hasta acertar. En una de las versiones de la máquina, los usuarios tenían que responder dos veces correctamente para cambiar de pregunta; de esta manera se aplicaban las tres leyes mencionadas. La máquina ideada por Pressey permitía a los estudiantes tener presente la pregunta hasta su selección y proporcionar un resultado inmediato de las respuestas. Esta fue la primera demostración de que una máquina tenía la capacidad de enseñar. No obstante, no fue hasta finalizada la Segunda Guerra Mundial que la enseñanza con máquinas ganara popularidad y atención con el desarrollo de las máquinas de enseñanza diseñadas por Skinner.

Es importante recalcar que las respuestas que se requerían en las máquinas de Pressey y Skinner eran diferentes. En las máquinas de Pressey se presentaban preguntas de opción múltiple y solo permitía una respuesta cerrada. En cambio, en las máquinas de enseñanza de Skinner se requería una respuesta “abierta”, escrita por el estudiante.

Skinner en 1953, visitó la clase de aritmética de su hija y observó la gran cantidad de materiales que presentaban los educadores en una única sesión, la infrecuencia de reforzamiento y de inmediatez; además de la falta de individualización de los materiales. Skinner concluyó que el mayor problema que había presenciado era que se ignoraban los componentes de los principios del condicionamiento operante y del aprendizaje. Decidió solventar esta situación y diseñó las “máquinas de enseñanza”. Las describió como dispositivos mecánicos útiles para aplicar los elementos críticos de una enseñanza eficaz. Las máquinas de enseñanza proporcionaban retroalimentación inmediata después de que el estudiante emitiera miles de respuestas, además de moldear respuestas complejas enseñando discriminaciones sutiles.

Esta nueva tecnología consistía en descomponer y presentar textos de complejidad creciente. Cada texto correspondía con un objetivo conductual medible; identificado previamente. Una vez ya organizado el material, se dividía en marcos secuenciales siguiendo el enfoque operante (contingencia de tres términos, ABC). El marco A, era el antecedente, información que el estudiante necesitaba para completar el siguiente marco, el B. Por lo general, podían ser frases u oraciones incompletas que funcionaban de estímulos discriminativos. El marco B, requería la respuesta del estudiante basado en la información proporcionada en el marco A. El último, el C (la consecuencia), permitía comprobar la respuesta anterior (marco B). Una presentación ordenada y secuenciada del material reducía el número de errores. La máquina anotaba un punto si el alumno respondía correctamente, en cambio, ante una respuesta incorrecta, presentaba de nuevo la serie ABC que había fallado o retrocedía varias series hasta rectificar la respuesta incorrecta. Este paradigma representaba la instrucción lineal programada. Actualmente, la mayoría de programaciones siguen un patrón de ramificación ante una respuesta incorrecta derivando al estudiante a una nueva rama diseñada solo para remediar el error. Los conceptos son desglosados y se presenta nueva información para que el estudiante elabore la respuesta correcta. Una vez responde adecuadamente, el programa le redirige a la secuencia inicial de marcos. Por lo tanto, la distinción entre los programas lineales y ramificados se diferencian en la metodología de responder ante una respuesta incorrecta. En un programa lineal, el estudiante repite el mismo marco que ha fallado y en un programa ramificado, el estudiante se expone a nuevos marcos que le permiten rectificar el error.

De este modo, las máquinas de enseñanza facilitan a los estudiantes avanzar con el contenido académico cuando emiten respuestas correctas y el propio sistema accede a la siguiente

pantalla funcionando como reforzador. También existen programas que entregan un reforzador condicionado, en forma de punto, por cada acierto y una vez acumulados, se canjean por un reforzador preferido. Cada respuesta es una medida directa de los objetivos del programa y la precisión de cada respuesta conlleva a una instrucción lineal o ramificada. Skinner resaltó que el propósito de tecnología mediada con ordenadores era un simple complemento a la enseñanza, no un mecanismo capaz de reemplazar la figura del docente. Estas máquinas presentaban el material académico de forma estructurada y gradual (de más sencillo a más difícil), progresando a medida que el estudiante adquiría conceptos cada vez más complejos.

A finales de la década de los 80, con la revolución de las nuevas tecnologías y el uso personal de los ordenadores, toda la enseñanza asistida por ordenadores estuvo influenciada por los mecanismos que utilizaban las máquinas de enseñanza de Skinner. Esta época fue testigo del auge del uso de ordenadores personales y del desarrollo del “software educativo”; concepto que reemplazó la “enseñanza mediada por ordenadores⁴”.

Hoy día ha desaparecido el uso de las máquinas de enseñanza tal y como las definió Skinner, pero permanece el uso de la tecnología en la educación. No obstante, la gran mayoría de procesos informatizados no se basan en objetivos medibles y secuenciales que sigan patrones lineales o ramificados predeterminados. En cambio, la exposición del contenido es más atractivo (visualmente) para el estudiante.

2.2 Instrucción programada

La presentación del material académico que utilizaban las máquinas de enseñanza diseñadas por Skinner (1958) se conoce como instrucción programada⁵. Una forma menos costosa de aplicar la instrucción programada fue mediante textos de enseñanza o libros bien programados. Tanto las máquinas de enseñanza como los libros programados se regían por los mismos principios conductuales de una instrucción eficaz⁶: (a) énfasis en descomponer una habilidad compleja en unidades que puedan entrenarse con más facilidad. Este proceso recibe el nombre de análisis de tareas, (b) el uso de ayudas para aumentar la probabilidad de que el estudiante emita la respuesta correcta ante un estímulo específico y pueda así, ser reforzada, (c) la aplicación de estrategias

4 *Computer-Aided Instruction*, en la literatura original.

5 *Programmed Instruction*, en la literatura original.

6 *Effective Instruction*, en la literatura original.

de intervención de antecedentes, organizando y presentando los materiales de forma secuencial y gradual, de conceptos simples a más complejos. Esta planificación prevé mayor posibilidad de transferencia de conocimientos. (d) La aplicación de reforzamiento diferencial, permite el reforzamiento de aproximaciones sucesivas hacia la respuesta correcta, procedimiento que recibe el nombre de moldeamiento. (e) El desvanecimiento o la transferencia de las ayudas hasta eliminarlas por completo y el estudiante sea capaz de responder correctamente ante el estímulo discriminativo específico y finalmente, (f) el realce en la retroalimentación inmediata. La adquisición de una respuesta correcta se logra tras recibir un reforzador inmediato, asegurando su probabilidad futura. La inmediatez de la retroalimentación permite mantener un ritmo de trabajo constante. Es importante, destacar la labor que desempeña el docente: estructurar, analizar, clasificar el contenido y programar las diferentes ramas de acción del sistema.

En 1968 Skinner acuñó el término de “tecnología de la enseñanza” para describir la enseñanza programada como una aplicación de la ciencia del aprendizaje a las técnicas prácticas de la enseñanza. A lo largo de los años 60, incrementó su uso y el interés en investigar la instrucción programada utilizando los libros programados y las máquinas de enseñanza. Ambos formatos compartían, esencialmente, las mismas características: presentación estructurada del contenido, exigencia de respuestas activas de los estudiantes y la presentación inmediata de un reforzador, en caso de una respuesta correcta o la exigencia de una nueva respuesta ante una corrección. A fin de cuentas, la instrucción programada es un método de autoinstrucción diseñado para enseñar mediante libros o máquinas. El objetivo es facilitar al estudiante la capacidad de construir la respuesta correcta a raíz de la información previamente presentada y no mediante un proceso de memorización respetando el ritmo de aprendizaje del estudiante.

Si bien las teorías de aprendizaje y las tendencias educativas actuales han cambiado, el efecto que ha perdurado en la tradición de la instrucción programada es la convergencia de las teorías del aprendizaje derivadas de la ciencia básica con la enseñanza práctica. El propio Skinner fue quien debatió este punto en su propio libro “La ciencia del aprendizaje y el arte de la enseñanza” (1954).

Actualmente, los orígenes conductuales y las nociones de la instrucción programada continúan siendo evidentes. Desde la identificación de conductas observables y cuantificables, la realización de un análisis de tareas, la creación de múltiples oportunidades para practicar y entregar retroalimentación a los estudiantes, hasta la evaluación final de todos los objetivos conduc-

tuales. La revolución y proliferación de las tecnologías e internet multiplica las oportunidades de formarse en cualquier momento y lugar, siguiendo el propio ritmo del estudiante. Por lo tanto, la llegada de los ordenadores amplía la idea característica de la instrucción programada.

Uno de los primeros ejemplos de programa informático basado en la instrucción programada es el sistema de “lógica programada para la operación automática de la enseñanza” o PLATO⁷ desarrollado por Donald Bitzer (1960). Este sistema de dominio público ofrecía planes de estudio basados en el ordenador, manteniendo la filosofía conductual de evaluar el progreso de los estudiantes, proporcionar retroalimentación y si es necesario, corrección. PLATO evolucionó hasta convertirse en un software que permitió crear comunidades en línea. A lo largo de las décadas transcurridas y la expansión de softwares y programas informáticos, la instrucción programada sigue siendo un enfoque empírico y eficaz (Huang et al., 2017). Además, han derivado nuevas tecnologías apoyadas en la instrucción programada, como el sistema personalizado de instrucción, la instrucción directa y la tutoría programada.

2.3 Sistema personalizado de instrucción

Fred S. Keller sugirió actualizar los métodos de enseñanza tradicionales inspirándose en la Instrucción Programada de Skinner. Keller junto Gil Sherman, Rodolpho Azzi y Martuscelli Bori perseguían la búsqueda de un patrón metodológico que maximizara el éxito y la satisfacción de los estudiantes progresando sistemáticamente hacia un resultado específico. Desarrollaron en 1963 el sistema personalizado de instrucción (PSI por sus siglas en inglés). Keller se convirtió en un defensor de este sistema, y es por este motivo que también se conoce como Plan o Método Keller. Fue en 1986 que alcanzó su mayor auge con la publicación “Adiós al profesor” (Keller, 1986).

Este diseño organiza y fracciona el contenido de un curso en unidades secuenciales y progresivas. Todas las actividades están individualizadas y los estudiantes determinan su propio ritmo y la cantidad de aprendizaje. Cada unidad constituye una meta a lograr alcanzando un criterio de dominio mediante un examen evaluativo (por ejemplo, 80%, 85%, 90% o 95% de aciertos). En caso contrario, los estudiantes deben de revisar de nuevo la unidad y volver a ejecutar la evaluación para demostrar el dominio del temario. Tanto las respuestas incorrectas o los intentos de exámenes fallidos, no se penalizan. Si la puntuación es satisfactoria, el supervisor le plantea la

7 *Programmed Logic for Automatic Teaching Operations*, en la literatura original.

siguiente lección, de lo contrario, revisan los errores y hacen los ajustes necesarios para volverla a realizar.

Originalmente, el sistema personalizado de instrucción se diseñó para ser impartido en educación superior, pero resultó ser un método de enseñanza innovador que se probó en educación primaria, secundaria e incluso en formación de empresas (Sherman, 1992). El éxito de este diseño se centra en las cinco características que lo definen (Liu, 2001): énfasis en la comunicación escrita, priorizar el ritmo de estudio del estudiante, el dominio del material, el apoyo que ofrecen los supervisores y el uso de conferencias con fines motivacionales.

2.3.1 Énfasis en la comunicación escrita

El objetivo de F. Keller era maximizar la cantidad de respuestas relacionadas con el estudio en contacto con reforzamiento. En un curso que utiliza el sistema personalizado de instrucción, los educadores entregan a los estudiantes una guía de estudio con los objetivos de estudio para ayudarles a focalizar la atención al material que tiene más probabilidad de aparecer en examen, también incluyen consejos para optimizar su rendimiento y las fechas de evaluación. También, incluye actividades para practicar lo estudiado, preguntas de reflexión para que exploren la materia y un listado de lecturas complementarias. En definitiva, entregar la guía del curso y los materiales de estudio en formato papel (por ejemplo, libros o textos) ofrece al estudiante un grado de libertad. El uso de utilizar materiales en formato escrito permite revisarlo en cualquier lugar y momento, detener la lectura y reanudarla tantas veces sea necesaria. Para los estudiantes que presentan dificultades en las habilidades de escritura, esta no sería la mejor estrategia, siendo más efectiva la expresión verbal vocal (oral).

2.3.2 El efecto del ritmo

En relación con el apartado anterior, los estudiantes reciben la guía del curso con las fechas límites de examen y el material didáctico que aprender. Este sistema da lugar a que cada estudiante administre su propio ritmo de estudio. Por lo tanto, una vez se inicia el curso siguiendo el sistema personalizado de instrucción, los estudiantes trabajarán en diferentes unidades según su ritmo de autoaprendizaje. Para lograr el dominio de esta habilidad y culminar con el curso, los estudiantes han de tener habilidades de autocontrol y evitar procrastinar. Por lo general, las instituciones que aplican este sistema suelen tener una tasa elevada de abandono.

2.3.3 El dominio del contenido

En el sistema personalizado de instrucción el contenido se subdivide en diferentes componentes llamados “unidades”. La característica del dominio requiere que los estudiantes demuestren que han aprendido el material de una unidad al alcanzar el criterio predeterminado en una prueba de evaluación. Pueden repetir la evaluación el número de veces que necesiten hasta alcanzar la puntuación predeterminada. Por lo tanto, el sistema de calificación se utiliza como incentivo para promover las conductas de estudio y de rendimiento y no como medida de comparación con el resto de compañeros. En un estudio realizado por O’Neill et al. (1975) examinaron el efecto que tiene dividir una unidad en pequeñas porciones de estudio, en este caso, en 30, 60 y 90 páginas para observar su efecto en el rendimiento. Los resultados demostraron que el tamaño de la unidad afectó al primer intento de las evaluaciones, pero no alteraba el intento final para cumplir criterio de dominio. También observaron que el tiempo en finalizar la unidad era considerablemente mayor que los grupos que tenían pequeñas porciones de material de estudio.

2.3.4 El apoyo de los supervisores

En el Plan Keller, las figuras de los supervisores pueden ser estudiantes de promociones anteriores que reciben créditos por realizar esta labor (supervisores externos) o estudiantes matriculados en el curso que ya han superado dicha unidad (supervisores internos). Se encargan de informarles sobre su rendimiento, ofrecen retroalimentación sobre sus áreas débiles y dirigen a los estudiantes a las siguientes unidades. La incorporación de esta figura es fomentar conductas de aprendizaje y perseverancia mediante la interacción social con los supervisores, intentando que funcionen como reforzadores sociales naturales. Este hecho se demostró en el estudio liderado por Farmer et al. (1972) al dividir 124 universitarios en grupos de cinco. Un grupo no recibía retroalimentación del educador y el resto recibió diferentes cantidades a lo largo de 20 unidades de instrucción. Los resultados corroboraron que la ausencia de supervisión dio lugar a puntuaciones de rendimiento significativamente inferiores en comparación al resto de estudiantes.

2.3.5 Conferencias motivacionales

Las conferencias dentro del Plan Keller se utilizan como una herramienta de motivación para despertar el interés de los estudiantes en temas relacionados con el contenido del curso. Son oportunidades de expandir el conocimiento en las conferencias de investigadores expertos o de renombre, es decir, no forman parte del claustro de profesores comunes del programa. Este

componente, no se ha demostrado que influya en el rendimiento de los estudiantes (Brothen & Wambach, 1998), una posible explicación a este suceso podría explicarse por la falta de evaluar el contenido al que se exponen en las conferencias. El contenido revisado en un examen y calificado inmediatamente tiene más probabilidad de permanecer y ser generalizado en próximas unidades.

Por sus propias características el sistema personalizado de instrucción se lleva a cabo en formaciones de posgrado o especialización profesional con adultos. Para que funcionen los cinco elementos que componen este plan, el sistema requiere la entrega de todo el material necesario al inicio del curso; además de conductas de disciplina y autocontrol del propio estudiante para progresar y avanzar. Desde el inicio, generó interés a cientos de investigadores y profesionales de la educación. Incluso, se creó una revista *Journal of Personalized Instruction* y un centro para intercambiar información con usuarios de este sistema, denominado centro personalizado de instrucción (Sherman, 1992). También, se han llevado a cabo meta análisis para constatar su eficacia en comparación a métodos tradicionales de exposición oral. Por ejemplo, Kulik et al. (1979) examinó 75 estudios sobre el uso del sistema personalizado de instrucción y concluyó que los estudiantes universitarios expuestos a este sistema de aprendizaje obtuvieron resultados estables y elevados sobre el rendimiento en comparación a los estudiantes que recibían otros métodos de instrucción.

2.4 Instrucción directa

La instrucción directa es un modelo sistemático de enseñanza basado en el trabajo de Siegfried Engelmann. A principios de la década de 1960, Engelmann trabajaba en el mundo del marketing y empezó a desarrollar técnicas para enseñar a sus hijos a retener información. Durante estos primeros ensayos llegó a la conclusión que existía una relación entre las estrategias de enseñanza y lo que aprendían. Aplicó este conocimiento juntamente con el investigador Carl Bereiter en el Instituto de Investigación sobre niños excepcionales de Champaign, Illinois (1964-1966). Durante este periodo de investigación, Engelmann desarrolló los procedimientos que hoy día conocemos como instrucción directa. Durante más de 30 años se ha perfeccionado e integrado elementos eficaces de instrucción: objetivos claros, selección de materiales, oportunidades de respuesta suficientes, entrega de guiones, evaluación continua, ritmo adecuado y una retroalimentación inmediata (Hempenstall, 2014). La instrucción directa incorpora todos estos elementos para satisfacer las necesidades de cualquier estudiante, que presentan problemas de aprendizaje, que

tienen necesidades especiales o pertenecen a grupos minoritarios. Este método de enseñanza se sustenta en el supuesto básico de que todo estudiante puede aprender, independientemente de su rendimiento anterior; el error no está en el alumno sino en los métodos de enseñanza.

Una de las características distintivas de este enfoque es el uso de estrategias de enseñanza que permitan un aprendizaje rápido, secuenciado, gradual y generalizable (Adams & Engelmann, 1996). Para lograrlo, es esencial que los educadores identifiquen estrategias generalizables o “grandes ideas⁸” para organizar el currículo académico (Kameenui & Carnine, 1986). Las “grandes Ideas” son aquellos conceptos o componentes centrales de un área que amplían y facilitan la adquisición de nuevos conocimientos. Por ejemplo, si queremos enseñar a un estudiante a leer, es más eficaz dedicar la enseñanza a aprender fonéticamente los sonidos comunes y la habilidad de combinarlos que no enseñar miles de palabras por separado. Antes de iniciar con la Instrucción Directa, los estudiantes realizan una prueba de nivel para identificar qué áreas y habilidades dominan y así determinar el grupo de instrucción al cuál pertenecerá. Generalmente, con grupos de entre 10 y 15 personas se llevan a cabo estrategias conductuales como la respuesta activa de los estudiantes, la presentación de una señal o estímulo discriminativo para responder, el uso de ayudas, el alcance de un criterio de adquisición, la retroalimentación inmediata y la corrección de errores (Slocum & Rolf, 2021).

Las lecciones están diseñadas siguiendo los siete pasos propuestos por Slavin (1995). (1) El primer componente requiere de la atención del estudiante utilizando el repaso del contenido tratado anteriormente (2) El segundo paso, relacionado con el primero, consiste en revisar y comprobar la adquisición del material o la necesidad de enseñarlo de nuevo. (3) El tercer componente consiste en presentar contenidos nuevos, es decir, identificar objetivos académicos, definirlos operacionalmente y desglosarlos en conductas sencillas fáciles de realizar por el estudiante con la ayuda mínima. Conceptualmente, la descomposición de una actividad compleja en subcomponentes, siguiendo un orden secuencial, recibe el nombre de análisis de tareas. Una de las razones de crear un análisis de tareas es determinar objetivamente el “guión” a seguir para el estudiante y especificar su punto de partida. Por lo tanto, el guión es un instrumento pedagógico que los educadores lo utilizan para fomentar la adquisición de nuevas habilidades; asegurando el dominio de conductas prerequisites. El cuarto y el quinto paso tratan de la enseñanza y la práctica independiente de los conocimientos. (4 y 5) El docente presenta el estímulo discriminativo, ofrece una

8 *Big Ideas*, en la literatura original.

ayuda de modelo para modelar la respuesta correcta de los estudiantes. A continuación, se ofrece la misma señal para que respondan al unísono (respuesta coral). Si la respuesta es correcta y responden a la vez, el docente refuerza inmediatamente las respuestas para que entren en contacto con el reforzador natural asociado a la destreza. Si un estudiante se equivoca, el docente corrige la respuesta y repite de nuevo el mismo estímulo discriminativo para repetir toda la secuencia. Si no responden al mismo tiempo, se presenta la misma señal hasta lograrlo. Todos los alumnos han de responder para que el profesor se asegure que nadie está repitiendo la respuesta del compañero. En un esfuerzo de enseñar lo máximo posible, un componente esencial de la instrucción directa es mantener un ritmo rápido de instrucción (Watkins & Slocum, 2004). Finalmente, los últimos componentes: evaluación del rendimiento con retroalimentación y práctica distribuida se trata de evaluaciones formales. (6)

El sexto paso evalúa los conocimientos mediante un examen sobre el contenido aprendido y se ofrece retroalimentación. (7) Y el último paso, la práctica distribuida está diseñada para que el estudiante emerja nuevas relaciones con contenidos nuevos y los integre con lecciones previas. Aproximadamente, una lección corresponde a diez respuestas activas por minuto. El docente determinará avanzar con el contenido según el rendimiento actual. Cada programa instruccional garantiza un índice elevado de respuestas activas y de interacciones con el educador.

Los principios de enseñanza utilizados en la instrucción directa se investigaron en un proyecto educativo de gran escala, denominado, Proyecto *Follow Through* (Adams & Engelmann, 1996). El objetivo era determinar qué métodos eran eficaces para el aprendizaje y el rendimiento de los estudiantes en relación a otros estilos de enseñanza. Watkins (1997) comparó veinte modelos de instrucción diferentes y resultó ser la instrucción directa la más efectiva en las habilidades escolares básicas (matemáticas, lectura, ortografía, lenguaje), en las habilidades cognitivas (comprensión de la lectura y resolución de problemas matemáticos) y en habilidades afectivas (autoconcepto).

3. EDUCACIÓN CONDUCTUAL FUNDAMENTADA EN LA EVIDENCIA: APLICACIONES EDUCATIVAS FOCALIZADAS

El análisis aplicado de conducta investiga las relaciones funcionales entre los estímulos del contexto y las respuestas conductuales. Su vasta literatura contiene técnicas eficaces, sistemáticas y pragmáticas para abordar cambios en conductas socialmente significativas. Los centros edu-

cativos están diseñados para provocar cambios medibles en los estudiantes haciendo uso de estrategias fundadas en la evidencia.

3.1 Estrategias con base en la evidencia

“Las estrategias basadas en la investigación son esenciales para identificar las prácticas educativas eficaces. La investigación, cuando se basa en observaciones y análisis científicos sólidos, proporciona información fiable sobre lo que funciona, por qué y cómo funciona. Esta información es esencial para diseñar una enseñanza eficaz y para demostrar que, de hecho, es eficaz. Por lo tanto, las decisiones responsables sobre lo que es bueno para los alumnos requieren estrategias basadas en la evidencia” (Reyna, 2004, pág. 47).

El conocimiento científico informa de las estrategias socialmente eficaces y válidas por la ciencia. El término “basado en la evidencia” se restringe a intervenciones que cumplen los criterios de un conjunto de estándares y el término “empírico” engloba las intervenciones respaldadas por el conocimiento científico. Por lo tanto, las “intervenciones basadas en la evidencia” hacen referencia a procedimientos objetivos, sistemáticos y rigurosos basados en la observación y experimentación. Su aplicación conduce a analizar datos fiables y válidos con tal de justificar las hipótesis planteadas. Utiliza como medio de evaluación el uso de diseños experimentales definidos al máximo detalle, con tal de poder ser replicados. Cualquier estrategia basada en la evidencia, como mínimo, ha sido previamente aceptada por una revista revisada por pares y aprobada por un comité de expertos.

El sector con mayor tradición en las prácticas basadas en la evidencia, es el sanitario. En 1990, surgió la necesidad de responder a preguntas: “¿Es un hecho que esta intervención es más eficaz y eficiente para los pacientes con un diagnóstico específico?, ¿la aplicación de este tratamiento mejora la calidad de vida de este paciente? En la profesión médica, la evidencia empírica constituye un conjunto de investigaciones sistemáticas y aleatorizadas teniendo en cuenta el bienestar del paciente (Spring, 2007). Al igual que la profesión médica, el uso de prácticas basadas en la evidencia de la educación garantiza la aplicación de técnicas exhaustivas, probadas sistemáticamente en múltiples investigaciones. En este sentido, sirven como guía y orientación para la toma de decisiones de los profesionales, basándose en el rigor científico y no en enfoques de ensayo y error o en opiniones subjetivas (Brown et al., 2017). Los profesionales de la docencia deben vincular la investigación y la enseñanza en las aulas (Bauer, 2012).

Comprometerse con la investigación en el sector educativo permite mejorar las estrategias de enseñanza de los docentes, indagar nuevos canales de comunicación con los estudiantes y observar su impacto y examinar nuevas perspectivas para futuras investigaciones en este ámbito. Esto también conduce a comprender aspectos que puedan obstaculizar el aprendizaje en la educación (Hempenstall, 2014).

3.2 Gestión del aula

Los factores contextuales de las aulas y las interacciones entre estudiantes y maestros influyen en la capacidad de impartir una enseñanza eficaz. Los profesionales que emplean estrategias de gestión del aula⁹, utilizan un conjunto de prácticas basadas en la evidencia con el fin de maximizar el tiempo de instrucción, minimizar el número de interrupciones o conductas disruptivas y evocar conductas cooperativas entre los estudiantes (Sugai & Horner, 2002). Existe un gran número de habilidades que garantizan la eficacia de la gestión en el aula. No obstante, pueden organizarse en cuatro bloques: (a) estrategias que actúan en los antecedentes (prevención), (b) estrategias de instrucción e interacción del maestro, (c) estrategias que actúan en las consecuencias y (d) estrategias de autogestión¹⁰ (Brandi Simonsen et al., 2008).

Este apartado se centra en las estrategias que actúan en los antecedentes más eficaces y útiles, desde la perspectiva de la evidencia conductual. El motivo por el cual se ha decidido enfatizar este componente es porque a largo de este capítulo se comentan con más detalle el resto. Por ejemplo, más adelante se menciona el procedimiento de videomodelado, que forma parte de las estrategias de instrucción e interacción del maestro, también habrá una sección destinada a explicar la importancia de las habilidades de autogestión y la retroalimentación.

El grupo categorizado como estrategias sobre los antecedentes incluyen procedimientos que el maestro aplica antes de que aparezca una conducta problema. De hecho, la literatura hace referencia a prácticas de prevención; es decir, compuestas por prácticas que se centran en la organización, estructura del aula, la disposición de los elementos del contexto, el establecimiento de reglas y expectativas. En concreto, profundizaremos en (a) añadir opciones de instrucción e (b) incrementar la tasa de oportunidades de respuestas de los estudiantes.

9 Classroom Management, en la literatura original.

10 Self-management, en la literatura original

3.2.1 Opciones de enseñanza

Incorporar opciones de enseñanza y que el estudiante tenga la posibilidad de elegir entre varias opciones, se trata de una conducta ética por parte del profesional que respeta la dignidad del estudiante (Bailey & Burch, 2019, pág. 22). La elección aumenta la motivación para realizar la tarea cuando el estudiante disfruta de formar parte del proceso y se beneficia del resultado (Patall, 2012). Las intervenciones de elección están diseñadas para que la tarea deje de ser un estímulo aversivo incorporando estímulos o actividades preferidas (Dunlap et al., 1994). Para ello, los maestros pueden ofrecer un menú de opciones para que los estudiantes elijan el orden en el que realizan las tareas o bien pueden ofrecerles la posibilidad de que seleccionen los materiales didácticos con el que trabajar la tarea o cambiar de ubicación dentro del mismo contexto (por ejemplo: en la mesa grande, en la alfombra, en una mesa pequeña, etc.). Se trata de una estrategia que requiere de tiempo y de disponer diferentes materiales, este motivo hace que sea fácil y de uso frecuente. Varios estudios sugieren que proporcionar opciones fomenta el compromiso con la tarea y reduce la aparición de interrupciones y conductas desafiantes en el aula (Skerbetz & Kostewicz, 2013). Por ejemplo, la revisión sistemática de Royer et al. (2017) analizaron 26 estudios, 12 de los cuales informaron de un incremento en la conducta de participación y finalización de tareas sin presentar interrupciones.

3.2.2 Tasa de oportunidades de respuesta

Las oportunidades de respuesta es una técnica del análisis de conducta, eficaz y eficiente que utilizan los maestros para aumentar la participación activa de los estudiantes en el aula y reducir conductas inapropiadas (Lane et al., 2015). Consiste en proporcionar frecuentes oportunidades de respuesta presentando contenidos ya adquiridos o de comprensión básica durante un tiempo predeterminado. Lo ideal es crear un ritmo de instrucción lo suficientemente rápido, por lo general se practican tres respuestas por minuto, para mantener la motivación de los alumnos (Conroy et al., 2008). Esta técnica se emplea al llevar a cabo la instrucción directa; ya que el maestro puede elegir que le respondan al unísono o individualmente. Además, puede variar el formato de respuesta de los alumnos indicando que sea verbal, escrita, mediante dispositivos electrónicos o señales con las manos ante respuestas correctas e incorrectas (pulgares arriba o abajo). Se ha demostrado que ofrecer un número elevado de oportunidades de respuesta fomenta la fluidez (Haydon et al., 2012). Un estudio realizado por Partin et al. (2009) constató que un elevado uso de oportunidades de respuesta promueve mayores niveles de elogio por parte del maestro, un

incremento en la conducta de participación de los estudiantes, respuestas más precisas en relación al contenido académico y menores tasas de conductas disruptivas tanto en alumnos de primaria, secundaria y bachillerato. Además, en un estudio reciente liderado por Bolt et al. (2019) corroboraron que el éxito de esta técnica es replicable con estudiantes de educación especial.

La aplicación de estrategias sobre la gestión del aula promueve un ambiente de trabajo cooperativo, fomenta la participación de los estudiantes con el contenido académico y reduce cualquier interrupción y la aparición de conductas disruptivas. Para que las prácticas sean eficaces, los maestros han de desarrollar las competencias necesarias en la aplicación del aula y recibir supervisión continua.

3.3 Aprendizaje cooperativo

Desde un enfoque analítico-conductual, el aprendizaje cooperativo¹¹ es un método de enseñanza basado en la evidencia que motiva a los estudiantes a compartir recursos utilizando un formato de enseñanza que utiliza la aplicación de contingencias grupales (Johnson & Johnson, 1982). Las contingencias grupales son aquellas en las que una consecuencia común es contingente a la conducta de un estudiante, de un conjunto de estudiantes o de todos los miembros del grupo. Existen tres tipos de contingencias grupales: las contingencias dependientes, independientes e interdependientes. En las contingencias independientes, se presenta la contingencia a todos los miembros del grupo, pero solo aquellos que cumplan con el criterio descrito en la contingencia recibirán el reforzador. En cambio, en la contingencia dependiente, el grupo entero accede al reforzador si un miembro del grupo es capaz de alcanzar el criterio. En el caso del aprendizaje cooperativo, se aplica la contingencia interdependiente, es decir, todos los miembros del grupo han de alcanzar el criterio, así pues, han de cumplir con una contingencia individual para que el grupo entero tenga acceso a la recompensa o reforzador (Popkin & Skinner, 2003). Este tipo de contingencia, da lugar a que todos los miembros del grupo se ayuden y cooperen entre ellos para que cada uno alcance su objetivo y todo el grupo tenga acceso al reforzador. Por ejemplo, si el reforzador de la contingencia grupal es tener cinco minutos más de recreo y su obtención depende de la respuesta individual de cada miembro, esta situación fomenta el aprendizaje individual y respuestas cooperativas como grupo. No obstante, es probable que el éxito de que haya aprendizaje individual y respuestas interactivas esté determinado por la entrega contingente de

11 Cooperative Learning, en la literatura original.

una consecuencia reforzante después de que cada miembro cumpla con el criterio de instrucción (Slavin, 1995). El análisis aplicado de conducta debería de poner énfasis en este modelo para averiguar qué componentes del aprendizaje cooperativo, o bien los reforzadores, la responsabilidad otorgada a los estudiantes o las conductas interactivas son las responsables de los resultados positivos.

Los estudiantes adquieren un rol activo de su propio aprendizaje, han de cumplir el criterio participando e intercambiando información y recursos durante la realización de la actividad propuesta (Ellerani & Gentile, 2013). No obstante, el educador: (a) identifica las conductas objetivo de cada miembro del grupo, decide el tamaño del grupo, los materiales y reforzadores necesarios para llevar a cabo la actividad, (b) detalla la tarea y el criterio de éxito; especifica qué respuestas espera de cada miembro del grupo, (c) interviene en los grupos, en caso de ser necesario, para proporcionar cualquier tipo de ayuda que les permita concluir la tarea con éxito, (d) recopila datos de las respuestas sociales y académicas que espera de cada estudiante trabajando en grupo, (e) supervisa y evalúa el producto final realizado. El aprendizaje cooperativo implica la realización de actividades cuidadosamente estructuradas que permitan a los miembros del grupo sugerir, comentar, reflexionar y ayudarse mutuamente (Bećirović et al., 2022). En los últimos años, esta técnica de enseñanza ha comenzado a aplicarse en centros de educación infantil, primaria, secundaria y superior porque fomenta la conducta social, la integración de estudiantes (Buchs & Butera, 2015) y el rendimiento académico (van Ryzin et al., 2020). Además, es útil con personas con y sin discapacidad, independientemente del nivel académico (Roseth et al., 2008).

3.4 Tutoría entre iguales

En las décadas del 1980 y 1990, cuando el análisis aplicado de conducta no se utilizaba en el contexto educativo, la tutoría entre iguales¹² se consideró una estrategia de enseñanza derivada del aprendizaje cooperativo. Con más de 40 años de investigación, esta técnica consiste en la enseñanza y el desarrollo de habilidades académicas entre dos estudiantes (tutor y alumno) durante la práctica (Choudhury, 2002). Posteriormente, Axelrod y Greer (1994) corroboraron que la tutoría entre iguales tenía componentes idénticos al aprendizaje cooperativo, hecho que argumentaba su eficacia demostrada en una variedad de áreas académicas, con estudiantes con y sin discapacidad (Bowman-Perrott et al., 2013) y de diferentes edades (Leung, 2015); incluido

12 Peer tutoring, en la literatura original.

preescolar (Morgan et al., 2020). La investigación de los factores comunes desde un enfoque conductual y desde una enseñanza efectiva resultó estar asociado a la respuesta activa del estudiante¹³ (Barbetta, 1993) mencionada en el apartado de instrucción directa (Kinder & Carnine, 1991), la oportunidad de responder¹⁴ (Sutherland & Wehby, 2001) y las unidades de aprendizaje (Greer & McDonough, 1999), comentadas anteriormente en el apartado de sistema personalizado de instrucción.

3.4.1 Respuesta activa del estudiante

La respuesta activa del estudiante es una medida directa para conceptualizar objetivamente la participación de los estudiantes en el aula. Existen diversas estrategias para incrementar la participación de los estudiantes e involucrarlos en las tareas, independientemente del número de alumnos: respuestas al unísono o respuesta coral¹⁵ (Haydon et al., 2013), uso de tarjetas de respuesta (Christie, 2003) y plantillas de guía (Haydon et al., 2011). Las aulas que generan respuestas académicas activas suponen una dedicación total del maestro en suscitar numerosas oportunidades de respuesta y una cantidad considerable de tiempo de los estudiantes respondiendo a materiales escritos y verbales. Cuántas más oportunidades tenga el estudiante para responder, más respuestas activas emitirá y mayor probabilidad de recibir retroalimentación correctiva; incrementando así la probabilidad de aumentar el rendimiento académico (States, 2019).

3.4.2 Oportunidades de respuesta

Los términos “oportunidades de respuesta” y “respuesta activa del estudiante” están interrelacionados entre sí. Debido a la presencia de una oportunidad para responder, el estudiante emite la respuesta. Greenwood (1984) descubrió que el aprendizaje era el producto de una interacción entre los estímulos antecedentes ambientales y las respuestas activas de los estudiantes. Existe literatura que demuestra que los maestros y padres que ofrecen mayor número de oportunidades de respuesta, el aprendizaje se duplica. Por ejemplo, la revisión de la literatura propuesta por Sutherland y Wehby (2001) concluye que las aulas con estudiantes con trastornos emocionales y de conducta tienden a presentar una tasa reducida de oportunidades de respuesta, lo que conlleva a bajos resultados sobre rendimiento académico y una alta presencia de conductas dis-

13 Active Student Response (ASR), en la literatura original.

14 Opportunity to Respond, (OTR), en la literatura original.

15 Choral Responding, en la literatura original.

ruptivas. Un posible método para aumentar la tasa de oportunidades de respuestas en este tipo de aulas es utilizando las tarjetas de respuesta (Armendariz & Umbreit, 1999). Greenwood et al. (1989) desarrollaron la tutoría entre iguales de toda la clase¹⁶ que duplicaban las respuestas activas de los estudiantes, mejoraban los resultados académicos y reducían las conductas inapropiadas en el aula.

3.4.3 Unidad de aprendizaje

Las unidades de aprendizaje consisten en contingencias entrelazadas entre el maestro y el estudiante. En su práctica, el estudiante aprende la contingencia de tres términos (antecedente, respuesta y consecuencia) dentro del marco de enseñanza. A lo largo del aprendizaje del estudiante, el maestro presenta unidades de aprendizaje cada vez más complejas, es decir, incorpora progresivamente todos los componentes necesarios para llevar a cabo una habilidad. Por ejemplo, para leer una oración, las primeras unidades de aprendizaje fueron enseñar los sonidos, más tarde, la lectura silábica, luego palabras de tres y cuatro sílabas, así sucesivamente hasta que el estudiante sea capaz de leer palabras compuestas de cinco y seis sílabas con una precisión del 100%. En la última etapa, la unidad de aprendizaje sería la lectura fluida y precisa de una oración formada por 6 palabras, de entre cuatro y cinco sílabas como máximo.

Es importante recalcar que la unidad de aprendizaje ha de contener la progresión de todos los componentes del estudiante y del maestro; para así garantizar el éxito de aprendizaje. Se ha demostrado que las tutorías entre iguales que contienen unidades de aprendizaje son más eficaces en comparación a aquellas que no las utilizan en el aula (Polirstok & Greer, 1986).

En definitiva, la evidencia hasta la fecha muestra que la tutoría entre pares aumenta el rendimiento de los estudiantes y los tutores, pero los que muestran mayor beneficio son los que actúan de tutor (Leung, 2019).

3.5 Economía de fichas

La economía de fichas es una estrategia común para gestionar el aula basándose en la evidencia conductual (Simonsen et al., 2008). Es un sistema para gestionar contingencias en el que las personas son recompensadas al emitir conductas objetivo previamente especificadas y preestablecidas (Petry, 2011). Implica el uso de un contrato entre el maestro y el estudiante dónde se

16 *Classwide peer-tutoring*, en la literatura original.

especifican las conductas objetivo que ha de alcanzar el estudiante y el número de fichas u objetos simbólicos que necesita para acceder a un objeto o actividad preferida por el estudiante. Cada emisión de una respuesta objetivo conlleva a la obtención de una ficha o un reforzador condicionado. Estos objetos simbólicos tienen el propósito de reforzar diferencialmente el desempeño de la conducta objetivo y hacer más llevadera la demora en obtener el reforzador final o el acceso a un privilegio. Al conseguir un número determinado de fichas, el estudiante podrá intercambiarlas por un reforzador específico. El objetivo de la economía de fichas es potenciar las conductas adecuadas o deseables y simultáneamente, reducir las conductas problemáticas.

Los investigadores también han utilizado estos programas con personas diagnosticadas con esquizofrenia (Dickerson et al., 2005), trastorno del espectro autista (Samburgo, 2017), trastorno negativista desafiante y otras discapacidades intelectuales. También se han utilizado con adultos que residen en centros residenciales (Murray & Sefchik, 1992), para fomentar la educación vial (Ludwig & Geller, 1991) y con adultos que presentan adicción a sustancias estupefacientes (Budney et al., 1991). Sin embargo, la mayoría de la literatura se centra en el uso de la economía de fichas dentro del contexto educativo. En este ámbito el objetivo es incrementar conductas apropiadas como las habilidades sociales, el rendimiento académico, el tiempo de estudio y la asistencia a la escuela (Kowalewicz & Coffee, 2014). Actualmente, Kim et al. (2021) ha realizado una revisión sistemática de los estudios publicados que utilizan el sistema de economía de fichas en el ámbito escolar entre el 2000 y el 2019 con estudiantes de infantil hasta secundaria. Los resultados sugieren que se trata de un procedimiento eficaz tanto en aulas de educación general como de educación especial, presentando mayores efectos en las aulas de educación especial en comparación con las de educación general. Este hallazgo amplía y especifica la hipótesis citada en la revisión sistemática liderada por Ivy et al. (2017).

3.6 Entrenamiento de habilidades conductuales

El entrenamiento en habilidades conductuales¹⁷ (EHC) es un enfoque de intervención basado en la evidencia para enseñar o capacitar a profesionales sobre una amplia gama de habilidades (Parsons, 2013). Está formado por cuatro componentes: instrucciones (verbales o escritas), modelado, ensayo y retroalimentación (Ward-horner et al., 2012). Se diferencia del resto de métodos tradicionales porque en el EHC el maestro proporciona oportunidades de desempeño al estu-

17 *Behavior skills training*, en la literatura original.

diante para que practique la conducta y reciba retroalimentación sobre su rendimiento durante el ensayo. Se ha demostrado que ambos componentes incrementan la generalización de las habilidades y la integridad de procedimiento (Sawyer et al., 2015). El objetivo final del entrenamiento en habilidades conductuales es que el estudiante adquiera nuevas habilidades y las emita en entornos naturales sin necesidad de ayudas externas ni supervisión.

3.6.1 Instrucciones

Las instrucciones consisten en la descripción exacta de la conducta apropiada y la importancia de la adquisición de la habilidad. Las instrucciones deben ser claras, concisas y deben especificar bajo qué circunstancias contextuales se espera que muestre la habilidad. Ofrecer instrucciones en formato escrito es una excelente ayuda visual para describir el análisis de tareas necesario para la adquisición de la conducta objetivo.

3.6.2 Modelado

Es una estrategia de cambio conductual que muestra o expresa exactamente la habilidad que se espera que el estudiante emita. Por lo tanto, los estudiantes aprenden nuevas habilidades imitando modelos que demuestren habilidades en vivo o a través de un video, foto o audio. Para que se logre el aprendizaje, el estudiante ha de tener la capacidad de imitar, es decir, prestar atención y reproducir la misma conducta que acaba de observar. La presentación del modelado en vivo implica que el profesional muestra la conducta exacta que ha de ejecutar el estudiante. Este modus operandi puede presentar una serie de desventajas, por ejemplo, la presencia de un público puede modificar la conducta del estudiante, también el profesional puede olvidarse de presentar un aspecto relevante de la habilidad por estar atendiendo a un estímulo externo no controlado. Es por este motivo, que la presentación de modelos simbólicos, como a través de un video son eficaces en estas situaciones. La presentación de un modelo utilizando un video, recibe el nombre de videomodelado. Este procedimiento se detalla más adelante. Es importante recalcar que la conducta modelada ha de ser entrenada o bien en un contexto simulado de role-play que incluya estímulos discriminativo relevantes del contexto apropiado dónde se espera que se realice la habilidad o bien en el contexto real. Después de observar el modelo, el estudiante ha de tener la oportunidad de ponerla en práctica para que entre en contacto con el reforzamiento lo antes posible.

3.6.3 Ensayo conductual

El ensayo conductual se trata de una oportunidad que le otorga el profesional al estudiante para que practique la conducta que acaba de observar o de recibir las instrucciones. Los estudios que emplean este procedimiento acostumbran a presentar múltiples ensayos, entre tres y veinte, de unos dos y cinco minutos. Este componente es de vital importancia porque se trata de una oportunidad para evaluar su desempeño y reforzar la conducta o bien, corregirla hasta moldearla.

3.6.4 Retroalimentación

La retroalimentación o feedback como último componente del EHC consiste en entregar reforzadores sociales (elogio, el más común) y utilizar instrucciones adicionales para moldear la correcta ejecución en ensayos posteriores. Desde el análisis de la conducta se trata de aplicar el procedimiento de reforzamiento diferencial, es decir, reforzar la conducta si la ejecución es correcta y ofrecer correcciones si el desempeño lo precisa. Para que la retroalimentación sea eficaz ha de tener una serie de características: inmediata, descriptiva y positiva. La retroalimentación inmediata ha demostrado ser eficaz y la preferida por los profesionales (Fajfar et al., 2012). La presentación de elogios cuando el desempeño es correcto consiste en describir con precisión lo que hizo o dijo el estudiante para enfatizar qué rasgos requiere la ejecución de la habilidad. Al proporcionar retroalimentación correctiva el comentario ha de empezar y terminar con aspectos positivos de su actuación. En el caso de que el estudiante haya realizado una serie de conductas incorrectas, el profesional se focalizará en una, hasta su correcta ejecución. Otra forma de entregar retroalimentación es utilizando y analizando gráficas sobre el rendimiento del estudiante o bien utilizando el lenguaje escrito.

3.6.5 Aplicaciones

Numerosos profesionales e investigadores han aplicado este paquete de intervención para enseñar una gran variedad de habilidades. Los estudios destinados a la población infantil con y sin discapacidad enseñaban habilidades para prevenir conductas de abuso sexual (Miltenberger, 1999), secuestros (Miltenberger & Olsen, 1996), juegos con armas (Miltenberger et al., 2004) y habilidades de emergencia en caso de incendios (Jones et al., 1981). Por ejemplo, Warzak y Page (1990) enseñaron conductas de rechazo ante propuestas sexuales en chicas adolescentes diagnosticadas con discapacidad intelectual.

Existe una amplia gama de estudios que han utilizado los procedimientos de entrenamiento en habilidades conductuales para formar a profesionales y padres en diferentes procedimientos del análisis aplicado de conducta, por ejemplo, en la aplicación de ensayos discretos (Clayton & Headley, 2019). El ensayo discreto es un método de enseñanza intensivo que se utiliza en las primeras fases de intervención temprana con niños con trastorno del espectro autista (TEA). Su aplicación facilita el desarrollo del control instruccional, es decir, la adquisición de conductas prerequisites para que el estudiante atienda al adulto y responda ante instrucciones de aprendizaje. El estudio reciente de Clayton y Headley (2019) sugiere que el paquete de EHC es útil para entrenar al personal en procedimientos de ensayos discretos de forma precisa, rápida (10 minutos de formación) y perdurable en el tiempo. Otros procedimientos que se han beneficiado del EHC son la evaluación de preferencias (Pence, 2012), en el entrenamiento de mandos (Nigro-Bruzzi & Sturmey, 2010) y en la realización de análisis funcional (Lloveras, 2022), entre otros. En este último estudio, Lloveras y colaboradores añaden un componente, en comparación a todos los estudios realizados hasta el momento. Los profesionales recibieron el entrenamiento a través de una pantalla, es decir, utilizaron el formato online. Los resultados del estudio fueron positivos; ya que todos los participantes lograron el dominio de las habilidades.

Kirkpatrick et al. (2019) elaboró una revisión sistemática y resultó que la mayoría de los estudios que habían utilizado el entrenamiento en habilidades conductuales estaban destinados a la formación de maestros de educación especial. Por ejemplo, un estudio reciente realizado por Ethington et al. (2021) evaluó los efectos del EHC en la enseñanza de habilidades de autoaprendizaje en niños de primaria de un aula de educación especial. La adquisición de esta conducta permite una mayor independencia a la persona, por ejemplo: aprender a montar un avión de madera siguiendo los pasos explicados en un video de YouTube. En este caso, los estudiantes tenían que navegar por el iPad y acceder a unos videos de danza. Los resultados demostraron que el uso del EHC contribuyó a la adquisición de una serie de habilidades necesarias para acceder a los videos de danza y completar las tareas. Además, observaron que aquellos estudiantes que recibieron entrenamiento tecnológico mediante el EHC, generalizaron las habilidades.

Por lo general, todos los estudios disponibles en la literatura emplean el paquete de intervención para la enseñanza de diferentes habilidades, independientemente de su complejidad. Un área de investigación futura podría destinarse a la evaluación e identificación de los componentes eficaces que forman el EHC. Parece poco probable que las instrucciones funcionen por si solas,

como elemento modificador de una conducta o el modelado no tendría la misma eficacia sin la retroalimentación. La posible realización de un análisis de componentes podría ahorrar tiempo en la formación de profesionales y padres.

3.7 Videomodelado

El Videomodelado¹⁸ es una estrategia innovadora con evidencia que requiere del uso de la tecnología (Hughes & Yakubova, 2019). Consiste en grabar un video de una habilidad específica de forma sistemática y explícita realizada por un profesional, un experto o un igual que domine la conducta. El objetivo implica mostrar el clip de video al estudiante antes de que complete la actividad con el fin de que imite la habilidad que visualiza en el video (Charlop-Christy et al., 2000). El modelo visual suele grabarse con cualquier dispositivo con cámara que luego permita editar el clip y eliminar cualquier distractor del ambiente (variables extrañas) (MacDonald et al., 2009). Esta estrategia ofrece al estudiante la oportunidad de observar la contingencia discreta: los antecedentes, la habilidad específica y las variables consecuentes, facilitando la comprensión de cuándo y porqué emitir dicha conducta.

Son muchas las investigaciones que han demostrado que se trata de una estrategia eficiente y eficaz que permite recrear entornos que no son posibles de acceder en vivo o ilimitadamente. Esta característica lleva veinte años siendo un procedimiento de preferencia para la enseñanza de habilidades en niños y adolescentes diagnosticados con trastorno del espectro autista (Ayres & Langone, 2005). Recientemente, Qui et al. (2018) llevó a cabo una revisión sistemática con estudios de caso único publicados entre el 1985 y el 2015. Se analizaron veinticuatro estudios y corroboraron que el videomodelado mejora considerablemente las habilidades comunicativas y sociales con personas con TEA. Adquirir una habilidad compleja utilizando videomodelado puede convertirse en una tarea sencilla al desglosarla en varios componentes y grabarla paso por paso. Esta característica permite que el estudiante centre la atención en la habilidad y promueva la imitación y la generalización de la conducta (Corbett & Abdullah, 2005). Según la revisión de Bellini y Akullian (2007) los niños con TEA pueden aprender habilidades sociales, de juego, comunitarias (Helbig et al., 2021) y académicas (Yakubova et al., 2020). Por ejemplo, Acar et al. (2017) muestra que el videomodelado es un procedimiento eficaz para enseñar habilidades sociales a niños con TEA y fácil de administrar por los padres. Otro contexto en el que se ha

18 Videomodeling, en la literatura original.

mostrado su utilidad ha sido en el juego simbólico (MacDonald et al., 2009) y social. Por ejemplo, Cardomy & Stauch (2020) combinó el videomodelado con el entrenamiento matricial¹⁹. El entrenamiento mediante una matriz permitía la transferencia de habilidades a aquellos que no habían recibido instrucción previa. El videomodelado en este estudio ofreció un ritmo de adquisición, cada vez más rápido.

Hasta la fecha, la investigación no ha examinado el potencial de esta estrategia en la enseñanza universitaria con estudiantes de desarrollo típico. Las características del videomodelado podrían beneficiar a estudiantes y profesores para modelar hábitos de estudio, entre otros.

3.8 Autogestión

La autogestión es una estrategia basada en la evidencia que enseña al estudiante a supervisar y modificar su propia conducta (Koegel et al., 1999). El objetivo final del profesional es dotar a los alumnos de estrategias que puedan aplicarlas de forma independiente en diversos contextos. Se ha demostrado su efectividad en áreas educativas, sociales y vocacionales. Sin embargo, en un estudio reciente realizado por Barczak & Cannella-Malone (2021) concluyeron que es necesaria más investigación en la enseñanza de habilidades profesionales en personas con discapacidad intelectual utilizando autogestión. Identificar qué componentes son los más eficaces permitiría avanzar la literatura y proveer cambios significativos. En cambio, el uso de la autogestión en entornos educativos ha adquirido mayor reconocimiento y alcance en los últimos años (McDougall et al., 2017). La mayoría de estudios demuestran su eficacia en contextos escolares con niños y adolescentes con discapacidad intelectual y otros trastornos del desarrollo. Por ejemplo, el metaanálisis de Carr et al. (2014) demostró que las intervenciones de autogestión son eficaces para mejorar habilidades sociales y académicas en estudiantes de entre cinco y veinte años con TEA. Sin embargo, existe poca evidencia sobre la enseñanza de habilidades académicas en secundaria con adolescentes diagnosticados con TEA. Un estudio reciente elaborado por Roberts et al. (2019) con jóvenes de instituto diagnosticados con TEA demuestra que las estrategias de autogestión son eficaces. Además, expresa que los resultados mejoran al combinarlas con un componente de reforzamiento mediado por los compañeros²⁰. Las estrategias de autogestión también facilitan la inclusión de estudiantes con discapacidad en escuelas de educación general debido a

19 *Matrix training*, en la literatura original

20 *Peer mediated-intervention*, en la literatura original.

la presencia de conductas objetivo en el aula y de una intervención directamente en el contexto (McDougall, 1998). Son escasos los estudios que utilizan técnicas de autogestión para mejorar habilidades académicas en estudiantes universitarios de desarrollo típico. En este caso sería beneficioso por la cantidad de ventajas que presentan. En primer lugar, reduce la cantidad de tiempo requerido por el instructor porque es el estudiante el responsable de monitorizar su propia conducta (McDougall, 1998). Se ha demostrado que este procedimiento tiene éxito a largo plazo porque facilita la independencia y el mantenimiento de habilidades y aumenta las probabilidades de aplicarlas en múltiples entornos (Carr et al., 2014). Además, se ha analizado la validez social con profesores (Guerasko-Moore et al., 2006) y estudiantes (Briesch & Daniels, 2013) resultando ser una intervención aceptada y apropiada para obtener cambios en conductas académicas.

Para desarrollar intervenciones que empleen estrategias de autogestión es necesario integrar múltiples componentes conductuales. Cooper (2019) sugirió que la máxima eficacia se conseguía observando y registrando la propia conducta, es decir, la automonitorización²¹, comparando la conducta observada con un criterio objetivo o estándar, es decir, autoevaluándose²². Una forma de promover la autoevaluación es identificando y definiendo las conductas a incrementar²³. Finalmente, el último componente de la autogestión es la recompensa de uno mismo al cumplir o superar su estándar de rendimiento, componente que recibe el nombre de autorreforzamiento²⁴.

3.8.1 Automonitorización

Las tareas de automonitorización requieren que el estudiante observe y registre la ocurrencia o no ocurrencia de sus conductas objetivo. Este procedimiento está formado por tres etapas: (1) Autoobservación. Etapa d/onde el estudiante discrimina su propia conducta, (2) Autorregistro²⁵. El estudiante toma datos sobre la frecuencia o duración de las ocurrencias de las respuestas objetivo. (3) Autograficado²⁶. Paso que requiere la transferencia de datos a una gráfica (Hallahan et al., 1979). La automonitorización promueve la autoevaluación porque ofrece una representación

21 Self-monitoring, en la literatura original.

22 Self-evaluation, en la literatura original.

23 Goal setting, en la literatura original.

24 Self-reinforcement, en la literatura original

25 Self-recording, en la literatura original

26 Self-graphing, en la literatura original

visual del progreso del estudiante. También puede funcionar como técnica de cambio de conducta y provoca un cambio de dirección en los datos hacia una tendencia deseada. Este fenómeno se conoce como “reactividad” y se debe al simple hecho de observar y tomar datos de la propia conducta (Rosenbaum & Drabman, 1979). Parece ser que la exactitud de los registros no es una variable que afecte a la tendencia; un registro poco preciso puede dar lugar a un cambio de conducta deseado (Reinecke et al., 1999).

La literatura se ha centrado en investigar qué es más eficaz, la automonitorización de respuestas de prestar atención o la automonitorización del rendimiento. Ambos procedimientos han dado lugar a resultados positivos sin diferencias significativas, siendo la automonitorización del rendimiento el de mayor preferencia por los estudiantes. En cambio, Harris et al. (2005) comparó ambos procedimientos con estudiantes con déficit de atención e hiperactividad y resultó obtener mayores avances utilizando el autorregistro de rendimiento. Shimabukuro et al. (1999) demostraron que la automonitorización es un procedimiento eficaz para mejorar el rendimiento académico y las habilidades atencionales. Los estudiantes diagnosticados con déficit de atención e hiperactividad tomaron datos sobre sus tareas de matemáticas, comprensión lectora y expresión escrita y autoevaluaron sus resultados con una plantilla con puntuaciones estándar. Este procedimiento se trata de una prerrequisita para la autoevaluación; siguiente paso del procedimiento de autogestión.

3.8.2 Autoevaluación

El estudiante juzga con precisión su rendimiento y lo compara con algún estándar o regla (Clearly y Zimmerman, 2001). Los profesores pueden pedir a los estudiantes que autocorrijan sus tareas ofreciéndoles una hoja con las soluciones teniendo acceso a una retroalimentación inmediata. Se ha demostrado que la inmediatez mejora el rendimiento académico. Por ejemplo, un estudio de Morton et al. (1998) utilizó el procedimiento de autoevaluación para reducir errores ortográficos en dictados de palabras. Los resultados demostraron que la autoevaluación inmediata, después de escribir una palabra era más eficaz para la adquisición de palabras nuevas. En cambio, los estudiantes que autoevaluaban su tarea, después de escribir diez palabras, se enlentecía la adquisición de nuevas palabras. Además, este procedimiento es útil en estudiantes de secundaria dónde la materia es más compleja. Sweeney et al. (1993) enseñaron a estudiantes a autoevaluar la legibilidad teniendo en cuenta las características de las grafías (inclinación, tamaño, presencia de rasgos característicos, etc.) La incorporación del procedimiento mejoró significativamente la

escritura y el mantenimiento de esta habilidad. Una forma para promover la autoevaluación es utilizar la estrategia de identificación de conductas objetivo.

3.8.3 Identificación de objetivos

La identificación de objetivos es una estrategia conductual que se utiliza para proporcionar información sobre un conjunto de conductas objetivo; previamente autoseleccionadas, que funcionan como estructura para controlar el progreso del estudiante hacia el cumplimiento del criterio final (Hirsch et al., 2013). Esta práctica permite al estudiante que identifique objetivos adecuados, supervise su propio progreso y fomente un aprendizaje independiente. La participación en seleccionar objetivos promueve habilidades de autogestión (Copeland & Hughes 2002) y aumenta el rendimiento académico (Moeller, 2012). Incluso se ha demostrado que los estudiantes que fijan sus propios objetivos, obtienen mejores resultados en comparación a los estudiantes con objetivos asignados por una tercera persona (Bruhn et al., 2016). Este suceso es consistente con la literatura que evidencia que ofrecer alternativas a los estudiantes repercute en la mejora del rendimiento académico (Begeny et al., 2010).

Durante varias décadas se han llevado a cabo numerosos estudios que utilizan el esta estrategia con una amplia gama de edades y áreas académicas. Los resultados sugieren que su uso mejora la lectura (Didion & Toste, 2021), la escritura (Hansen & Wills, 2014), las matemáticas (Sides & Cuevas, 2020) y el estudio de lenguas extranjeras (Shih & Reynolds, 2018). La identificación de objetivos puede funcionar como componente adicional para la gestión del aula, siendo válida y eficaz en la reducción de conductas disruptivas (Kumm & Maggin, 2021). La identificación de objetivos, como tarea aislada, no hace que mejoren los resultados académicos. Más bien, la especificidad de los objetivos al definirlos de forma clara, explícita y completa aumenta la probabilidad de mejorar los hábitos de estudio. Estrapala y Reed (2019) propusieron un conjunto de directrices para mejorar el rendimiento académico en alumnos de secundaria. Durante todo el proceso el educador mantiene un rol de guía y utilizaba un registro de observación para garantizar la fidelidad de los pasos que llevaba a cabo el estudiante. Los componentes básicos eran: (a) Recopilación de datos. Este primer paso el estudiante reúne información escrita sobre tareas, calificaciones del curso, exámenes, expedientes académicos, registros de asistencia, etc. El objetivo es obtener evidencia de sus puntos fuertes y sus déficits. (b) Establecer un plan de acción. El docente en este paso guía al estudiante para que identifique el nivel actual de rendimiento que presenta. (c) Identificar los principales cambios conductuales a realizar. Mediante las ayudas

del docente el estudiante ha de identificar independientemente y priorizar qué cambios han de realizar para abordar sus necesidades. (d) Redacción de objetivos mediables y alcanzables. Este componente es esencial para que los estudiantes comprendan que definir cuantitativamente los objetivos les permitirán mejorar su rendimiento gradualmente. Para ello, el docente ha de proporcionar ejemplos de definiciones operacionales y enfatizar en la precisión del redactado. (e) Monitorizar el progreso y decidir cómo avanzar. El último componente subraya la importancia de tomar datos y monitorizar el progreso para tomar decisiones futuras sobre las conductas.

Podemos observar que a lo largo del entrenamiento de la identificación de objetivos el modelo es un procedimiento clave para mostrar cómo llevar a cabo los procedimientos. Locke y Latham, (1990) constataron que las personas tienden a prestar atención a los modelos cuando creen que la adquisición de las conductas modeladas los puede llevar a alcanzar sus metas. Actualmente, el modelado se utiliza como procedimiento para demostrar la ejecución de los pasos a seguir, pero todavía no existe literatura que haya utilizado el videomodelado para moldear conductas requisitas en el procedimiento de autogestión.

3.8.4 Autorreforzamiento

Tanto la automonitorización como la autoevaluación son habilidades prerequisites para enseñar a los estudiantes a administrarse consecuencias por sí mismos contingentemente después de emitir una conducta objetivo. Este fenómeno recibe el nombre de autorreforzamiento (Southall & Gast, 2011). Se ha demostrado que los reforzadores autoseleccionados y administrados por la propia persona que los va a consumir pueden ser más eficaces que aquellos que son seleccionados por agentes externos (Hayes et al., 1985). Para ello, para que los estudiantes se familiaricen con la mecánica de gestión de consecuencias, el educador debería de encargarse de administrar el reforzador de forma contingente tras la realización de la conducta objetivo. Después de varios ensayos con éxito, se puede proceder a desvanecer la figura del educador para que sea el propio estudiante que gestione el autorreforzamiento (King-Sears, 2008). El objetivo final de hacer partícipe a los estudiantes del proceso de autogestión es fomentar el uso de los procedimientos de forma independiente para gestionar sus propias conductas y alcanzar cualquier meta propuesta.

3.9 Instrucción sobre equivalencia

Murray Sidman y sus colegas introdujeron la tecnología de la instrucción basada en la equivalencia²⁷ introduciendo este fenómeno en la enseñanza de comprensión lectora en estudiantes con discapacidad intelectual y síndrome de Down (Sidman, 2009). Fue a partir de este momento que se empezó a estudiar esta tecnología en la enseñanza académica. Gran parte de la investigación se ha llevado a cabo en contextos altamente controlados (Mace & Critchfield, 2010). Actualmente, se utiliza para diseñar aprendizajes generativos, es decir, para la creación de nuevos repertorios conductuales complejos a partir de contingencias preexistentes, sin necesidad de un entrenamiento explícito. Este hecho es posible debido a que las nuevas contingencias creadas a partir de interconexiones previas, comparten componentes que son comunes con las contingencias originales (Sidman, 1971). La instrucción basada en la equivalencia incorpora los principios de equivalencia, es decir, los estímulos que no comparten similitudes topográficas entre sí, pero forman parte de una misma clase de estímulo, significa que forman una clase de equivalencia (Sidman & Tailby, 1982). Para demostrarlo, todos los estímulos que forman parte de esta clase de equivalencia han de tener propiedades reflexivas ($A=A$), de simetría (si $A=B$, entonces $B=A$) y de transitividad (Si $A=B$ y $B=C$). Además, este marco instructivo utiliza el entrenamiento de emparejamiento a la muestra y el entrenamiento de discriminaciones condicionales, como procedimientos conductuales para establecer clases de equivalencia. La equivalencia de estímulos ha sido eficaz en la enseñanza de la ortografía (De Rose et al., 1996) y las matemáticas (Dixon, 2016). Además, existe una amplia literatura que demuestra su utilidad en un número de personas con diferentes habilidades. Por ejemplo, es útil en la enseñanza de personas con síndrome de Down (Grisante, 2014), el síndrome X frágil (Hall et al., 2006), discapacidad intelectual severa (Sidman & Cresson, 1973) y para el aprendizaje en personas con trastorno del espectro autista (de Melo et al., 2020) y lesiones cerebrales (Fienup, 2010).

En los últimos años varios estudios se han centrado en investigar el uso de la instrucción basada en la equivalencia de estímulos aplicada en la enseñanza universitaria. En primer lugar, se ha demostrado la adaptabilidad y el potencial de la tecnología, independientemente del canal de enseñanza. Por ejemplo, Walker y Rehfeldt (2012) exploraron la eficacia de la instrucción presentando el protocolo de enseñanza y de sondeo a través de la web, en concreto con Blackboard®. Esta plataforma permitía a los profesores y estudiantes de máster participar en actividades sín-

27 *Equivalence-based instruction (EBI)*, en la literatura original.

cronas y asíncronas y evaluar las respuestas intraverbales. Sella et al. (2014) evaluó si un protocolo de entrenamiento de enseñanza de tareas de emparejamiento a la muestra emergía nuevas respuestas escritas basadas en la topografía presentando todas las tareas, tanto los sondeos como los ensayos de entrenamiento a través de Adobe Connect®. Recientemente, Albright et al. (2015, 2016) utilizó la instrucción basada en la equivalencia mediante tutoriales informáticos automatizados (softwares) para enseñar estadística y la interpretación de los resultados gráficos de un análisis funcional a estudiantes universitarios de psicología. Los resultados sugirieron la formación de clases de equivalencia, un aumento de las puntuaciones entre las fases de línea base y post tratamiento y la generalización y mantenimiento de las habilidades complejas adquiridas. Brodsky y Fienup (2018) publicaron una revisión sistemática y un metaanálisis de los 28 últimos estudios publicados hasta agosto de 2016 relacionados con la enseñanza universitaria y la equivalencia de estímulos. Los resultados indicaron que existe una base sólida de evidencia que demuestra la eficacia de la instrucción basada en la equivalencia en estudiantes universitarios. Los hallazgos argumentaban un incremento en las respuestas de participación de los estudiantes y la adquisición de habilidades complejas. En cuanto a la integración de las nuevas tecnologías como canal de enseñanza, la literatura propone la falta de formación de maestros sobre el diseño de softwares para hacer posible que la instrucción basada en la equivalencia se convierta en una corriente principal de enseñanza.

3.10 Retroalimentación

El término de retroalimentación²⁸ se utiliza con mucha frecuencia en el contexto de la educación, sin embargo, no existe un consenso para su definición (Houmanfar, 2013). En entornos educativos la retroalimentación consiste en transmitir información sobre el desempeño de una persona en comparación a unos objetivos conductuales o criterios previamente establecidos. En la literatura podemos encontrar definiciones enfatizando la cantidad y calidad de la información proporcionada (Prue y Fairbank, 1981), la descripción precisa de qué y cómo se ha realizado el desempeño y los ajustes para mejorar su rendimiento (Duncan y Bruwelheide, 1985).

A pesar de ser una intervención popularmente aplicada en el contexto del análisis de conducta, existe un desacuerdo en cuánto a los principios conductuales encargados del cambio conductual. La retroalimentación podría funcionar como: (a) una instrucción (Hirst et al., 2013),

28 Feedback, en la literatura original.

(b) un estímulo discriminativo (Peterson, 1982), (c) una ayuda (Salmoni et al., 1984), (d) una operación de establecimiento (Agnew, 1998), (e) una regla (Haas & Hayes., 2006); (f) un reforzador o estímulo punitivo (Carpenter & Vul, 2011) dependiendo del contexto donde se utilice (Alvero et al., 2001). Por ejemplo, los resultados del estudio de Mangiapanello y Hemmes (2015) afirman que si después de recibir retroalimentación, el receptor genera reglas a partir de la información proporcionada y provoca una modificación en la conducta, conduce a que la retroalimentación funciona como una regla. En cambio, si la retroalimentación evoca inmediatamente una clase de respuestas relacionadas con la información de la interacción, la retroalimentación funciona como estímulo discriminativo. Actualmente, ninguna función ha sido probada científicamente que corrobore cual es la más eficaz.

Varios autores asumen que la retroalimentación funciona como reforzador porque existe un solapamiento entre los procedimientos operantes y la retroalimentación. En ambos procesos la consecuencia es contingente e inmediata a las propiedades de la respuesta (Weatherly & Malott, 2008). Este hecho explica que la retroalimentación inmediata mejora la adquisición de aprendizaje en comparación a una retroalimentación demorada. Por ejemplo, el metaanálisis realizado por Kulik y Kulik (1988) demostró que los estudios llevados a cabo en aulas, la retroalimentación inmediata provocaba mejores resultados que la demorada. En cambio, los estudios que simulaban situaciones de aprendizaje en entornos experimentales, no había diferencia de resultados entre ambos tipos de retroalimentación.

La retroalimentación se considera un componente crucial para la enseñanza y el aprendizaje, es por ello, que se incorpora como estrategia en la educación (Noell et al., 2014). Además, suele incorporarse en áreas como (a) el entrenamiento en habilidades conductuales, (b) la auto-gestión, (d) la tutoría con iguales y (c) la gestión de la conducta organizativa. Este último ámbito se dedica a aumentar la productividad, reducir los malos hábitos y el absentismo.

En la literatura la retroalimentación puede aplicarse como una estrategia unitaria, es decir, los estudiantes tan solo reciben información sobre su desempeño; o bien en combinación con otros procedimientos. Por ejemplo (a) el uso de retroalimentación con ayudas o reglas (componentes antecedentes), (b) la retroalimentación en conjunto con consecuencias conductuales (reforzadores tangibles o de actividades), (c) la retroalimentación y el uso de identificar objetivos

y (d) la retroalimentación como componentes de un paquete de intervención formado por estrategias antecedentes, establecimientos de objetivos e intervenciones sobre las consecuencias.

La entrega de la retroalimentación puede variar en función de una serie de dimensiones identificadas por Balcazar et al. (1985). Las características que pueden variar la eficacia de la retroalimentación son: (a) la fuente de retroalimentación (supervisores, investigadores, generada automáticamente por un dispositivo electrónico, clientes, etc.), (b) el canal de comunicación, es decir, mediante lenguaje verbal, escrito, visual (gráficas) o el uso de videos, (c) los participantes que reciben la retroalimentación (individual, grupal o en combinación), (d) la privacidad de la retroalimentación, (e) el contenido, es decir, según el rendimiento comparándolo o bien con el rendimiento del mismo estudiante, de otro o mediante un criterio estándar, (f) la frecuencia de la retroalimentación, ya sea a diario, de forma semanal, mensual o trimestral, etc. La inmediatez de la retroalimentación, es decir, el tiempo que transcurre entre la recepción de la retroalimentación y el desempeño influye en la eficacia de la estrategia, tal y como hemos mencionado anteriormente.

La literatura ha focalizado el interés en la retroalimentación individual y en privado, adoptando el rol de corregir el desempeño de una persona. La retroalimentación correctiva tiene el propósito de modificar o reducir una clase de respuesta específica (Alvero et al., 2001). La retroalimentación puede funcionar como un estímulo que provoque eventos privados aversivos en la persona receptora. Es por este motivo que la literatura reciente se ha dedicado a moldear las habilidades fundamentales para recibir retroalimentación. Por ejemplo, Eldrich, (2020) ha utilizado los procedimientos de entrenamiento en habilidades conductuales (EHC) para reducir cualquier respuesta aversiva.

La retroalimentación es un procedimiento fácil de implementar y eficaz para provocar una modificación en el desempeño, independientemente de la población y del contexto. En general, la investigación se centra en la retroalimentación entregada verbalmente, a pesar de existir la posibilidad de mecanizar el procedimiento. Por ejemplo, Geiger et al. (2018) comparó el coste de tiempo total llevando a cabo dos formaciones idénticas, incluyendo la retroalimentación como parte de la formación y variando el canal de transmisión. Un grupo fue inmerso a un entrenamiento de habilidades conductuales (EHC) y el otro recibió los módulos a través de un ordenador. Los participantes que recibieron la retroalimentación mediante el EHC obtuvieron una mejora de tan solo el 4%, en comparación a aquellos que obtuvieron la retroalimentación mediante un dis-

positivo electrónico. Por lo tanto, puede resultar beneficioso la formación mediante ordenadores porque es más eficiente en tiempo; ya que no precisa tiempo del personal.

4. EDUCACIÓN CONDUCTUAL FUNDADA EN LA EVIDENCIA: APLICACIONES EDUCATIVAS INTEGRALES

4.1 Enseñanza de Precisión

La enseñanza de precisión²⁹ se basa en los conocimientos de condicionamiento operante libre y el análisis de las tasas de respuesta que Odgen Lindsley adaptó durante su época de estudiante de Skinner en 1960 (Lindsley, 1992). Se trata de un método de medición conductual que permite determinar la eficacia de las estrategias de enseñanza (West et al., 1990). Por lo tanto, no se considera un sistema de enseñanza tal y como su nombre indica, sino un método que evalúa la eficacia de las técnicas de instrucción que se llevan a cabo en el aula. Por este motivo, puede complementarse con otros procedimientos o métodos de enseñanza (Kubina, 2021). El uso de este sistema facilita a los educadores la toma de decisiones basándose en la competencia de los estudiantes. La enseñanza de precisión se rige por cuatro principios fundamentales.

4.1.1 El estudiante siempre tiene la razón

El primero y fundamental, es que el estudiante siempre tiene la razón. Esto significa que las respuestas del estudiante son el mejor indicador para evaluar si las estrategias de instrucción contribuyen a mejoras observables. Este enfoque comienza evaluando las habilidades que presenta cada alumno en el repertorio antes de intervenir y después de practicar para luego obtener medidas sobre los resultados de aprendizaje. El estudiante practica tanto como necesita y progresa a medida que se le ofrece retroalimentación sobre su rendimiento. Este enfoque nos recuerda que el aprendizaje es individual, sea cual sea el rendimiento del estudiante, la responsabilidad de progreso recae en el educador.

4.1.2 Uso de medidas directas

El segundo principio, relacionado con el anterior, enfatiza en tomar medidas directas y frecuentes de conductas observables a lo largo del tiempo. La selección de las conductas depende de los objetivos preestablecidos por el docente, seleccionando respuestas contables y repetibles a lo

29 *Precisión Teaching*, en la literatura original.

largo del día. Una de las características de este método es evaluar el rendimiento de las habilidades clasificadas previamente en orden lógico, desde los componentes más básicos hasta las habilidades más complejas. Normalmente, se acostumbra a medir respuestas académicas como: el deletreo, la lectura de palabras u oraciones, la emisión de conceptos y sus definiciones o bien la enumeración de pasos necesarios para resolver problemas matemáticos, ya sean oralmente o textualmente o cualquier otra respuesta que requiera de fluidez.

4.1.3 Registrar la fluidez

Para que sea lo más eficaz posible a efectos de obtener información, los profesionales han de tomar medidas. El tercer principio de la enseñanza de precisión es el uso de la unidad de medida estándar, la tasa de respuesta. Esta dimensión de la conducta consiste en hacer un recuento o tomar datos sobre la frecuencia de la respuesta objetivo dividiéndolo por el tiempo, siendo un minuto la medida por excelencia. No obstante, el tiempo puede variar en función de la complejidad de la tarea.

Los docentes o los propios estudiantes toman datos sobre su rendimiento varias veces al día o después de un periodo determinado de enseñanza. Por ejemplo, después de un periodo determinado practicando una habilidad, el educador toma datos mediante “sondeos” o ensayos breves en el que el estudiante ha de emitir una respuesta el mayor número de ocasiones. El rendimiento más elevado a lo largo del día será el que quede graficado. El aspecto más importante sobre el uso de medidas de sondeos se trata de un método eficaz para obtener información sobre la tasa de respuestas correctas e incorrectas y permite conocer el nivel de ejecución del estudiante y el grado de fluidez (Binder 1996; Johnson & Street 2012). Por lo general, el sistema educativo utiliza el porcentaje de respuestas correctas como medida estándar, obteniendo meramente la medida de precisión. Por ejemplo, dos estudiantes pueden obtener un 100% al finalizar la tarea, pero uno de ellos ha tardado 15 minutos y el compañero 30 minutos. Utilizando la medida de la tasa y no el porcentaje de respuesta, obtenemos información sobre la fluidez y podemos concluir que el primero ha logrado un dominio fluido de la tarea, mientras que el último, simplemente ha ejecutado un rendimiento preciso.

El desarrollo de la fluidez se caracteriza por obtener tanto la precisión como la velocidad apropiada en la realización de tareas. El grado de fluidez al que una tarea debe realizarse, se centra predominantemente en la velocidad o el ritmo de respuesta (Binder & Watkins 2013). Para ello, el educador tiene que hacer uso de: (1) datos normalizados, según la edad de los alumnos y

la tarea, (2) el nivel de rendimiento del grupo, como dato de referencia, (3) el nivel adecuado para que la habilidad se considere funcional en el contexto natural y (4) el nivel de ejecución del propio profesional. Por ejemplo, el educador podría basar un objetivo de rendimiento para estudiantes de posgrado matriculados en un curso de Introducción al análisis de conducta. Si el instructor es capaz de nombrar 50 términos relacionados con el tema explicado, durante un minuto, éste sería un objetivo de rendimiento adecuado para los alumnos. La continua adquisición de habilidades objetivo sienta las bases para alcanzar la fluidez en los componentes necesarios para mostrar competencia en habilidades más complejas. Si el ritmo de aprendizaje es lento, se administran muchas oportunidades de enseñanza hasta que el estudiante domine una secuencia de componentes hasta desarrollar una habilidad concreta.

4.1.4 Uso del gráfico de aceleración estándar

Para que la recogida de datos sea informativa y eficaz, es importante graficar. El último principio importante de la enseñanza de precisión es el uso de gráficos de aceleración estándar (GAE) como herramienta visual que muestra los cambios de rendimiento y permite que el profesional tome decisiones respecto a sus estrategias de enseñanza (Johnson y Street, 2012). El gráfico de aceleración estándar contiene una escala de tiempo en el eje de la X que puede acomodar 140 días naturales (días de un curso escolar) en base a una escala multiplicadora o logarítmica en el eje de la Y para la frecuencia de las respuestas (1,10,100,1000, etc.) (Calkin, 2005). El gráfico está compuesto por seis ciclos, motivo que explica que a veces se denomine “gráficos de seis ciclos”. Son útiles para medir la aceleración; la rapidez con la que el estudiante responde en ensayos consecutivos o la desaceleración cuando la respuesta a ensayos consecutivos enlentece. La evidencia sugiere que la aceleración es una medida directa de la adquisición y de la fluidez del aprendizaje (Cooper, 2020). Del mismo modo, se puede observar una aceleración de las respuestas correctas y mantener el mismo número de errores. Este dato no se obtendría en caso de utilizar medidas de porcentaje. La visualización de los cambios de respuestas de rendimiento a lo largo del tiempo hace que sea un gráfico rápido y fácil de analizar (Lindsley, 1992). A pesar de la elaboración sofisticada del registro, no es necesario emplear el mismo nivel de matemáticas exactas para tomar decisiones de enseñanza según los datos recogidos. El docente traza una línea uniendo los puntos o las “X” con el fin de observar la pendiente de aprendizaje. Si la pendiente tiende a ascender, lo más probable que se mantenga la misma estrategia de instrucción. Por el contrario, si presenta una línea plana, una tendencia descendiente o en dirección equivo-

cada, sería necesario cambiar la estrategia de enseñanza actual o las contingencias presentes durante la instrucción (Carl Hughes et al., 2007). Se han desarrollado modelos de aprendizaje³⁰ que indican la complejidad de la tarea (muy fácil o muy difícil), el nivel de instrucción que requiere el alumno (más o menos ensayos), la necesidad de progresar o retroceder a otras habilidades y la eficacia de los métodos de enseñanza aplicados. Se considera una deficiencia en la instrucción y no un déficit del estudiante cuando no se alcanza el índice de celeridad esperado (Chiesa & Robertson, 2000). En resumen, son una guía excelente para tomar decisiones con rapidez sobre la efectividad de la enseñanza.

4.1.5 Tutorías

A pesar de que la instrucción precisa se considere, principalmente, un método evaluativo, incorpora principios basados en la enseñanza eficaz, como la evaluación de los estudiantes mediante tutorías diarias. Los tutores pueden ser compañeros de clase o el propio educador. Su función es registrar la tasa de respuestas correctas durante un minuto, aproximadamente, en áreas que todavía no ha adquirido la fluidez. En caso de ser necesario, ejecutan la tarea, presentando el modelo de forma precisa y competente y proporcionan retroalimentación inmediata según las respuestas del estudiante, pautando el ritmo que ha de seguir para el siguiente sondeo. Esta evaluación permite a los educadores juzgar objetivamente el progreso de los estudiantes, optimizar su rendimiento y descartar planes de intervención ineficaces para la siguiente semana.

4.1.6 Aplicaciones prácticas

Una de las aplicaciones a gran escala de la enseñanza de precisión tuvo lugar en la escuela de Sacajawea de Great Falls, Montana, en 1975. Tres escuelas de primaria incorporaron los principios de enseñanza de precisión, el uso diario de gráficas de aceleración estándar y la evaluación continua (condición experimental). Los datos de los estudiantes los compararon con otras tres escuelas de primaria que no habían llegado a cabo este método (condición control). Después de tres años, los estudiantes de las escuelas experimentales obtuvieron entre 20 y 40 puntos de percentil más en las pruebas de rendimiento de habilidades básicas entre estudiantes de aulas de educación ordinaria como especial, en comparación con las escuelas de la condición control (Beck & Clement, 1991). Otro ejemplo de la eficacia de este enfoque, es el Modelo Morningside de Instrucción Generativa (Johnson & Street, 2012). Se trata de una escuela de primaria y secun-

30 *Learning Pictures*, en la literatura original.

daria situada en Seattle que utiliza los principios de instrucción de precisión para garantizar el dominio de las habilidades de lectura, lengua, arte y matemáticas. Los estudiantes que atienden a estas escuelas suelen ser etiquetados con “problemas de aprendizaje” o “déficit de atención e hiperactividad”. Los resultados de las investigaciones que han llevado a cabo este proyecto revelan un crecimiento medio de dos cursos por cada año escolar. El Modelo Morningside de Instrucción Generativa es un método derivado de la instrucción de precisión eficiente en relación al coste y beneficio de tiempo y aprendizaje.

Este enfoque se ha aplicado con éxito con una diversidad de estudiantes y áreas educativas debido a los indicadores de calidad que posee como enseñanza exitosa: (a) adaptable en función de las habilidades de los estudiantes, (b) múltiples oportunidades para practicar con tiempo limitado, (c) retroalimentación inmediata, (d) autorregistro y (e) acceso a reforzamiento (Kleinert et al., 2018). Se trata de un enfoque eficaz para enseñar a personas de desarrollo típico habilidades de lectoescritura (Brosnan et al., 2018), matemáticas (Sleeman et al., 2021; McTiernan et al., 2018) y habilidades verbales vocal (tactos e intraverbales) (Cihon et al., 2017). También, se ha demostrado su eficiencia con personas con discapacidad para enseñar a sumar (Vostanis et al., 2021) y a hablar (Aravamudhan & Awasthi, 2021).

4.2 CABAS

El análisis aplicado de conducta se ha centrado en desarrollar una tecnología que permita medir objetivamente y proporcionar una enseñanza eficaz. Un modelo que aplica estrategias conductuales de forma integral en todo el sistema educativo, incluyendo estudiantes con discapacidad, es la aplicación integral del análisis de la conducta a la enseñanza escolar³¹ (CABAS, son sus siglas). El modelo CABAS fue desarrollado por R. Douglas Greer (1989) y sus colegas en la Universidad de Columbia. Sus orígenes se remontan a la década de 1980, con la incorporación de la Ley de Educación para Personas con Discapacidad (IDEA, las siglas en inglés). Greer y sus colegas se esforzaron en crear un sistema educativo dirigido a un grupo de estudiantes diagnosticados con discapacidad del desarrollo, trastornos del espectro autista y alumnos con conductas autolesivas y agresivas. En este entorno trataron de averiguar si un enfoque basado en la ciencia, focalizado en el estudiante era posible, eficaz y práctico. Para el desarrollo de las escuelas CABAS se incorporaron fundamentos del análisis de aplicado de la conducta y de algunos métodos de enseñan-

31 Comprehensive Application of Behavior Analysis to Schooling, en la literatura original.

za eficaz. Se basaron en la Instrucción Directa (Englemann & Carnine, 1982), además del sistema personalizado de instrucción (Keller, 1968) de dónde extrajeron la importancia de la formación de los profesores y padres, también se influenciaron de la Instrucción Programada (Skinner, 1984) dónde contemplaron la posibilidad de utilizar ordenadores como canal de enseñanza. Además, se beneficiaron de la Enseñanza de Precisión enfatizando la fluidez y la precisión de los datos (Lindsley, 1990). Por otro lado, aprovecharon las contribuciones que ofrecen los modelos aplicados como: la equivalencia de estímulos, el análisis de la conducta verbal y la teoría del marco relacional (Greer & McDonough, 1999). En efecto, desarrollaron y aplicaron planes de intervención y fueron conscientes de la importancia de apoyar sus prácticas con formación y supervisión continua a todas las personas que convivían con los estudiantes en el día a día (Selinske et al., 1991).

CABAS es un enfoque dinámico e integral de enseñanza basado en datos y en la aplicación de prácticas educativas con evidencia científica, en el que el rendimiento de todos los miembros que componen una escuela puede medirse, analizarse y ser modificados en beneficio al estudiante (Greer, 1997). El término “escuela” en este modelo, engloba a cualquier lugar físico que proporcione servicios educativos (p.ej., una clínica, un aula, un centro de estudios, un hogar), lo mismo sucede con el término “maestro”, se refiere a cualquier persona que tiene el rol de impartir conocimientos, ya sea un terapeuta, un clínico, un analista de conducta, etc.

Este modelo aplica el análisis de la conducta a todos los miembros que forman la escuela: padres, profesores, supervisores y mentores. Todos son responsables de las necesidades individuales y del éxito de los alumnos (Singer-Dudek et al., 2010). Este enfoque sitúa la figura del estudiante en el centro de la matriz; siendo el programa de formación y el resto de miembros quienes se han de adaptar a las necesidades de aprendizaje de los estudiantes y a no preocuparse por intereses externos (p.ej., dinero, recursos, días festivos, etc.).

4.2.1 Componentes de CABAS

El modelo educativo de CABAS está definido por una matriz organizativa necesaria para ser completa y autocorregible. Todos los siguientes componentes son necesarios para que funcione el sistema.

- **Los estudiantes:** El núcleo de la matriz es el estudiante y el resto de componentes giran en torno a él. Las escuelas atienden a niños desde preescolar hasta la primaria. Se ofrecen aulas para estudiantes de desarrollo típico y con necesidades de educación

especial. Los estudiantes son agrupados según sus habilidades verbales y su rendimiento académico, existiendo la posibilidad de permanecer en un aula inclusiva o de recibir enseñanza individual (1:1). Se asignan roles específicos a diversos roles de enseñanza como presentamos de forma abreviada a continuación.

- *Los padres.* A los padres se les involucra en la educación de sus hijos, formándoles sobre la importancia de utilizar al unísono las mismas estrategias conductuales que emplean los maestros aplicándolas en el hogar.
- *El maestro.* Los maestros funcionan como científicos aplicados siendo responsables de proporcionar a los estudiantes las unidades de aprendizaje con precisión y sin errores. Deben formarse y a superar módulos para ir progresando. La filosofía de este enfoque afirma que los estudiantes aprenden tan rápido como los maestros pueden enseñarles. Greer (1997) afirmó que cuántos más conocimientos tenga el maestro, mayores resultados obtendrán los estudiantes.
- *El supervisor.* Son maestros que han superado 30 módulos de tres niveles de formación. Mediante observaciones sistemáticas, supervisan y entregan formación a los maestros en el aula. Además, comprueban que cumplen con los criterios programados semanalmente y revisan que estén tomando las decisiones adecuadas en relación a sus alumnos. También, están encargados de supervisar el rendimiento de los maestros e identificar cualquier problema y buscar solución.
- *El mentor.* Son profesionales que dominan los principios de la ciencia básica y del análisis aplicado de conducta. Normalmente, son profesores universitarios y/o investigadores que aportan los avances de vanguardia de procedimiento conductuales. Su rol consiste en analizar continuamente todos los componentes del programa; ya que su responsabilidad recae en el funcionamiento integral del sistema y de cada una de sus partes.

4.2.2 Formación y supervisión

El análisis de la conducta verbal expuesto por Skinner (1957) es la base de la formación y la supervisión de maestros y supervisores en el sistema CABAS. Este programa utiliza un sistema de medición intensivo; ya que se toman datos de forma continua de todo el proceso de formación. Los profesores deben alcanzar criterios de formación programados por sus mentores. El contenido de la formación de los docentes está compuesto por tres módulos conformado bajo tres

epígrafes: (a) el análisis de la conducta verbal de Skinner, (b) los procedimientos de enseñanza, (c) la toma de decisiones según los datos. Cada módulo integra una serie de competencias gradualmente ordenadas de menor a mayor complejidad. A cada profesor se le asignan un número de competencias o módulos en función de sus habilidades observadas en el aula por el supervisor. La finalización de los módulos y obtener un mérito depende de la adquisición de criterios de dominio y alcanzar un nivel de fluidez. Este sistema de formación y supervisión refleja el número total de unidades logradas, la tasa de rendimiento alcanzada del profesor y la precisión al tomar decisiones fundamentados en datos.

La formación de padres es otra característica importante de este modelo. El objetivo es ofrecer formación individual sobre la entrega de unidades de aprendizaje en casa y enseñarles estrategias conductuales que aumenten las habilidades de lenguaje espontáneo y de interacción social (Greer, 1994).

4.2.3 Modelos de medición continua

El sistema CABAS está compuesto por currículos basados en los hitos del desarrollo verbal ordenados de forma lógica. Los maestros, padres y el resto de partes implicadas elaboran conjuntamente por escrito los planes de intervención individualizados. Cada plan está guionizado por objetivos de aprendizaje, a corto y largo plazo y por sus criterios de adquisición. Para obtener datos de los objetivos de aprendizaje desarrollaron una unidad de medida, denominada, unidad de aprendizaje (Greer & McDonough, 1999). Está compuesta por una contingencia de cuatro términos, en primer lugar, se tiene en cuenta las operaciones motivadoras y la historia ontogenética que podría estar afectando al estudiante, el siguiente estímulo es la aparición de una señal que indica la disponibilidad de responder (Greer & McDonough, 1999), a continuación, el estudiante responde a dicho estímulo dentro de los tres próximos segundos; siendo ésta la medida que se registra. El último elemento de la contingencia, es la respuesta del maestro proporcionando una retroalimentación correctiva o bien, de reforzamiento (Ingham & Greer, 1992). Las unidades de aprendizaje se presentan en el contexto de instrucción y se entrelazan con las contingencias del maestro, implicando múltiples operantes a analizar (Skinner, 1986). Cumplen con cuatro funciones (a) medir directamente el rendimiento del estudiante, (b) facilitar que el maestro supervise el progreso y realice los ajustes necesarios, (c) proporcionar información para resolver las dificultades en el aprendizaje, (d) refinar los procedimientos que ponen en práctica en el aula (Greer, 2002). Por lo tanto, se consideran un índice de aprendizaje que permite medir la eficacia de la

enseñanza del educador y el aprendizaje del estudiante (Keohane et al., 2008). Se ha demostrado que mejoran el rendimiento académico (Bahadourian et al., 2006), incrementan el número de respuestas correctas y, por lo tanto, aumenta la adquisición de objetivos programados en el plan de intervención diseñado (Selinske et al., 1991).

Los datos de las Unidades de Aprendizaje recopilados, se grafican y se analizan a diario para tomar decisiones precisas de acuerdo al denominado Protocolo de Decisión (Keohane & Greer, 2005). Se trata de un algoritmo analítico que lleva a los maestros a formularse preguntas estratégicas sobre el análisis visual de los datos. Este consiste, primero, en realizar un análisis visual de los datos e identificar si es necesario tomar una decisión al respecto. El segundo y último paso se trata de responder una serie de preguntas estratégicas relacionadas con estrategias gobernadas por reglas que facilitan la toma de decisiones. El uso del Protocolo de Decisión se introdujo en el modelo CABAS para esbozar estrategias de resolución de problemas y resolver las dificultades de aprendizaje de los estudiantes.

La literatura ha demostrado que los profesores aumentan su precisión con la aplicación de la Escala de Precisión y Rendimiento del Profesor³² y por consecuencia aumenta el número de respuestas correctas de los estudiantes (Ingham & Greer, 1992). Los supervisores evalúan el desempeño de los profesores in situ. Antes de observar cómo el profesor ejecuta la sesión, los mentores revisan los datos graficados de la sesión anterior, la selección de objetivos a intervenir, la preparación de los materiales necesarios y ofrece retroalimentación a modo de sugerencias. Una vez iniciada la sesión en el aula, el mentor observa y registra la tasa de unidades de aprendizaje presentadas por el profesor y la respuesta de los alumnos. Por cada unidad entregada, los supervisores proporcionan retroalimentación in situ sobre la fluidez de instrucción, la adquisición de procedimientos y la precisión en la recogida de datos (Greer, 1994).

A lo largo de los últimos años, el modelo CABAS se ha puesto en práctica en escuela que atienden a alumnos con discapacidad intelectual, visual, problemas de aprendizaje y trastornos emocionales. Al tratarse de un enfoque sistémico, las características del modelo han ido evolucionando y adaptándose en función de cada escuela.

32 *Teacher Performance Rate and Accuracy Scale (TPRA)*, en la literatura original.

4.3 Instrucción Directa

La instrucción directa se ha aplicado en diversas áreas académicas: lenguaje, lectura, matemáticas, escritura. Existen investigaciones publicadas con personas de desarrollo neurotípico y con poblaciones de educación especial. Específicamente con estudiantes que presentan problemas del lenguaje, de la comunicación y conductas disruptivas, incluso, personas con trastorno del espectro autista y lesiones cerebrales traumáticas. Este enfoque puede satisfacer las necesidades de todo tipo de estudiante utilizando un currículo ya completo con todos los programas, contenidos e instrucciones específicas para el profesor. Probablemente, el programa más conocido se conoce con las siglas en inglés DISTAR³³ que se adapta a las áreas del lenguaje, lectura y matemáticas.

4.3.1 Lenguaje

La adquisición del lenguaje es un eslabón importante para el desarrollo de la comunicación. Los estudios han demostrado una correlación entre el retraso del lenguaje (24 y 31 meses) y deficiencias en la lectura y el rendimiento académico (Rescorla, 2002). También, existe una conexión directa entre el desarrollo del lenguaje oral y la consciencia fonológica (Metsala, 1999). La enseñanza que se centra en el lenguaje oral, favorece el desarrollo de la conciencia fonológica, de la lectura, del vocabulario y de la adquisición de estructuras gramaticales, siendo estas habilidades imprescindibles para el éxito académico (Biemiller, 2006). Otro aspecto esencial para el desarrollo del niño, es la comprensión. Existen numerosos estudios que demuestran que el lenguaje oral está vinculado con la comprensión lectora (Brett et al.,1996).

La instrucción directa ha desarrollado una serie de currículos que favorecen las habilidades del lenguaje oral. Para que sean eficaces, los estudiantes han de tener una serie de habilidades prerequisites: imitación, responder a oraciones sencillas, señalar, seguir órdenes y nombrar objetos. El primer nivel, DISTAR I o Lenguaje para Aprender³⁴ está compuesto por 150 lecciones que se centra en enseñar vocabulario, palabras descriptivas y oraciones. La enseñanza se extiende al segundo nivel, DISTAR II o Lenguaje para Pensar³⁵. Ofrece 140 lecciones centradas en la enseñanza de vocabulario y de estructuras más complejas, que aportan la enseñanza de discriminar

33 *Direct Instruction for the Teaching of Arithmetic and Reading*, en la literatura original.

34 *Language for Learning*, en la literatura original.

35 *Language for Thinking*, en la literatura original.

entre el significado preciso e inferencial del lenguaje hablado y escrito. El último nivel, DISTAR III o Lenguaje para Escribir³⁶ conocido como sienta las bases de las habilidades comunicativas. Los estudiantes que completan los tres niveles tienen las habilidades y capacidades de analizar estructuras orales y textuales. Al inicio, existen muchas oportunidades de interacción con el educador, que a medida que avanzan se desvanecen y se convierte en trabajo independiente.

4.3.2 Lectura

La lectura es una habilidad fundamentada en el lenguaje que ha sido intervenida con éxito en numerosas investigaciones de instrucción directa (Slavin et al., 1995). Se ha demostrado empíricamente que el éxito en la lectura depende de la conciencia fonológica, la fonética, la fluidez y la comprensión de textos (Hillman, 2003). Los niños que obtienen un buen resultado en tareas de conciencia fonológica, es decir, en identificar los sonidos en el lenguaje oral se convierten en lectores exitosos, en cambio, aquellos que obtienen resultados bajos, presentan dificultades para aprender a leer (Mann & Foy, 2003). En cambio, la fonética aborda estas habilidades con palabras escritas. La adquisición de la fluidez se logra mediante el uso de guiones, característicos de la instrucción directa (Stockard et al., 2018). La comprensión trata de extraer el significado del texto. Messer et al. (2004) averiguaron que los niños que presentaban dificultades en el lenguaje, también tenían problemas con la lectura y la comprensión de textos. El primer programa que abarcó todos estos componentes se conoció como DISTAR para la lectura³⁷ o Dominio de la Lectura³⁸ (Cooke et al., 2004). Más tarde, Tobin (2004) creó el programa Horizons formado por tres niveles (A, B, C-D) y comparó el rendimiento de lectura de estudiantes de primer grado, con ambos programas. Los resultados mostraron que los estudiantes que recibieron el programa Horizons obtuvieron mayor rendimiento y superaron significativamente a los estudiantes formados mediante el programa *Reading Mastery*.

4.3.3 Matemáticas

Las habilidades matemáticas son esenciales en las actividades de la vida diaria, por este motivo, es importante comprenderlas y no memorizar los procedimientos. Para facilitar el aprendizaje a

36 *Language for Writing*, en la literatura original.

37 *Reading DISTAR*, en la literatura original.

38 *Reading Mastery*, en la literatura original.

estudiantes con dificultades en esta área, se desarrolló el programa Aritmética DISTAR³⁹. Está compuesto por dos niveles, el primero se utiliza con alumnos de preescolar y el segundo con los de primer grado. Los conceptos se vinculan y se practican de una lección a otra, incrementando el nivel de complejidad de forma gradual. Los estudiantes que completen ambos niveles tendrán adquiridas las destrezas de sumar, restar, multiplicar, operar con fracciones y resolver cualquier caso práctico que involucre estas habilidades. Incluso, se creó un programa para jóvenes para desarrollar habilidades de resolución de programas. El enfoque Conectando Conceptos Matemáticos⁴⁰ estaba formado por seis niveles, con 30 lecciones por nivel planificado para aprender gradualmente e integrar los conceptos a la aplicación práctica. En esta área se ha realizado abundantes investigaciones, con tamaños de efecto muy grandes (Crawford & Snider, 2000). Los resultados obtenidos se han generalizado a estudiantes con discapacidad.

4.3.4 Escritura

La expresión escrita requiere atención a la mecánica de escribir, la ortografía y la construcción de frases, párrafos y relatos (McCutchen, 1966). Existen programas para esta área: Escritura Expresiva (Walker, 2007), Ortografía mediante Morfemas (Grossen, 2004) y el programa de Razonamiento y Escritura (Datchuk & Rodgers, 2019). Estos programas han demostrado tener éxito, por ejemplo, Viel-Ruma et al. (2010) constataron resultados favorables en el rendimiento de aquellos estudiantes que recibieron Instrucción Directa midiendo la secuencia correcta de las palabras y la longitud del texto. Son escasas las investigaciones de esta área con personas con discapacidad.

4.3.5 Investigación

La instrucción directa es un enfoque educativo que posee una amplia validación científica; ya bien por los componentes o técnicas que utiliza, como por el estudio de los programas que han derivado de sus principios o del propio enfoque como una unidad a evaluar.

Las investigaciones que están relacionadas indirectamente con la instrucción directa se basan en la eficacia y el entrenamiento de los profesores (Demant & Yates, 2003), la interacción entre alumno-profesor y la organización sistemática y gradual de los programas de instrucción. Rosenshine (1986) detalló cómo los docentes tenían que presentar los guiones, la importancia

39 DISTAR Arithmetic, en la literatura original.

40 Connecting Math Concepts, en la literatura original.

de practicar y de dar correcciones en caso de ser necesario; además de revisar semanalmente y mensualmente todos los conceptos adquiridos.

Otra fuente de investigación que recibe un respaldo empírico importante son las áreas temáticas específicas (lenguaje, lectura, escritura y comprensión). Stockard et al. (2018) examinó la eficacia de este enfoque revisando la literatura publicada a lo largo de 50 años. Se recopilaron 328 estudios abarcando una amplia gama de sujetos, entornos y enfoques metodológicos. Se estudiaron alrededor de 4000 efectos sobre el rendimiento académico en las áreas de matemáticas, lenguaje, lectura y escritura. Todos los efectos fueron estadísticamente positivos. Recientemente, Datchuk et al. (2019) ha comprobado que la instrucción directa mejora la capacidad de escritura. La incorporación de las nuevas tecnologías ha provocado que los autores utilicen la instrucción directa para enseñar nuevas habilidades. Bruno et al. (2021) utilizó este enfoque con estudiantes universitarios con discapacidad intelectual y del desarrollo a manipular un dispositivo electrónico, en concreto, un lector; dispositivo que facilita la lectura.

Por otro lado, existe literatura que incluyen programas específicos de la instrucción directa, como el uso de tarjetas⁴¹. Es el más utilizado en las aulas de educación especial para enseñar habilidades académicas y no académicas. Este sistema consiste en presentar tarjetas, una tras otra, dejando unos segundos para la respuesta del estudiante. En caso de responder incorrectamente, se presenta de nuevo el estímulo discriminativo, se ofrece ayudas, modelo y retroalimentación (Marchand-Martella et al., 2004). Las tarjetas que hayan necesitado ayuda, se sitúan al final del conjunto, para presentarlas en un nuevo ensayo. Han sido numerosos los estudios que han evaluado la eficacia del programa con poblaciones de preescolar (Mangundayao et al., 2013), de primaria (Erbey et al., 2011), secundaria (Becker et al., 2009) y con adolescentes (Brasch et al., 2008). También, hay una amplia literatura que estudia la eficacia de los nueve enfoques principales de la educación recopilados en el Proyecto Follow Through (1968 -1976). Este proyecto duró ocho años, involucró a 10.000 estudiantes y costó 500 millones de dólares (Adams & Engelmann, 1996). Se obtuvieron miles de datos, que permitió la posibilidad de hacer numerosos análisis.

Hattie (2008) examinó cuatro meta-análisis formados por un total de 304 estudios con más de 42.000 estudiantes. Los informes de todos los estudios tenían un tamaño medio de efecto del 0,59 siendo significativamente mayor que el resto de estudios realizados por Hattie. Los

41 Flashcards, en la literatura original.

resultados indicaron que la instrucción directa es eficaz con estudiantes de educación general y especial, en los niveles de grado y en diferentes áreas académicas.

4.4 Sistema personalizado de instrucción

Los cursos basados en el sistema personalizado de instrucción se caracterizan por respetar el ritmo de aprendizaje de cada estudiante mediante el uso de materiales secuenciados, por la necesidad de alcanzar el dominio de las unidades mediante prácticas repetidas, por el énfasis en materiales textuales, por el uso de conferencias motivacionales y el apoyo de supervisores que entregan retroalimentación de forma continua (Keller, 1968). Todos estos componentes, pueden observarse en otros métodos de enseñanza basados en la evidencia, pero en lo que realmente destaca este sistema es el vínculo de cada componente con la teoría del reforzamiento. Bufford (1976) corroboró que la tasa de estudio incrementa solo después de recibir reforzadores (bonificaciones) y disminuye después de una contingencia de castigo (bajas calificaciones o retirada de bonificaciones).

El sistema personalizado de instrucción ha sido un método de enseñanza eficaz, relativamente fácil de entender y aplicar en una gran variedad de entornos, poblaciones y temas. Taveggia en 1976 llevó a cabo una revisión de catorce estudios realizados en diferentes sectores (psicología, química, ingeniería, antropología, etc.). Los resultados desvelaron que los estudiantes con mayor promedio de rendimiento recibieron la enseñanza aplicada al Método Keller y no con métodos tradicionales. Kulik et al. (1979) también desarrolló un meta-análisis con 72 estudios centrándose en el rendimiento final general, las calificaciones de exámenes finales, las horas de estudio y el nivel de satisfacción. Los resultados mostraron que los cursos que utilizaban el sistema personalizado de instrucción recibieron calificaciones elevadas en todas las variables analizadas.

4.4.1 Componentes esenciales según la literatura reciente

A pesar de su época de apogeo y de la multitud de publicaciones y cursos impartidos, fue a partir del 1979 que el centro de enseñanza personalizada desapareció y toda la actividad educativa e investigadora disminuyó considerablemente. Las causas probablemente fueron debidas al cambio de enfoque de la enseñanza tradicional centrándose en el calendario académico y no en el ritmo de los estudiantes. En definitiva, se centró en los intereses de la entidad educativa y la inercia general de la enseñanza mediante la instrucción oral (Sherman, 1992).

La literatura empírica revela que no todos los componentes identificados originalmente en el listado clásico son críticos y esenciales para que funcione el Sistema personalizado de instrucción. Un aspecto esencial para lograr el “dominio” de una unidad académica es tener la capacidad de autogestionar el tiempo y cumplir con las entregas dentro de un plazo fijo. La dificultad principal de los estudiantes, supervisores y de la propia entidad es la procrastinación del estudiante que afecta al rendimiento y, generalmente, conlleva al abandono académico (Kulik et al., 1990). En este sentido, los educadores han intentado poner en práctica estrategias de economías de fichas (Eppler & Ironsmith, 2004), contratos de contingencias (Brooke & Ruthven, 1984) e incluso el uso de estrategias de autogestión para establecer plazos de entrega fijos (Roberts & Semb, 1989). Sin embargo, si el educador determina los tiempos de entrega, no se puede considerar que se está aplicando el Método Keller por no respetar el propio ritmo que dicta el estudiante. Otro aspecto que adquiere importancia es el rol del supervisor en el sistema personalizado de instrucción, en concreto, la función de retroalimentación. Kulik et al. (1979) afirmó que la retroalimentación inmediata tiene un impacto significativo en el rendimiento, independientemente del canal de entrega. Por ejemplo, Crosbie & Kelly (1993) afirman que la retroalimentación proporcionada a través de un ordenador también es efectiva.

4.4.2 Áreas de investigación

Las tecnologías de la información y la comunicación e internet han facilitado la adaptación de los sistemas personalizados de instrucción creando nuevos programas. Por ejemplo, el sistema personalizado de instrucción asistido por ordenador (CAPSI)⁴² utilizado para el desarrollo de habilidades de orden superior: habilidades verbales complejas, comprensión, pensamiento crítico y pensamiento de orden superior (Pear & Crone-Todd, 1999). Los estudiantes entran al sistema y reciben una serie de preguntas cortas sobre las lecturas que han realizado siguiendo su propio ritmo de aprendizaje. La realización de una unidad se mide como una nota de rendimiento. Cuando los estudiantes consideran que ha adquirido los conocimientos de una unidad completa solicitan una prueba de evaluación dentro del programa. El ordenador genera una prueba con preguntas aleatorias. Al finalizarla se entrega a los supervisores para que se la califiquen. Se ha examinado ampliamente el sistema CAPSI, Martin et al. (2002b) evaluaron la eficacia de la retroalimentación de los supervisores en los estudiantes y el mismo año, Martin et al. (2002a) analizaron la exactitud y la precisión de los comentarios de los supervisores mediante un análisis

42 Computer-aided Personalized System of Instruction, en la literatura original.

cuantitativo. Dentro del área del análisis aplicado de conducta se ha demostrado su eficacia en la enseñanza de ensayos discretos (Zaragoza Scherman et al., 2015) y para el uso de emparejamiento a la muestra (Oliveira et al., 2013).

Existe otro sistema informático que emplea muchos de los componentes principales del Método Keller diseñado para el aprendizaje de las matemáticas y estadística, ALEKS®. Muchas de las plataformas de *e-learning* más populares presentan componentes basados en el dominio que pueden adaptarse al Sistema personalizado de instrucción, como Moodle®, Blackboard®, WeBCT®. La aparición de internet ha facilitado que el sistema personalizado de instrucción siga aplicándose de forma eficaz gracias a las calificaciones automáticas de las pruebas de evaluación y la entrega inmediata de retroalimentación.

4.5 Interteaching

A pesar de los prometedores resultados obtenidos de los métodos de enseñanza integrales y focalizados, comentados a lo largo de este capítulo, a día de hoy las estrategias más utilizadas en la enseñanza universitaria constan de clases magistrales basadas en conferencias del educador y el rol pasivo del alumnado (Dipiro, 2009). Los posibles obstáculos que pueden explicar la falta de integración de los métodos de enseñanza conductual pueden ser (a) la cantidad de tiempo que el instructor ha de emplear para preparar los materiales y las sesiones, (b) la falta de encaje con el calendario académico universitario, por ejemplo, el sistema personalizado de instrucción (Keller, 1968) aboga por respetar el ritmo del estudiante, hecho incompatible con el cierre de expedientes semestrales, (c) la poca consistencia y sistematicidad en la aplicación de los métodos de enseñanza, ha provocado que se concluyan como ineficaces. Basándose, en estas limitaciones y extrayendo las ventajas de las aplicaciones educativas integrales y focalizadas, Boyce y Hineline (2002) introdujeron el concepto de interteaching como enfoque conductual de enseñanza terciaria que trata de remediar los supuestos problemas mencionados. En concreto, se trata de un método novedoso y flexible de enseñanza conductual respaldado empíricamente con múltiples componentes que fomentan la participación activa y la responsabilidad del estudiante en relación a los contenidos del curso académico.

Debido a su flexibilidad, no existe una estructura estandarizada que describa el uso específico de los componentes que integran una sesión de interteaching. Hurtado -Parrado et al. (2022) llevaron a cabo una revisión sistemática de la literatura e identificaron que los componentes más utilizados eran: las guías de preparación, la tutoría entre iguales, los formularios de

retroalimentación, las conferencias aclaratorias y los sondeos de evaluación. A continuación, se describirán todos los componentes que podrían aplicarse en una sesión: (a) guía de preparación, (b) tutoría con iguales, (c) contingencias de reforzamiento, (d) formulario de retroalimentación, (e) conferencias aclaratorias, (f) evaluaciones frecuentes y (g) puntos de calidad.

La estructura general de una clase típica de interteaching comienza con la elaboración de una guía preparatoria. El educador diseña esta guía teniendo en cuenta los objetivos de aprendizaje del curso y estructura el contenido según los temas que se impartirán en las clases. Cada documento contiene entre diez y doce preguntas organizadas por dificultad, desde las más sencillas hasta las más complejas de sintetizar y/o aplicar Saville et al., 2011. Normalmente, cada guía abarca quince hojas de lectura obligatoria y está disponible, en formato papel u online, varios días antes de la clase. Un tercio de la duración de la clase, el educador lo dedica a aclarar las dudas de la guía preparatoria de la sesión del día anterior.

El resto del tiempo se emplea para las sesiones de tutoría con iguales que reciben el nombre de sesiones interteaching (interteach sesion) (Sturmey et al., 2015) y ocupan dos tercios del periodo de clase (Querol et al., 2015). Durante este tiempo, los estudiantes pueden ser emparejados al azar o explícitamente, según el criterio del educador. El objetivo consiste en debatir los aspectos principales de la lectura, aclarar dudas y ayudar al compañero a comprender y aprender el temario. Durante la sesión interteaching con iguales, el educador mantiene un rol activo, observa y reorienta las interacciones, interrumpiendo los debates y redirigiendo interpretaciones erróneas. También registra qué grupos terminan antes o presentan dudas con frecuencia, y así reorganizar algunas parejas para futuras sesiones. En esta fase de aprendizaje se crean contingencias de reforzamiento que fomentan la interacción y la participación de los estudiantes y el aprendizaje analítico y práctico del contenido. En un estudio reciente liderado por Jimenez y Gayman (2021) demostraron que este componente es fundamental para impactar positivamente en el rendimiento académico de los estudiantes.

Al finalizar la clase, completan un formulario de retroalimentación que combina preguntas abiertas (por ejemplo, ¿Quién ha sido tu compañero?, ¿Qué preguntas han sido las más difícil de responder, ¿qué temas te gustaría que se abordaran en las conferencias aclaratorias siguientes?, etc.) y preguntas que evalúan la sesión en general utilizando la escala de Likert, dónde la puntuación con un “1” equivale a pobre y la puntuación de 10 equivale a excelente (Boyce & Hineline, 2002). La información que proporcionan los formularios del alumnado es útil para el educador

porque revisa el éxito de las sesiones de interteaching e identifica los temas que debe abordar en la próxima clase. Al principio de la siguiente clase, el educador realiza una conferencia aclaratoria diseñada mediante la información obtenida de los formularios de retroalimentación generados al final de las tutorías entre iguales.

Por último, el último componente esencial de interteaching es el uso de evaluaciones frecuentes o sondeos de evaluación. Consiste en evaluar el dominio del contenido adquirido a lo largo del semestre, enfatizando el material previamente preparado en las tutorías con iguales basados en las guías de preparación. Querol et al., 2015. Saville et al. (2011) realizaron una revisión sobre este componente para confirmar que la demostración frecuente de los conceptos adquiridos y la entrega de retroalimentación favorece la adquisición y la retención a largo plazo. Boyce y Hineline (2002) sugieren realizar cinco pruebas por semestre y priorizar las puntuaciones elevadas.

Además de la estructura general descrita, una sesión de interteaching puede constar de puntos de calidad, siguiendo los mismos requisitos que la práctica de aprendizaje cooperativo. En esencia, de cada pareja de estudiantes, cada uno puede ganar un punto extra o de "calidad" si ambos alcanzan el criterio máximo de puntuación en una pregunta. De lo contrario, si tan solo lo recibe uno de los dos o ninguno, nadie recibirá un punto de calidad. Boyce y Hineline (2002) sugieren el 10% de la nota global del curso del alumno puede estar destinada a los puntos de calidad.

Interteaching integra componentes de los sistemas educativos integrales; la instrucción programada (Skinner, 1986) y el sistema personalizado de instrucción (Keller, 1970). Ambos ofrecen información previamente estructurada respetando el ritmo de aprendizaje del estudiante y el instructor dedica gran parte del tiempo a elaborar recursos y planificar el temario. En cambio, interteaching ha incorporado los elementos esenciales de ambos sistemas y modificado algunos aspectos para contrarrestar los elementos débiles. Al inicio de cada sesión el instructor presenta la guía de preparación como documento que facilita a los estudiantes que trabajen a su propio ritmo, pero incorpora plazos de entregas evitando posibles abandonos o conductas de procrastinación. Por otro lado, el educador ha de centrarse en proporcionar retroalimentación en los debates que surgen en las tutorías con iguales y esto implica una reducción en el tiempo destinado a la elaboración de materiales. Por último, las conferencias aclaratorias tienden a ser breves debido a que están destinadas a resolver dudas. También, interteaching integra dos sistemas educativos

focalizados; como son el aprendizaje cooperativo, la tutoría con iguales y la retroalimentación, tal y como se han descrito anteriormente.

Desde la introducción de este nuevo enfoque de enseñanza, un número creciente de investigaciones empíricas ha estudiado su eficacia. En 2005, Saville et al. (2005) llevaron a cabo el primer análisis experimental y compararon el método de interteaching con clases magistrales y lecturas. Para ello, crearon tres grupos de estudiantes (1) los que leyeron el artículo y realizaron la evaluación del contenido, una semana más tarde (grupo control), (2) aquellos que leyeron el artículo, asistieron a una clase magistral sobre el contenido del artículo y más tarde pusieron sus conocimientos a prueba con una evaluación y (3) los restantes que ejecutaron los mismos pasos que el grupo anterior, pero en lugar de asistir a una clase magistral estándar, recibieron una clase de interteaching. Los resultados demostraron que el grupo que recibió el nuevo enfoque educativo obtuvieron mejores resultados en las evaluaciones en comparación al resto de grupos. En un estudio posterior realizado en el aula universitaria, Saville et al. (2006) alternaron sesiones de interteaching con clases magistrales y las puntuaciones de las evaluaciones realizadas después de las clases de interteaching fueron más altas que las clases basadas en conferencias. Posteriormente, Arntzen y Hoium (2010) compararon una sesión de cada una de las modalidades de enseñanza (clase magistral y clase de interteaching) y los resultados de satisfacción de los alumnos informaron una clara preferencia por las sesiones de interteaching. Sin embargo, un estudio similar, con 19 estudiantes que recibían sesiones alternas de cada enfoque, no se hallaron diferencias significativas en las puntuaciones de las evaluaciones (Zayac & Paulk; 2013). Sin embargo, un estudio realizado por Saville et al. (2014) con una muestra más amplia (134 estudiantes) demostró que los conocimientos obtenidos durante las sesiones de interteaching tienden a retenerse y recordarse con mayor facilidad que los conceptos recibidos en clases magistrales.

Recientemente, Rieken et al (2018) compararon sesiones de interteaching adaptadas a una plataforma e-learning (Blackboard®) con sesiones online de enseñanza tradicional. Las sesiones que ponían en práctica el nuevo enfoque educativo utilizaban la herramienta de videoconferencia para llevar a cabo la tutoría con iguales, rellenaban el formulario de retroalimentación y escuchaban las conferencias aclaratorias en modalidad asíncrona. Los resultados replican los hallazgos anteriores ya que los estudiantes que recibieron las sesiones de interteaching obtuvieron un rendimiento superior a los estudiantes que recibieron instrucción online tradicional.

El creciente interés de este enfoque en la educación superior ha comportado un aumento de estudios que analizan su eficacia. Según las revisiones anteriores podemos concluir que el enfoque interteaching produce mejores resultados que las sesiones de enseñanza tradicional en estudiantes universitarios del área de psicología y es adaptable a funcionalidades de plataformas de e-learning. No obstante, para futuras investigaciones sería recomendable extender su aplicación con otras modalidades de enseñanza, como la enseñanza invertida⁴³, la integración de plataformas de redes sociales, cursos híbridos, etc. Además de incluir estudios longitudinales que demuestren una retención permanente de los conocimientos y ensayos controlado aleatorizados para eliminar cualquier sesgo en los efectos de la intervención.

5. CONSIDERACIONES FINALES

A lo largo de este capítulo se ha mencionado el recorrido de los primeros investigadores que contribuyeron a la aplicación sistemática de principios conductuales para abordar conductas socialmente significativas en el ámbito académico. Desde la creación de las máquinas de enseñanza, incorporando dispositivos electrónicos como canal para proporcionar retroalimentación a los estudiantes. A la introducción de la instrucción programada, una técnica de enseñanza precursora y autodidáctica que se rige por los principios conductuales de una instrucción eficaz. Hasta la presentación de modelos educativos conductuales que incorporan los principios de instrucción eficaz: el sistema personalizado de instrucción, la enseñanza de precisión y la instrucción directa. Todos ellos hacen hincapié en la identificación de objetivos claros, completos y precisos, en mantener una tasa elevada de oportunidades de respuesta con el fin de fomentar una respuesta activa del estudiante y así tener la posibilidad de entregar retroalimentación por el desempeño del alumno. Estos enfoques educativos se pueden llevar a cabo mediante softwares educativos o bien en un aula aplicando la enseñanza tradicional. El resultado final de emplear mecanismos basados en la enseñanza eficaz es transmitir conocimientos críticos (por ejemplo, escritura, lectura, matemáticas, etc.) a los estudiantes para que se conviertan en estudiantes autónomos capaces de estar aprendiendo constantemente de cualquier fuente de información. Teniendo en cuenta las aplicaciones educativas integrales revisadas en este capítulo, tan solo dos se han focalizado en la educación terciaria: el sistema personalizado de instrucción e interteaching.

43 *Flipped classroom*, en la literatura original

Después de profundizar en los principios conductuales que se rigen los métodos educativos de la enseñanza eficaz, avanzamos en la literatura con las técnicas o procedimientos derivadas de los principios de la ciencia. Todos ellos se tratan de procedimientos basados en la evidencia y se utilizan con frecuencia en las aulas, normalmente desde preescolar hasta secundaria, con personas de desarrollo típico o con discapacidad. Esto nos permite concluir que es escasa la literatura en el ámbito analítico-conductual que utiliza aplicaciones educativas focalizadas con estudiantes de posgrado.

Finalmente, el capítulo termina con la descripción y descomposición de las características que integran los enfoques educativos integrales, comentados con anterioridad. La aplicación de los principios conductuales y sus técnicas derivadas en la educación permiten que el alumnado tenga acceso a enseñanzas objetivamente eficaces.

CAPÍTULO 4. Aplicaciones analítico-conductuales en el ámbito educativo mediante TICS

1. APLICACIÓN DE SOFTWARES EDUCATIVOS EN EDUCACIÓN TERCIARIA

Los orígenes de la instrucción automatizada desde las máquinas de enseñanza de Pressey (1924) y Skinner (1954) han evolucionado y se han adaptado a las nuevas tecnologías de la información y la comunicación (TIC). Al ritmo que avanza la tecnología surgen nuevas posibilidades de crear canales de enseñanza automatizados. Este capítulo tiene como objetivo introducir el concepto de la enseñanza asistida por ordenador y explicar cuáles son los principales factores que contribuyen a una enseñanza eficaz utilizando softwares educativos. Por otro lado, se expone una breve introducción al uso de la telesalud en los últimos años, enfatizando su aplicación en el análisis aplicado de conducta. Y finalmente, se presenta la importancia que han adquirido las redes sociales en nuestra sociedad y en el ámbito educativo. Ante la carencia de estudios en el área del análisis de conducta que integren las redes sociales como canal de enseñanza para aplicar procedimientos analítico-conductuales, se dedica un apartado a comentar las posibles funcionalidades que aportan las redes sociales y pueden compatibilizarse con las intervenciones focalizadas e integrales comentadas en el capítulo anterior.

1.1 Enseñanza asistida por ordenador o software educativos

En la década del 1960 y 1970, los sistemas de enseñanza computarizados basados en la instrucción programada comenzaron a popularizarse entre los desarrolladores de software educativos debido a la capacidad de proporcionar estímulos y retroalimentación inmediata de forma sistemática.

El concepto software educativo es un término que engloba cualquier programa informático que haya sido diseñado explícitamente para la enseñanza y el aprendizaje; esto excluye los procesadores de palabras y base de datos. Este concepto también se conoce como plataformas educativas, programas educativos o informática educativa. Un sistema de software educativo ha de cumplir con una serie de requisitos como (a) la necesidad de utilizar un dispositivo electrónico,

(b) la facilidad de manejar el programa, sin conocimientos previos, (c) la existencia de funciones interactivas que promuevan la participación de los estudiantes y la posibilidad de enviar retroalimentación, además de (d) permitir la individualización del ritmo de aprendizaje. También, cada software tiene diferentes propósitos según las necesidades académicas que esté destinado a cubrir, en consecuencia, dispone de sus propias funciones. Las principales funciones son los programas de ejercicios y práctica⁴⁴, tutoriales y simulacros. El educador ha de seleccionar el tipo de software que corresponda según los objetivos académicos.

1.1.1 Software basado en ejercicios y práctica

Este programa de enseñanza sienta sus bases en el principio de reforzamiento. Está caracterizado por la presentación repetida de estímulos y respuestas, las respuestas correctas conducen a un reforzador inmediato y como consecuencia, fortalecen la conducta futura de responder correctamente, facilitando así la fluidez del aprendizaje. Concretamente, mide la velocidad de respuesta y el número de respuestas correctas respondidas en un tiempo determinado (Lovitz et al., 2020). Weigandt (2021) evaluó empíricamente la eficacia de dos softwares de ejercicios y práctica (por ejemplo, Quizlet y Cram) utilizados para adquirir fluidez en el aprendizaje de conceptos. Los estudiantes que utilizaron los softwares educativos obtuvieron mejores resultados en las puntuaciones de las evaluaciones finales en comparación al rendimiento de los estudiantes que utilizaron papel y lápiz.

1.1.2 Software basado en tutoriales

Los tutoriales son un tipo de software que presenta una secuencia de enseñanza completa sobre un tema concreto y requieren que el estudiante practique y reciba retroalimentación. Pueden adquirir un formato de texto, audio o video utilizando las tecnologías actuales, por ejemplo, YouTube. Los tutoriales pueden enseñar conocimientos declarativos, como el estudio realizado por Reed y Kaplan (2011) que hicieron uso de softwares utilizando tutoriales con el fin de enseñar los principios conceptuales y las posibles interpretaciones prácticas sobre el concepto de la “ley de igualación”.

Además, son útiles para enseñar aprendizajes procedimentales, por ejemplo, Vanselow y Bourret (2012) concluyeron que el uso de tutoriales es más eficaz y eficiente que otros métodos

44 *Drill and practice*, en la literatura original.

más tradicionales (lectura independiente o clases magistrales) para enseñar a construir gráficos a profesionales del análisis de conducta.

1.1.3 Software basado en simulacros

Los programas basados en simulacros permiten crear y presentar modelos de situaciones realistas diseñados para enseñar, poner en práctica y reforzar habilidades objetivo. Este recurso didáctico promueve la generalización de habilidades entrenadas en situaciones de aprendizaje artificial permitiendo la exposición repetida de la persona en el contexto natural (Chernikova et al., 2020). A diferencia de los dos softwares anteriores que estaban formados por paquetes de enseñanza estructurados, los estudiantes que utilizan los simulacros escogen el orden de la secuencia hasta emitir la habilidad compleja. El uso de softwares de simulacros se utiliza con frecuencia en la educación superior en áreas como la medicina, ingeniería, tecnología, etc. (Chernikova et al., 2019).

Desrochers et al., (2000) evaluó empíricamente la eficacia de un software en base a simulacros que proporciona a estudiantes de psicología aplicada cómo diseñar evaluaciones e intervenciones conductuales. El programa informático recibe el nombre de simulacros en el ámbito de discapacidades del desarrollo⁴⁵(SIDD, sus siglas en inglés). El rendimiento obtenido por el grupo experimental fue más elevado que el grupo control; que no recibió entrenamiento con el software. Es más, los estudiantes que lo utilizaron valoraron positivamente el programa informático. Así pues, el software SIDD es una herramienta útil para practicar habilidades conductuales.

1.2 Componentes eficaces para elaborar una enseñanza asistida por ordenador

Las características esenciales para que los softwares educativos sean herramientas eficaces han de presentar tres componentes: (a) la presentación de estímulos discriminativos mediante el uso de instrucciones o reglas gobernadas junto con la presentación y desvanecimiento de ayudas, (b) la oportunidad del estudiante de emitir las habilidades objetivo y (c) la posibilidad de que el estudiante reciba retroalimentación inmediatamente después de la respuesta.

1.2.1 Estrategias de enseñanza en base a los antecedentes

El primer componente de intervención con base en los antecedentes es la presentación de una instrucción. Las instrucciones describen reglas y pueden ejecutarse mediante conductas moldeadas.

45 *Simulation in Developmental Disabilities (SIDD)*, en la literatura original.

das por las contingencias (Skinner, 1974) o conductas gobernadas por la regla (Skinner, 1969). Por lo general, los estudiantes siguen reglas porque presentan una historia de reforzamiento relacionadas con las contingencias descritas en la regla, este hecho describe el aprendizaje de conductas moldeadas por las contingencias. Por ejemplo, el aprendizaje de conceptos o contenido proposicional. En cambio, el aprendizaje sobre habilidades procedimentales; que requieren de una cadena de pasos, pueden adquirirse mediante conductas gobernadas por la regla. Skinner las describió como aquellas conductas que se emiten como resultado de la descripción de las contingencias proporcionadas en la regla y no debido al contacto previo con las consecuencias. Por ejemplo, Emurian (2005) demostró empíricamente la eficacia del aprendizaje basado en reglas para enseñar a programar a estudiantes universitarios utilizando un software específico de programación, Java.

Por otro lado, para garantizar la adquisición de nuevas habilidades, se pueden integrar en los softwares educativos, estímulos que funcionen como ayuda. Por ejemplo, modificar el tamaño y el color de los caracteres, utilizar el subrayado del texto para resaltar información, añadir imágenes o iconos para atraer la atención de los estudiantes. Cualquier tipo de ayuda o pista, debe retirarse gradualmente a medida que los estudiantes respondan correctamente al estímulo discriminativo para evitar dependencias a las ayudas (Gorgan & Kodak, 2019). Este procedimiento recibe el nombre de desvanecimiento (Morris et al., 2005).

1.2.2. Estrategia de enseñanza en base a la conducta deseada

El uso de softwares educativos, en concreto, los basados en ejercicios y práctica que presentan un elevado número de oportunidades de respuesta promueven respuestas activas de los estudiantes; ambos conceptos comentados en el capítulo anterior. Así pues, la enseñanza mediada por softwares educativos facilita la obtención de datos sobre la fluidez de respuesta de los estudiantes (tiempo y número de respuestas correctas) para así, observar la capacidad de mantener las habilidades adquiridas a lo largo del tiempo; extrayendo la medida sobre el mantenimiento de las respuestas.

Por otro lado, cualquier software educativo permite al educador crear instrucciones de preguntas abiertas o de opción múltiple, intercambiándolas entre sí para fomentar la atención de los estudiantes. No obstante, Funk & Dickson (2011) analizaron experimentalmente los efectos de ambos tipos de preguntas en el rendimiento de estudiantes universitarios. Los resultados concluyeron una mejora en los ítems presentados con preguntas de opción múltiple.

1.2.3 Estrategias de enseñanza en base a las consecuencias

Una de las características principales de la enseñanza asistida por ordenador es la capacidad de ofrecer retroalimentación inmediata e individualizada a los estudiantes. Debido a su importancia, en el capítulo anterior se ha descrito minuciosamente el impacto que puede tener en el rendimiento.

1.3 Sistemas asistidos por ordenador en análisis aplicado de conducta

Existe una larga historia sobre el desarrollo inicial de la enseñanza automatizada atribuida al análisis de conducta. Todo comenzó con las máquinas de enseñanza desarrolladas por Sidney L. Pressey en los años veinte donde los estudiantes respondían a preguntas de opción múltiple. En la década de los 50 y 60, B.F Skinner refinó el hardware de las máquinas de enseñanza de Pressey y personalizó el tipo de preguntas que se presentaban a los estudiantes. En este caso, eran los alumnos que generaban sus propias respuestas. Según Skinner, el aprendizaje “real” se desarrollaba cuando los estudiantes escribían las respuestas; más allá de seleccionar una opción presentada por la máquina. La ventaja principal de este sistema era la capacidad que tenía de ofrecer retroalimentación inmediata al estudiante, ya bien reforzando las respuestas correctas o corrigiendo las incorrectas. Este acontecimiento originó nuevas contingencias respecto la aplicación sistemática de técnicas conductuales de aprendizaje con base empírica en la educación superior (Terenzini & Pasacarella, 1991).

El modelo conductual que combinó un sistema de enseñanza y la tecnología digital fue el sistema personalizado de instrucción desarrollado por un colega de Skinner, F.S. Keller (Keller, 1968). El proceso de aprendizaje de este modelo se fundamenta en los principios de reforzamiento sustituyendo el uso de consecuencias de castigo que acostumbra a utilizar el modelo tradicional (Grant & Spencer, 2003). Los componentes esenciales del SPI son: (a) énfasis en la comunicación escrita, por lo tanto, el educador selecciona libros y materiales textuales (b) desglose del material textual de estudio en pequeñas unidades, (c) preguntas de estudio relacionadas con el material textual proporcionado, (d) demostración del dominio del material del curso, (e) autogestión del ritmo del estudiante, (d) uso de estudiantes más avanzados para ejecutar el rol de “supervisor” y administrar pruebas de evaluación y ofrecer retroalimentación. Por lo tanto, se trata de un enfoque de enseñanza estructurado que facilita la organización de los objetivos y permite la individualización del contenido, es por este motivo, que los estudiantes mejoran su rendimiento (Kulik et al., 1990).

A raíz del incremento generalizado de ordenadores personales y el auge exponencial de internet, Pear y sus colegas de la Universidad de Manitoba, desarrollaron en 1980 un programa informático online basado en los principios del SPI (Pear & Kinsner, 1988). El software se conoce como Sistema Personalizado de Instrucción asistido por ordenador (CAPSI®). CAPSI® es la versión informatizada del SPI más utilizada en la educación superior (Springer et al., 2008). Se trata de un programa flexible y basado en internet que permite el acceso desde cualquier ordenador con conexión; indiferentemente del lugar dónde se encuentre el estudiante (Pear & Novak, 1996). El diseño de un curso impartido por CAPSI® comienza por la identificación y el desglose de objetivos de estudio, denominados unidades y le sigue la elaboración de preguntas cortas sobre el contenido a evaluar. Además, al tratarse de un sistema electrónico que facilita datos de los días y horas de acceso, permite configurarlo para registrar patrones conductuales, como el porcentaje obtenido después de cada evaluación, la cantidad de unidades dominadas al día, los tiempos entre respuestas y las puntuaciones obtenidas en cada unidad, permitiendo observar el progreso del estudiante.

Cuando los estudiantes consideran que dominan el temario, entran al sistema de CAPSI® y el propio software aleatoriza un número de preguntas e inicia la evaluación. Una vez el alumno responde, automáticamente le aparece una calificación en comparación al rendimiento de sus compañeros. Ambos métodos de aprendizaje comparten características principales como la responsabilidad del estudiante de gestionar su ritmo de aprendizaje escogiendo dónde y cuándo estudiar. También, CAPSI® permite que los alumnos decidan qué unidad de contenido realizar; sin necesidad de seguir un orden pautado. Las preguntas del contenido son breves y claras, incluyendo solo el temario escrito proporcionado, esto excluye el contenido que se haya presentado en formato verbal. Para continuar con unidades nuevas, primero se ha de aprobar una breve evaluación, de lo contrario, han de repetirla hasta dominar el temario y obtener un aprobado, sin recibir penalizaciones por el número de intentos. Los cursos que utilizan CAPSI® incorporan más de un supervisor para evaluar el aprobado de una unidad, necesitando una respuesta de aprobado independiente de cada uno de ellos. Ambos programas se basan en la incorporación de la figura de un supervisor o un estudiante que domina y que ha completado el temario evaluado y, por lo tanto, tiene las habilidades de subministrar retroalimentación a sus compañeros. También existen diferencias entre ambos programas, la modalidad SPI puede utilizar preguntas de opción múltiple o de verdadero y falso y la modalidad CAPSI® demuestra el dominio del contenido a través de las respuestas breves escritas. Una característica única de este último programa es la transferencia

de rol en aquellos estudiantes que demuestran dominar una unidad de contenido mediante un aprobado, considerando que tienen las habilidades suficientes para evaluar exámenes de sus compañeros. Esta variable garantiza que el número de estudiantes no es un inconveniente ni resta eficacia al sistema porque a medida que aumentan los alumnos que superan una unidad, aumenta exponencialmente el número de supervisores; garantizando una retroalimentación frecuente y rápida.

Este sistema ofrece una variedad de áreas de investigación como la eficacia de la retroalimentación del programa CAPSI® (Martin et al., 2002b), el uso de conductas gobernadas por reglas como procedimiento para aumentar la precisión en la formación de supervisores y en el redactado de preguntas de examen por los supervisores (Wirth, 2004). También, existen estudios centrados en la tasa de respuestas escritas de los estudiantes (Crone-Todd et al., 2007) y en el incremento de la participación en estudiantes de psicología general (Dempsey, 2005).

Se conocen otras plataformas que se basan en el dominio y pueden configurarse para integrar algunos componentes del SPI, como por ejemplo Moodle® y Blackboard®. Introducir un número máximo de intentos en unidad de contenido, establecer un porcentaje de aprobado, añadir una fecha máxima de realización, etc. También se conocen sistemas informáticos que utilizan los principios del SPI, diseñados para presentar preguntas de opción múltiple, como por ejemplo ALEKS® programado con el fin de facilitar la comprensión de conceptos matemáticos en estudiantes universitarios.

Actualmente, no hay ningún programa informático que emplee CAPSI® en el mercado. El único que existe es una versión actual que recibe el nombre DOS® y funciona desde sus inicios, en la Universidad de Manitoba, pero es difícil de transferir o incorporar en otros programas online o basados en la web.

2. ENSEÑANZA ASISTIDA POR ORDENADOR Y TELESALUD

2.1 Telesalud

Tradicionalmente, los servicios basados en análisis aplicado de conducta se impartían en formato presencial. Sin embargo, a principios del 2020 debido a la reciente pandemia mundial ocasionada por la enfermedad del Coronavirus (Covid-19) muchos profesionales que ejercían en el ámbito educativo recurrieron a presentar servicios utilizando el modelo de telesalud.

El término telesalud hace referencia al uso global de las TIC con el fin de prestar servicios sanitarios y de educación sanitaria cuando instructores y alumnos se encuentran en lugares geográficos diferentes (Organización Mundial de la Salud, 2021). Actualmente, no existe ninguna definición universal para este concepto. Uno de los servicios que ofrece la “telesalud” incluye la “telemedicina”; conceptos que acostumbran a intercambiarse erróneamente. La telemedicina, a nivel práctico, no contiene un componente educativo (Moore, 1999), por este motivo a lo largo de la tesis solo se hará referencia al término telesalud.

Este concepto abarca los servicios de consulta, diagnóstico, formación, evaluación e intervención (Hayiroğlu, 2019). En concreto, nos centraremos en el área que está relacionada la presente tesis, la formación. La entrega de este servicio online consiste en ofrecer herramientas de formación sobre temas especializados utilizando herramientas tecnológicas que permitan debatir entre los iguales y proporcionar recursos educativos a los estudiantes (American Telemedicine Association, 2016).

En la literatura se distinguen dos modalidades apoyadas en las contingencias de interacción entre estudiantes e instructores y la temporalidad. Ofrecer servicios de forma síncrona implica una interacción en tiempo real utilizando cualquier plataforma online o dispositivo electrónico. Normalmente, se tratan de consultas o de formaciones utilizando herramientas de videoconferencia, como Skype® (Pantermuehl & Lechago, 2015) o Facetime (Boutain, 2020). El generar una comunicación bidireccional en tiempo real se asemeja a las interacciones entre estudiantes y profesorado que se originan en aulas de enseñanza tradicional. En el caso de la modalidad asíncrona la transmisión de la comunicación e información se realiza en diferentes momentos temporales. Este formato ofrece flexibilidad a ambas partes de la interacción y promueve que los estudiantes utilicen habilidades de autogestión. Es común el uso de aplicaciones como el correo electrónico o Dropbox® para compartir productos permanentes entre profesorado y alumnado (Neely et al., 2020). También, plataformas educativas como Moodle® o Blackboard® han resultado ser eficaces como herramientas de formación asíncrona (Gamage et al., 2022).

Un estudio reciente de Zhu et al., (2021) concluye que la retroalimentación recibida a través de una videoconferencia es más efectiva que no mediante correo electrónico. Lo cual se podría concluir que la modalidad síncrona es más efectiva que la asíncrona. Sin embargo, Farros et al., (2020) expuso a estudiantes de posgrado en formación sobre el análisis de conducta, a dos

tipos de retroalimentación: síncrona y asíncrona. El objetivo era determinar cuál era más efectiva. Los resultados concluyeron que la modalidad no impactaba en el rendimiento de los estudiantes.

2.2 Telesalud en el análisis aplicado de conducta

Es fundamental promover servicios de formación e intervención conductual de calidad en el ámbito de la educación y del análisis aplicado de conducta (Tomlinson et al., 2018). Durante las últimas dos décadas, la telesalud ha demostrado ser una herramienta útil para proporcionar servicios analítico conductuales y ofrecer formación a los profesionales sobre la aplicación de procedimientos específicos (Wacker et al., 2017). La primera aplicación tuvo lugar en 1987 en la universidad de Lend, en el estado de Iowa. Se trataba de un proyecto financiado con el fin de ofrecer servicios a niños con discapacidad residentes en zonas rurales. Los primeros estudios consistían en evaluar funcionalmente conductas graves y autolesivas de menores (Barretto et al., 2006), a posteriori, proponían intervenciones funcionales después de evaluar las conductas desafiantes. Por ejemplo, Wacker et al. (2013) utilizó la herramienta de telesalud para formar diecisiete padres a utilizar la estrategia de entrenamiento en comunicación funcional con sus hijos con TEA.

Actualmente, el uso de las plataformas de telesalud ha sido útil para la prestación de servicios analítico conductuales a personas que están a cargo de niños con discapacidad intelectual y TEA (Ferguson et al., 2019; Ellison et al., 2021). Por ejemplo, ha resultado eficaz en evaluaciones de preferencias (Ausenhuis, 2019), evaluaciones conductuales y análisis funcionales de conductas desafiantes y autolesivas (Schieltz, 2020). También, es útil en la formación a padres, terapeutas, profesores o cuidadores de niños o adolescentes, con el objetivo que pongan en práctica procedimientos que faciliten las habilidades imitativas (Wainer et al., 2015), comunicativas (Akemoglu et al., 2020) y de la vida diaria como, por ejemplo, reducir conductas de insomnio (McLay et al., 2020). Por otro lado, una creciente base de estudios científicos indica la posibilidad de prestar directamente los servicios conductuales a los clientes utilizando cualquier sistema de telesalud en modalidad síncrona. Por ejemplo, Pellegrino y Reed (2020) enseñaron a dos adultos, mediante videoconferencia, a realizar una comida utilizando procedimientos de encadenamiento, ayudas y desvanecimiento. Sin embargo, hasta la fecha no existe ninguna investigación analítico-conductual que utilice la modalidad de telesalud para formar directamente a estudiantes de desarrollo típico que cursen en educación terciaria.

3. ENSEÑANZA ASISTIDA EN LA EDUCACIÓN SUPERIOR CON SOPORTE DE REDES SOCIALES

3.1 Las redes sociales

En el capítulo anterior mencionamos que la aparición de internet en el 2003 y la innovación del software PLATO®, un programa informático basado en la instrucción programada permitió a cualquier persona con acceso a un ordenador con internet pudiera acceder a la universidad (Bitzer y Johnson, 1971). Después de la incorporación de la enseñanza mediada por ordenadores en los contextos educativos y la introducción de ordenadores de menor tamaño y económicamente asequibles, en 1975 el uso de la tecnología se consideró una necesidad en la educación (Lee, 2004). No fue, hasta el 1980 con la llegada de ordenadores personales y portátiles que se consideró obligatoria la incorporación de la tecnología en las aulas de educación superior (Prescott, 2015).

Desde el 1989 hasta inicios del 2004, el aprendizaje mediado por ordenadores empezó a crecer exponencialmente con llegada de la Web®. Un recurso de búsqueda que permitía el acceso a cantidades masivas de información con posibilidad de lectura, conocida como Web 1.0® (Kujur y Chhetri, 2015). Este avance en la era de la tecnología permitió que las universidades comenzaran a ofrecer cursos híbridos, impartiendo algunas materias a través de internet. En 2004, Dale Dougherty refinó la Web® y surgió la segunda generación, la Web 2.0®; definida como una plataforma online con posibilidad de interactuar, crear, colaborar y compartir innumerables contenidos digitales. Los avances de las TICs incluyen tanto el desarrollo de las redes sociales como el uso de la nube, provocando que las instituciones educativas de enseñanza superior se esfuercen por integrarlas en los procesos de enseñanza y aprendizaje.

3.1.1 Definición

En 1954, el antropólogo Barnes, fue el primero en definir el uso de dos conceptos que se superponen en inglés “social media” y “social networking sites”, traducidos al español con el mismo término, “redes sociales”, de forma indistinta. El término en inglés “social media” son estructuras o medios compuestos por personas o entidades conectadas por motivos diversos: laboral, intercambios económicos, amistad (Aichner et al., 2021). En cambio, el concepto en inglés “social networking sites”, se emplea para denotar las relaciones sociales entre miembros de un grupo. Es decir, una vez se ha decidido qué medio utilizar, cualquier persona puede escoger una aplicación para interactuar con las personas a través de internet. Por ejemplo, Facebook® o Twitter® son estructuras sociales que unen a personas con un interés común y ofrecen la libertad de crear

relaciones, buscar seguidores, generar debates o conectarse con su audiencia en línea. A lo largo de este trabajo, nos referiremos en todo momento al término “social networking sites” y lo traduciremos como “redes sociales”.

Las redes sociales son aplicaciones que se fundamentan en los principios de la Web 2.0® y proporcionan a los usuarios espacios sociales de colaboración e intercambio de contenidos sin necesidad de estar presentes en un espacio físico (Kaplan y Haenlein, 2010). Su proliferación ha sido extensa y generalizada y ya forma parte de la vida cotidiana de muchas personas, independientemente de la edad. Actualmente, 3.960 millones de usuarios de la población mundial son usuarios activos de las redes sociales (Statista, 2021). Este crecimiento ha provocado cambios en las contingencias de comunicación e interacción social con la introducción de nuevas plataformas. En 2003, apareció LinkedIn®, en 2004 Flickr® y Facebook®, un año después le siguió YouTube® y en el 2006® lanzaron Twitter (Baruah, 2012). De entre los más populares, Facebook® se sitúa el primero con 2.919 millones de usuarios, le sigue YouTube® (2.562 millones), WhatsApp® (2.000) e Instagram® (1.478 millones) (Statista, 2022).

3.1.2 Características

En general, las redes sociales invitan a los usuarios a crear una presencia en línea, compartir imágenes, videos, publicar actualizaciones de forma periódica, compartir información y conversar de forma asíncrona con los contactos de la misma plataforma (Boyd y Ellison, 2007). Según Mayfield (2008) las redes sociales comparten cinco características. Las redes sociales (a) fomentan las respuestas de interacción y participación, (b) están formadas por componentes que facilitan la conversación bidireccional, (c) permiten la interconexión de recursos con la función de agregar enlaces a personas, recursos u otras plataformas, por lo tanto, fomentan la conectividad, (d) crean un entorno propicio para crear comunidades con usuarios que tienen algún elemento en común y (e) favorecer que los usuarios intercambien y aporten opiniones desde el compromiso personal (apertura).

Las redes sociales tienen un impacto en la vida de niños, adolescentes y adultos, hasta el punto, de convertirse en el canal de interacción social cotidiano (Boyd, 2007). Este hecho se debe al fácil acceso, el amplio alcance de conversar sin límites de tiempo ni distancia, el bajo coste y la inmediatez de respuestas del resto de usuarios. No obstante, a pesar de todas las ventajas mencionadas, el uso de las redes sociales se ha de controlar y gestionar un tiempo específico

para evitar cualquier adicción; sobre todo cuando los usuarios son menores de edad (Chassiakos et al., 2016). Para contrarrestar el efecto negativo de las redes sociales en menores de edad, es necesaria la supervisión de los padres.

3.1.3 Facebook como modelo de red social con aplicación educativa

La red social con más usuarios activos del mundo fue desarrollada por un joven estudiante de la Universidad de Harvard, Mark Zuckerberg. Pretendía ofrecer un medio de socialización para los futuros estudiantes de su universidad. Facebook® es una herramienta originaria de la Web 2.0® que permite el intercambio de información y de recursos de forma síncrona y asíncrona a través de una plataforma tecnológica (Milošević et al., 2015). Las personas pueden dialogar enviando mensajes en un foro abierto, llamado “muro” o bien enviar mensajes por privado de recepción inmediata. Facebook®, ofrece la posibilidad de crear grupos privados, sin necesidad de ser “amigo” ni de exponer información personal. Originalmente, está destinada a interactuar socialmente enviando comentarios al resto de usuarios, planteando preguntas o compartiendo cualquier enlace de información mediante fotos, imágenes, documentos o videos. Al ser la plataforma más popular, se han realizado varios estudios desde diferentes disciplinas sobre su uso en áreas como biblioteconomía, medios de comunicación, psicología, información de gestión, sistemas de negocios, ciencias políticas, marketing, comercio y contabilidad. En el ámbito educativo no se ha vislumbrado totalmente su validez como herramienta de enseñanza (Prescott et al., 2013).

No obstante, su uso podría beneficiar tanto a estudiantes como docentes. Facebook, es un espacio propicio para que se produzca un entorno de colaboración e interacción social (Demir, 2018), aumentando así el trabajo cooperativo entre estudiantes (Al-Rahmi & Zeki, 2017), también facilita la comunicación bidireccional con los compañeros e instructores (Qureshi et al., 2014) y favorece el desarrollo de habilidades comunicativas. Los estudios que han utilizado esta plataforma en aulas destinadas al aprendizaje de un segundo idioma han demostrado que mejoran las habilidades de lectura y escritura (Buga et al., 2014). También, facilita la retroalimentación inmediata utilizando las herramientas de “me gusta” y “comentarios” que generan una notificación instantánea al receptor. Un estudio de Erjavec (2013) encontró que los estudiantes incrementaron el número de tareas publicadas al utilizar Facebook®, con el fin de recibir “me gustas” del resto de miembros del grupo. Los docentes pueden beneficiarse porque Facebook® facilita que los estudiantes se involucren en las tareas fuera del aula iniciando debates en los grupos cerrados (Nicolai et al., 2017), además, la información que publica el docente llega a todos los

miembros por igual, sin distinciones personales (por ejemplo: diferente volumen o tono de voz por estar ubicado al fondo del aula). También ha resultado útil como recurso para mantener la comunicación y crear comunidad ante una catástrofe natural; como por ejemplo el terremoto de Canterbury del 2010 (Dabner, 2012). Sin embargo, se ha destacado que una de las características esenciales de Facebook es la capacidad de fomentar cohesión y comunidad con los miembros de un grupo cerrado (Hurt et al., 2012).

Facebook, al ser una herramienta de la Web® destinada principalmente a la interacción social, a pesar de que los miembros del grupo la utilicen para debatir los temas estudiados en el aula o para compartir información académica, los estudiantes siguen utilizándola como herramientas de uso personal y para mantener el contacto con sus amigos (Grosbeck et al., 2011)

3.2 Educación superior y redes sociales

Los inicios de internet se regían por la única función unidireccional, la posibilidad de consultar y descargar información. En su lugar, desde el 1970 la Web® adquiere un carácter social con las primeras aplicaciones de internet que permiten intercambiar información, crear blogs y conocer a personas. Más tarde, las aplicaciones de la actual Web 2.0®, conocidas como “redes sociales” permiten a los usuarios interactuar, crear, compartir contenidos visuales, auditivos, textuales e incluso permiten “etiquetar” y “recomendar” contenido ya existente. Siendo la “interacción sin límites” la característica principal. Hoy día, las redes sociales promueven actividades participativas, colectivas, creativas y de convivencia.

En medio de estos avances tecnológicos se encuentran las instituciones de educación superior y de formación profesional con el reto de adaptarse a los usuarios de las redes sociales. Coincide que los usuarios más activos se encuentran en la franja de entre los 25 y 44 años, dedicando aproximadamente una hora (59,85 minutos) al día al consumo de las redes sociales (Statista, 2021). Esto comporta que los estudiantes de la era digital del siglo XXI son usuarios activos de las redes sociales que desarrollan una serie de cualidades conectadas con las aplicaciones de estas prácticas. Por ejemplo, la tendencia actual de los jóvenes es lidiar con la multitarea, mientras consumen contenidos visuales, llamativos y dinámicos (Karle, 2013).

No obstante, el reto actual más apremiante para la enseñanza terciaria, es hallar la mejor forma de emplear las características de las redes sociales para apoyar nuevas estrategias de aprendizaje asentadas en la creación de una comunidad para compartir y debatir conocimientos

(Alkhatlan & Al-Daraiseh, 2017). Uno de los primeros cambios a realizar sería cambiar el foco de los principios centrales de la educación formal. La introducción de las redes sociales en el sistema educativo implica que los estudiantes pasan a ser “creadores activos de información” en lugar de ser “consumidores pasivos” (Chen, 2014). La esencia de las redes sociales consiste en impulsar un rol activo y autónomo permitiendo a los estudiantes tomar decisiones sobre qué, cuándo aprender y cómo aprenderlo. Actualmente, se ha hecho hincapié en integrar las redes sociales desde que las universidades han empezado a distribuir contenidos en YouTube®, iTunes®, Google Earth®, etc. Son muchas universidades que han apostado por la incorporación de algunas redes sociales, con el fin de crear una comunidad entre profesores y estudiantes, para así contribuir en debates y fomentar un aprendizaje enriquecedor y motivante.

La red social más popular entre los estudiantes universitarios y de posgrado es Facebook (Cheng et al., 2016) es por este motivo que las universidades han aceptado integrarlo como canal de enseñanza y aprendizaje; ya que la gran mayoría pertenecen al este espacio virtual (Woodley & Meredith, 2012). Tal y como hemos mencionado anteriormente, Facebook® es una red social compuesta por recursos innovadores que permiten presentar materiales multimedia y ofrecer un espacio de diálogo virtual. Son numerosos los estudios que abogan por su integración en la educación para promover la participación y la comunicación entre los estudiantes permitiendo desarrollar habilidades comunicativas y sociales (Luaran et al., 2014). Las múltiples interacciones propuestas por todos los miembros y la disponibilidad de un canal de comunicación inmediato provocan un incremento en el número de respuestas de cada uno de los miembros del grupo de Facebook® (Kabilan et al., 2010). También se ha observado que el uso de esta infraestructura virtual fortalece el hecho de pertenecer a una comunidad y de forma indirecta favorece la adquisición de conocimientos académicos (Kocdar et al., 2018). Estudios recientes sugieren que el carácter participativo y de interacción que promueve Facebook® puede impactar positivamente en el rendimiento académico (Goh et al., 2019). Sin embargo, poca es la evidencia científica que corrobora experimentalmente que mejora el rendimiento académico con la integración de grupos de Facebook.

Facebook también ofrece ventajas a los docentes, porque además de crear vínculos con los estudiantes y promover la enseñanza, también puede adaptarse a una plataforma de tutoría (Donlan, 2014). No obstante, por lo general el profesorado de la educación terciaria se muestra adverso a incorporar esta plataforma como canal de enseñanza, sosteniendo la posibilidad de

que se trate de una fuente de distracción que reduzca hábitos de estudio (Wolniczak et al., 2013). El tema que presenta más controversia respecto al uso de Facebook® como herramienta educativa es la protección de privacidad y seguridad (Freishtat & Sandlin, 2010). Sin embargo, para pertenecer y participar en un grupo de Facebook® no es obligatorio “ser amigo” de ninguno de los integrantes. A la luz de las desventajas de Facebook, estas cuestiones deben ser estudiadas minuciosamente antes de incorporarlo en el sistema educativo. Sin embargo, se requieren más estudios que examinen experimentalmente las respuestas de interacción y participación y la posibilidad de mejora del rendimiento académico.

4. COMPATIBILIDAD DE LAS INTERVENCIONES FOCALIZADAS E INTEGRALES EN LAS REDES SOCIALES

En los últimos años se han desarrollado investigaciones dentro del ámbito del análisis aplicado de conducta para abordar la viabilidad y eficacia de integrar aplicaciones tecnológicas para cubrir los servicios de formación e intervención directa, tal y como se ha comentado en el apartado de telesalud. En cambio, a pesar de demostrar el potencial de esta ciencia como una tecnología capaz de integrar sistemáticamente softwares educativos en la enseñanza y de demostrar su aceptabilidad y eficacia, no existe ningún estudio en la actualidad que haya tratado de compatibilizar estrategias conductuales con las funcionalidades que aportan las redes sociales.

La mayoría de jóvenes acceden a diario a diversas aplicaciones de redes sociales, incluso de forma continua, para comprobar si tienen nuevas notificaciones en los dispositivos móviles. Esta conducta ya aprendida y generalizada puede adaptarse en contextos educativos. Es más probable que los jóvenes consulten una red social con fines académicos, que no inicien sesión regularmente en programas con software educativos, como, por ejemplo, Moodle® o Blackboard®.

Existen varias funciones que comparten muchas de las redes sociales y podrían adaptarse en contextos educativos:

- *Privacidad:* Existen redes sociales que permiten crear grupos cerrados con el fin de fomentar la comunicación entre un grupo de personas con intereses comunes, en este caso, entre estudiantes e instructores. Fundar un grupo cerrado e invitar a otros usuarios puede realizarse por cualquier persona que tenga una cuenta en la red social, sin necesidad de que los miembros “sean amigos” entre ellos. Este recurso ofrece un en-

torno profesional y privado, exento de acceder o exponer eventos personales, creando una separación entre el espacio íntimo y de aprendizaje. Los principales usos de los grupos cerrados consisten en plantear preguntas, debatir sobre temas en común e intercambiar ideas y compartir materiales en el foro.

- *Etiquetado*: La función de “etiquetar” a usuarios específicos permite informar a un miembro específico la participación en un debate o bien solicitar una respuesta a una pregunta previamente formulada por otro usuario. Esta acción permite alertar a los instructores sobre las preguntas o comentarios que han realizado miembros del grupo con el fin de resolver sus dudas. El etiquetado otorga la responsabilidad a los estudiantes para participar en los debates.
- *Notificación automática*: Las notificaciones automáticas son mensajes instantáneos que reciben todos los miembros de un grupo cerrado o cualquier usuario que pertenezca a una red social concreta, informando sobre las últimas publicaciones, comentarios o etiquetados, desde el último acceso a la red social. Esta función promueve mantener debates sólidos y facilita la inmediatez de la respuesta debido a la automatización de la acción.
- *Inmediatez*: La mayoría de redes sociales presentan como función central el intercambio de mensajes o archivos casi a tiempo real. En el caso de plataformas que permiten la creación de grupos cerrados, cualquier publicación, comentario o “etiquetado” genera una notificación automática que informa a todos los miembros del grupo o al receptor etiquetado en forma de mensaje al dispositivo móvil o email. Esta función de inmediatez facilita iniciar y mantener debates síncronos, independientemente del espacio y del tiempo. Karl y Peluchette (2011) concluyeron que era más rápido utilizar la red social de Facebook para comunicarse con los estudiantes, en lugar de intercambiar correos electrónicos.
- *Colaboración con iguales*: Las redes sociales permiten crear vínculos y socializar con usuarios que comparten intereses y rompen las brechas del tiempo y el lugar. Por lo general, las plataformas web que disponen de recursos para crear grupos cerrados promueven el “trabajo en grupo”, construyendo un espacio activo y compartido con otros iguales que fomenta la comunicación y el apoyo entre ellos. Kabilan et al. (2010) afirma que la creación de relaciones activas entre estudiantes, de edades e intereses similares, incrementa la participación, el compromiso y la motivación por la tarea.

Durante el periodo de aprendizaje y enseñanza, la retroalimentación entre pares es crucial (Petrovic et al., 2014).

- *Compartir recursos:* Algunas redes sociales tienen la capacidad de difundir recursos con imágenes, textos, videos o enlaces a páginas web. Esta función permite a todos los usuarios, independientemente de pertenecer a un grupo cerrado, publicar materiales creando un espacio online ofreciendo las mismas oportunidades para los estudiantes y los instructores. En general, los softwares educativos solo permiten a los instructores compartir materiales con el resto de estudiantes, limitando dicha opción a los estudiantes.
- *Debates síncronos y asíncronos:* A diferencia de los contextos educativos que finalizan los debates entre instructor y estudiantes cuando termina el periodo de clase, los coloquios pueden mediante cualquier red social que suscite a ello. La marca del tiempo la imponen los propios usuarios o miembros de un grupo privado, fomentando así las habilidades de pensamiento crítico (Larson y Sung, 2009). Comúnmente, las redes sociales posibilitan que los usuarios publiquen y comenten para que los receptores lo lean y respondan más tarde, dando lugar a un debate asíncrono en línea. Esta modalidad, facilita que los debates sean más ricos y se mantengan semanas o meses más tarde; ofreciendo la posibilidad de adaptarse al ritmo de cada estudiante y aportar conocimientos anexos después de haber tenido experiencias que conectan con el tema (Bernand et al., 2009). También, existe la modalidad de comunicarse de forma síncrona, gracias a la mensajería instantánea y las notificaciones automáticas que generan ciertas redes sociales. En este caso, los usuarios envían mensajes y los receptores responden de forma inmediata o casi inmediata. Por lo general, esta modalidad es común en redes sociales que tienen la opción de organizar grupos cerrados.

4.1 Intervenciones educativas integrales

A continuación, se presenta una propuesta de cómo las intervenciones educativas integrales podrían adaptarse a una red social utilizando las herramientas comentadas anteriormente (privacidad, etiquetado, inmediatez, contacto con iguales, recursos y modalidad de debate). Al final del capítulo se presenta una tabla que resume las características esenciales de cada intervención y su adaptación en función de los recursos que ofrecen las redes sociales (véase Tabla 2).

4.1.1 Enseñanza de precisión

La enseñanza de precisión es un sistema de medición que facilita a los profesionales discernir objetivamente si el procedimiento aplicado en un aula es eficaz. Está diseñado para mejorar la fluidez y la precisión de habilidades básicas como la lectura, escritura, matemáticas y el habla. Por lo general, los ensayos de enseñanza son intensivos, repetitivos, flexibles y breves; aproximadamente de diez minutos. Las sesiones interactivas pueden desarrollarse individualmente o con un grupo reducido de estudiantes. La transferencia del sistema de enseñanza de precisión a plataformas de redes sociales puede ser útil para reforzar las habilidades de escritura o matemáticas de un grupo reducido de estudiantes. Por este motivo, es importante escoger una red social que permita crear un grupo privado (por ejemplo, Facebook). Antes de comenzar con la dinámica de instrucción, el educador debe concretar una hora con el alumnado para iniciar los ensayos de enseñanza que requieren mantener un debate síncrono. También, el instructor puede utilizar el recurso de etiquetado antes de iniciar la actividad, con el fin de avisar a los estudiantes que está en línea. El tiempo de un ensayo estipulado es de un minuto y el material que ha de presentar en cada ensayo puede adaptarse empleando imágenes y publicándolas en el muro del grupo cerrado. Una vez publicadas las imágenes (estímulos discriminativos) los estudiantes responden en los comentarios del post. Por ejemplo, si el instructor sube una imagen de un sombrero con la pregunta textual “¿qué es?”, los alumnos escribirán en comentarios “sombrero”. En cambio, si el educador presenta una operación matemática “ $30 \times 7 = ?$ ”, los estudiantes escribirán el resultado en los comentarios “210”. Una vez recibidos todos los comentarios de cada estudiante, el educador escribe textualmente la respuesta correcta y ofrece retroalimentación al grupo en general. Por lo tanto, la retroalimentación inmediata del instructor tras recibir los comentarios, es un factor crucial en la enseñanza de precisión y es compatible con las herramientas de mensajería que disponen las redes sociales. Generalmente, las sesiones están formadas por un cómputo de diez ensayos por respuesta objetivo. Al finalizar la sesión, cada estudiante registra el número de respuestas correctas e incorrectas que ha realizado en cada ensayo y grafican el resultado más exitoso, es decir, el más cercano al criterio de adquisición. Al utilizar una red social como canal de enseñanza, el criterio de adquisición ha de modificarse debido a los segundos de latencia que existe entre el tiempo que transcurre desde el envío de la publicación y la aparición en el dispositivo electrónico del estudiante.

4.1.2 Instrucción Directa

La instrucción directa es un método de enseñanza que exige un rol activo del educador. Su actuación consiste en presentar lecciones bien estructuradas, explícitas, claras y adaptadas a sus necesidades. Para lograrlo, existe un conjunto de pasos a seguir: (1) captar la atención de los estudiantes, (2) comprobar los conocimientos adquiridos relacionados con la lección, (3) presentar los guiones de enseñanza, (4) respuesta coral, (5) retroalimentación (práctica guiada), (6) evaluación formal del contenido enseñado en la lección y (7) evaluación formal del aprendizaje emergente (práctica distribuida). La explicación detallada de cada componente se menciona en el capítulo anterior. Comúnmente, los grupos de enseñanza son reducidos, entre diez y quince estudiantes. Cada uno emite diez respuestas activas por minuto en una lección. Llevar a cabo el sistema de instrucción directa utilizando como canal de enseñanza una plataforma de red social, es posible si se modifican los criterios de adquisición teniendo en cuenta el periodo de latencia debido al uso de aparatos tecnológicos. El educador puede utilizar una red social que permita crear un grupo cerrado con los quince estudiantes, como máximo. El desarrollo de una lección requiere la respuesta activa de los miembros del grupo, esto solo sería posible si se realiza en modalidad síncrona, es decir, a tiempo real. Así pues, es necesario utilizar una red social que permita escribir y enviar mensajes o comentarios de forma instantánea. Para lograr que todos los estudiantes participen en la dinámica, el instructor puede concretar un día y una hora de inicio y publicarla con antelación. También, puede avisar a todos los miembros empleando el recurso de etiquetado en la misma publicación. De esta forma, recibirán una notificación automática a sus dispositivos móviles o al correo electrónico personal. Una vez iniciada la lección, los tres primeros pasos pueden realizarse dentro del grupo privado utilizando los recursos de publicación de material (imágenes, documentos, videos, etc.) dependiendo de la lección seleccionada. También, un aspecto importante a tener en cuenta respecto al tercer y cuarto componente, es que la respuesta coral y la retroalimentación ha de ser textual y no oral. Por ejemplo, el educador puede publicar un problema matemático utilizando un formato escrito, los estudiantes envían su respuesta escrita utilizando los comentarios de la propia publicación original y después de recibir todas las respuestas, el instructor ofrece feedback individual, tanto a los que han acertado como a los que han fallado. Este procedimiento ha de realizarse en cada uno de los guiones o estímulos discriminativos que presente el instructor. De este modo, los estudiantes al estar conectados a tiempo real en la misma plataforma de red social, reciben una retroalimentación instantánea en cada uno de los guiones presentados.

4.1.3 Sistema personalizado de instrucción

El sistema personalizado de instrucción es un sistema de enseñanza empleado en contextos educativos, desde infantil hasta la educación superior. Se diseñó para adaptarse al ritmo y a las necesidades de cada estudiante enfatizando cinco características: (a) la comunicación escrita, (b) el ritmo del estudiante, (c) el material de estudio, (d) la figura del supervisor y (e) el uso de conferencias motivacionales. Todos estos componentes pueden proporcionarse y adaptarse haciendo uso de las redes sociales, en concreto, de aquellas que permitan (a) crear grupos privados, (b) que tengan herramientas de compartir recursos de cualquier formato (textual, visual, documentos, etc.), (c) que permitan entablar conversaciones (textuales u orales) con los supervisores y (d) que dispongan de envío de mensajes instantáneos; para así poder proporcionar retroalimentación. Por ejemplo, el educador crea un grupo cerrado y privado con los estudiantes del curso y les comparte la guía de estudio con los objetivos del curso y un documento con actividades y preguntas de reflexión crítica. A partir de este momento, el alumno puede descargarse el material que ha publicado el instructor, planificar su tiempo en relación a las fechas de evaluación del temario e ir estudiando el contenido. Esta estrategia fomenta que el estudiante administre su propio ritmo. Durante el tiempo de estudio, el alumnado puede utilizar el grupo privado para preguntar, tanto a los instructores como al resto de compañeros, cualquier duda que tengan referente al contenido. Cualquier miembro del grupo tiene la posibilidad de crear publicaciones y etiquetar al resto de miembros para avisarles que hay una publicación que requiere su atención. La resolución de dudas puede suceder a tiempo real (síncrono) o bien en formato asíncrono; es decir, cuando el receptor lea la pregunta y disponga de tiempo para responder. Por otro lado, el instructor puede presentar las preguntas de evaluación del contenido en el grupo privado y aquellos estudiantes con rol de supervisor, por haber superado la materia en concreto, se encargan del desempeño de algunos estudiantes, registrando las respuestas correctas e incorrectas y ofreciendo feedback inmediato. Para garantizar que la retroalimentación se ofrece inmediatamente, ambos estudiantes (estudiante-supervisor y estudiante-aprendiz) han de permanecer en línea al mismo tiempo.

4.1.4 CABAS

El enfoque CABAS integra los fundamentos de diferentes sistemas de enseñanza basados en la evidencia, entre ellos, los comentados anteriormente (enseñanza de precisión, instrucción directa y sistema personalizado de enseñanza). La adaptación de este sistema de enseñanza a una red social dependerá de las unidades de aprendizaje a enseñar y de las habilidades del alumnado,

teniendo en cuenta que se trata de un sistema educativo dirigido a personas con discapacidad, diagnosticadas con TEA y problemas severos de conducta. No obstante, sería posible la enseñanza o mejora de habilidades relacionadas con las áreas de escritura, matemáticas o conversacionales. En caso de realizar una sesión de enseñanza CABAS mediante las redes sociales, los pasos a seguir podrían ser los siguientes: (a) el instructor crea un grupo privado con los estudiantes, (b) les comunica la dinámica y les convoca un día y una hora determinada para realizar los ensayos en tiempo real (modalidad síncrona), (c) presenta unidades de aprendizaje a través de publicaciones con imágenes o textuales, (d) avisa a los miembros del grupo etiquetándolos, (e) el alumno responde textualmente al unísono, (f) el instructor entrega retroalimentación inmediatamente después de cada respuesta coral.

4.1.5 Interteaching

Interteaching es un enfoque de enseñanza conductual integral formado por componentes del análisis aplicado de conducta, como: tutoría con iguales, retroalimentación inmediata, aprendizaje cooperativo y algunos componentes esenciales de la instrucción programada, de la enseñanza de precisión y del sistema de instrucción personalizada. Recientemente, un estudio realizado por Rieken et al. (2018) adaptaron la estructura típica de una sesión de interteaching empleando un sistema de e-learning, en concreto, con Blackboard®. El objetivo del estudio era comparar este nuevo sistema de enseñanza con las estrategias convencionales. Los resultados apoyaron su eficacia; ya que los estudiantes que recibieron sesiones de interteaching obtuvieron mejores resultados en las evaluaciones, en comparación del resto. No obstante, hasta la fecha no existen estudios que hayan adaptado este paquete de intervención conductual en las plataformas de redes sociales.

Primeramente, el educador debería seleccionar una red social que le permita crear un grupo privado con los estudiantes que van a recibir este nuevo enfoque (por ejemplo, los grupos de Facebook). Antes de comenzar las sesiones, el educador debería de subir al grupo cerrado un documento que explicara la estructura de la dinámica, la necesidad de trabajar en parejas y la adjudicación de cada una de ellas. Una vez resuelto el tema de las tutorías con iguales, el instructor inicia la sesión enviando al grupo un documento descargable con la guía de preparación, es decir, con el número de preguntas teóricas que los alumnos han de responder mientras debaten con su respectiva pareja de estudio. La adaptación del componente de tutoría con iguales en un

grupo de Facebook requiere transferir los debates orales a un medio textual. Dentro de un grupo cerrado es posible llevar a cabo debates a tiempo real.

Cada pareja tiene una hora asignada para debatir en el grupo, así que el educador etiqueta a cada pareja, es decir, crea una publicación con los nombres de los miembros de la pareja funcionando para los estudiantes como estímulo discriminativo para iniciar el debate. Mientras se desarrolla el debate, el educador está pendiente de la conversación con el fin de ofrecer retroalimentación positiva y correctiva. El mero suceso que ambos roles (educador-alumnado) estén debatiendo en modalidad síncrona, comporta que la retroalimentación se reciba de forma inmediata.

Al final de la sesión, el educador publica en el grupo un formulario de retroalimentación para que los estudiantes anoten los puntos fuertes y débiles de la sesión y del contenido académico. Esta estructura serviría de ejemplo para el resto de parejas de estudiantes; que deberían de realizar la dinámica el mismo día. Este hecho, requeriría mucho tiempo del educador; además del que necesita para preparar el material y las sesiones de conferencias aclaratorias.

Para integrar interteaching en un grupo de Facebook sería conveniente adaptar y modificar el componente de tutoría con iguales y sustituirlo por aprendizaje cooperativo, con el fin de reducir el tiempo activo del educador y poder así dedicarlo a otras tareas académicas, también importantes.

4.2 Intervenciones educativas focalizadas

A continuación, se expone cómo las intervenciones educativas focalizadas pueden adaptarse a una red social utilizando las herramientas comentadas anteriormente (privacidad, etiquetado, inmediatez, contacto con iguales, recursos y modalidad de debate). Al final del capítulo se presenta una tabla que resume las características esenciales de cada intervención y su adaptación en función de los recursos que ofrecen las redes sociales (véase Tabla 2).

Las estrategias de autogestión, aprendizaje cooperativo y reforzamiento formarán parte de la intervención multicomponente (BE-Social) creada y aplicada en la sección de estudios experimentales-aplicados.

4.2.1 Gestión del aula

Los maestros que aplican estrategias de gestión de aula podrían transferir y simular un aula utilizando redes sociales que permitan crear espacios privados, es decir, grupos cerrados (por ejemplo, grupos de Facebook) con estudiantes universitarios o de posgrado. Para crear un ritmo de enseñanza dinámico es útil emplear la respuesta coral, técnica empleada en la instrucción directa. Para adaptarla en un medio online sin posibilidad de hacer videoconferencia con todos los estudiantes, ni que respondan verbalmente todos al unísono, puede utilizar un formato de respuesta textual. Es decir, el instructor escribe o realiza un video explicando que va a estar en línea durante un tiempo específico y va a realizar preguntas en formato escrito sobre el contenido académico para que respondan una vez visualicen la publicación. Para que funcione esta dinámica tanto instructor como estudiantes deben estar conectados en la plataforma de la red social y así poder realizar un debate síncrono, es decir, en tiempo real. Para aumentar la participación, los instructores pueden llevar a cabo la estrategia de tasa de oportunidades de respuesta. Consiste en proporcionar diferentes oportunidades de respuesta utilizando contenidos ya adquiridos o previamente consolidados. Así pues, una vez los estudiantes reciben una notificación con la publicación nueva del instructor (estímulo discriminativo), pueden comentarla a fin de enviar sus respuestas.

El resultado final consiste en visualizar una publicación con un comentario por estudiante. En general, todas las respuestas deberían de coincidir, en caso contrario, el instructor puede observar con facilidad qué el estudiante se ha equivocado. Este tipo de dinámicas se realizan en grupos con iguales, es decir, con alumnos con niveles académicos y edades similares. El factor que fortalece y aligera las actividades propuestas es la retroalimentación inmediata del instructor sobre el rendimiento de los estudiantes, añadiendo al final de la publicación un comentario con la respuesta correcta. En el caso de los grupos de Facebook, el educador tiene la opción de acceder a cualquier publicación y bloquear la opción de recibir más comentarios, dejando así la retroalimentación con la respuesta correcta del instructor como comentario final. Este recurso facilita a los estudiantes acceder en cualquier momento y observar la pregunta propuesta y la respuesta correcta, siendo una herramienta útil de repaso.

Otra de las tácticas que actúan en los antecedentes y forman parte de las estrategias de gestión de aula, es la posibilidad de ofrecer opciones de tareas. La adaptación a un grupo cerrado virtual con estudiantes es fácil de llevar a cabo; ya que el instructor puede crear una publicación

ofreciendo una variedad de dinámicas o tareas pidiendo a los estudiantes que escriban en los comentarios su preferencia.

4.2.2 Aprendizaje cooperativo

La aplicación de la táctica de aprendizaje cooperativo desde una perspectiva conductual se traduce en poner en práctica la contingencia grupal interdependiente, tal y como se ha explicado en el capítulo anterior. En este caso, el educador puede utilizar cualquier red social que permita crear un grupo privado con los estudiantes que compartan habilidades académicas y edades similares. Adaptar una contingencia grupal interdependiente en un grupo cerrado puede ser fácil con las funciones de las redes sociales de “etiquetado”, “compartir recursos” y “respuestas instantáneas”. Para que todo el grupo tenga acceso a un reforzador o recompensa, cada uno de los miembros ha de alcanzar un criterio propuesto por el instructor. Este tipo de contingencia fomenta el trabajo en equipo y compartir recursos; ya que la adquisición del criterio de un compañero predispone a cada uno de los estudiantes a estar más cerca de obtener el reforzador. Dentro de un grupo cerrado los estudiantes pueden publicar y compartir recursos textuales, visuales (imágenes y videos cortos) y enlaces de otras páginas web con el fin de lograr el objetivo individual y común. En ocasiones, los estudiantes necesitan debatir sobre el tema de la tarea a entregar y pueden utilizar la función de “etiquetar” al resto de compañeros o a una persona concreta. Cada estudiante etiquetado, recibe una notificación automática en su dispositivo móvil; en caso de tener la red social descargada desde una aplicación o bien desde un correo electrónico. Este recurso facilita que se cree un debate síncrono en formato escrito, dónde de forma ilimitada pueden participar todos los miembros del grupo a la vez, sin restricciones de número de comentarios. De ser así, obtener respuestas inmediatas o casi inmediatas de los iguales promueve la interacción y participación activa y enriquece el coloquio. No obstante, aquellos miembros que no hayan participado en el debate síncrono, pueden incorporarse más tarde aportando su contribución. Por lo tanto, poner en práctica la técnica de aprendizaje cooperativo no limita a los estudiantes estar en tiempo real activos, sino que promueve la cooperación y la interacción respetando el ritmo de cada alumno.

4.2.3 Tutoría con iguales

La tutoría entre iguales es una táctica de enseñanza que facilita las habilidades académicas entre dos estudiantes, uno de ellos ejerce el rol de tutor, es decir, se encarga de ofrecer breves explicaciones del contenido (formato textual u oral), evaluar la comprensión mediante una serie de preguntas y ofrecer retroalimentación sobre el rendimiento del estudiante. Cuando esta estrate-

gia se aplica con un grupo de estudiantes, recibe el nombre de tutoría con iguales. Estas tutorías consisten en emparejar a todos los estudiantes, el instructor entrega las unidades de aprendizaje que han de enseñar los estudiantes con rol de tutor y supervisa que la dinámica se desarrolle correctamente.

Es posible transferir esta estrategia a una plataforma de red social que permita crear un grupo privado con todos los estudiantes del curso. Para que funcione correctamente, sería necesario que el educador creara previamente las parejas y organizara el orden de conexión de cada pareja, especificando la franja horaria, el nombre del estudiante que actúa como “tutor” y las unidades de aprendizaje que ha de enseñar a su igual. De esta forma, a una hora específica se conectarían dos estudiantes, al grupo privado y podrían empezar la lección en tiempo real (modalidad síncrona).

El estudiante que ejerce el rol de tutor ha de crear oportunidades de respuesta, en este medio, podría desarrollar publicaciones textuales o visuales (imágenes o videos) tratando de explicar o generar un debate sobre la unidad de aprendizaje adjudicada por el instructor. También, el tutor ha de generar respuestas activas del estudiante-aprendiz y para mantenerlo alerta puede utilizar la función de “etiquetado”; ya que genera y envía una notificación automática al receptor. Para completar la unidad de aprendizaje del tutor, es necesario que ofrezca retroalimentación inmediata; comentando cada una de las respuestas del estudiante-aprendiz. Tanto el etiquetado como responder comentarios generan notificaciones que llegan inmediatamente al dispositivo electrónico del receptor. Si ambos estudiantes permanecen activos en una franja horaria determinada existe más probabilidad de recibir las notificaciones automáticas del receptor y crear una conversación (textual) casi en tiempo real.

4.2.4 Economía de fichas

La economía de fichas es un sistema que permite gestionar las contingencias en contextos educativos y mejora las habilidades académicas. Durante su desarrollo, los participantes obtienen reforzadores condicionados (fichas o puntos) por emitir conductas apropiadas, para luego, intercambiarlos por reforzadores previamente pactados (objetos, actividades o atención). En las primeras fases de aplicación del sistema, la inmediatez es un factor crucial para que el estudiante comprenda cómo funciona el acceso a reforzadores. Por ejemplo, durante los primeros ensayos es importante que el instructor ofrezca contingentemente después de la emisión de una respuesta objetivo del estudiante, un reforzador condicionado (ficha o punto). También deberá de

ofrecer un intercambio por las fichas ganadas, y entregar el reforzador pactado con anterioridad. Sin embargo, a medida que el estudiante adquiere una historia de reforzamiento con la economía de fichas, el educador puede aligerar los programas de reforzamiento y anunciar el recuento total de puntos o fichas al final del día, en días alternos o a final de semana. Aquellos alumnos que reúnan las capacidades para emitir las respuestas objetivos, día tras día, sin conocer la cantidad de puntos acumulados de forma inmediata ni acceder al reforzador final, podrían formar parte de una formación online que integrara el sistema de economía de fichas en una plataforma de redes sociales.

La adaptación de este sistema conductual de gestión de contingencias podría emplearse en una red social que permita crear un grupo cerrado de estudiantes, por ejemplo, los grupos de Facebook. Para facilitar la tarea del instructor sobre el recuento de conductas objetivo por estudiante, sería conveniente limitar el número de miembros (por ejemplo, diez). Normalmente, los sistemas de economía de fichas empleados en contextos educativos se utilizan para fomentar conductas adaptativas (por ejemplo, permanecer bien sentado, estar en silencio mientras se atiende a una explicación, tomar notas, etc.) o bien para promover conductas de trabajo grupal (por ejemplo, compartir materiales de interés, debatir con los compañeros, etc.). En este caso, el instructor podría seleccionar las conductas objetivo y adaptarlas a los recursos que ofrece la red social. Por ejemplo, en el caso de los grupos de Facebook, las conductas objetivo que permiten promover respuestas de trabajo cooperativo son: etiquetar, responder inmediatamente a los compañeros, compartir recursos académicos con el resto de miembros e iniciar y mantener un debate utilizando la creación de publicaciones o respondiendo con comentarios (véase ejemplo en la tabla 4.1).

Introducir la economía de fichas dentro de una actividad grupal debería comenzar por (a) la explicación detallada del instructor a los estudiantes sobre el contenido académico que cubrir, (b) la metodología conductual que se pretende aplicar, (c) la selección de las repuestas de cooperación que van a ser observadas y registradas, (d) la puntuación total mínima que se requiere para adquirir un reforzador y (e) el menú de reforzadores disponibles que pueden intercambiar (por ejemplo, reducir el tiempo de clase o el número de tareas a la semana, incorporar temas o materiales más atractivos y visuales, etc.) (Véase ejemplo en la tabla 4.1).

Así pues, el educador puede registrar el número total de ocurrencias diarias o semanales de la conducta de cooperar: (a) etiquetar a un compañero o igual, (b) responder inmediatamente

(tiempo entre respuesta de 2 minutos) a una petición o comentario, (c) compartir información y recursos académicos, (d) participar o iniciar un debate al día o a la semana.

Una vez el instructor ha recolectado los datos sobre las respuestas de cooperación de cada estudiante, obtiene la frecuencia total (diaria o semanal) y evalúa si ha alcanzado el criterio de reforzamiento. En caso afirmativo, puede anunciar los resultados enviando una imagen al grupo privado, del número total de puntos recopilados por cada estudiante y de la disponibilidad del reforzador.

En el caso de ejemplo de la tabla 4.1, todos los estudiantes obtienen una puntuación igual o superior al criterio de acceso a los reforzadores, previamente pactados con el instructor.

Tabla 4.1 Aplicación del sistema de economía de fichas integrada en una red social para mejorar la conducta de trabajo en grupo.

	Etiquetar	Inmediatez	Compartir recursos	Iniciar o participar en debates	Puntuación total	Criterio de acceso al R+
Estudiante 1	IIII	IIII	III	III	15	15
Estudiante 2	IIII	III	IIIII	II	16	15
Estudiante 3	IIII	IIIII	IIII	IIIII	18	15
Estudiante 4	III	IIIII	IIIII	IIIII	19	15
Estudiante 5	II	IIII	IIII	IIIII	15	15
Estudiante 6	III	IIII	IIIII	IIIIII	20	15

4.2.5 Videmodelado

El videomodelado es una táctica de enseñanza que requiere de dispositivos que generen videos, por lo tanto, puede resultar sencillo integrarla en una plataforma de redes sociales. Por lo general, las redes sociales permiten publicar videos; algunas tienen la opción de realizar un video en directo o bien, publicar un video previamente grabado (Instagram®, Facebook®, YouTube®, etc.)

Normalmente, esta estrategia es útil en la enseñanza de habilidades específicas, dónde un profesional, un experto o una persona de edad similar al aprendiz, se graba realizando correc-

tamente una conducta concreta, para luego mostrar el clip del video al estudiante y así, reproducir las respuestas modelo que visualiza. Por este motivo, sería conveniente escoger una red social que permita crear un grupo cerrado, solo para los instructores y aprendices, y además tenga la posibilidad de publicar videos previamente grabados. El procedimiento de videomodelado es fácilmente transferible a las redes sociales.

4.2.6 Entrenamiento en habilidades conductuales

El entrenamiento en habilidades conductuales es una intervención basada en los principios del análisis de conducta, compuesta por cuatro componentes: instrucciones, modelado, ensayo y retroalimentación. El EHC se focaliza en la mejora de habilidades comunicativas, sociales y de gestión de autonomía personal en personas con discapacidad (Boutain et al., 2020), pero también ha demostrado ser una estrategia eficaz para formar a profesionales en la enseñanza. Recientemente, a raíz de la pandemia mundial (COVID-19) se han publicado numerosos estudios que han empleado esta intervención utilizando dispositivos que permiten realizar videoconferencias, conversaciones telefónicas o mantener una comunicación asíncrona (correos electrónicos). Aunque, hasta la fecha se desconocen estudios que apliquen cualquier plataforma de red social para la formación de profesionales utilizando el EHC.

Es posible adaptar los cuatro componentes del EHC a una red social que permita mantener una comunicación fluida con el instructor y los aprendices, por este motivo, el uso de un grupo privado sería útil. El educador puede publicar las instrucciones (primer componente) de forma escrita como, por ejemplo, mediante una publicación o un documento compartido, o también podría ser verbal vocal, grabando una explicación en un video para publicarlo en el grupo privado. Con el fin de alertar a los aprendices, ya pueden ser estudiantes, terapeutas o padres, es posible utilizar la función de etiquetado para que genere una notificación automática y reciban una actualización en sus dispositivos electrónicos. El siguiente componente, el modelado, puede realizarse a través de un video, por lo tanto, se trataría de aplicar el procedimiento de videomodelado. Este paso implica que el instructor se graba a sí mismo mostrando los pasos correctos y necesarios para llevar a cabo un procedimiento. Así pues, el video generado puede publicarse en el grupo privado. Los dos últimos pasos (ensayo y retroalimentación) implica que los aprendices practican las habilidades presentadas en las instrucciones y el videomodelado. Estos dos componentes son esenciales para evaluar la adquisición de la habilidad. Si los aprendices pueden grabarse haciendo role play con algún miembro de su familia, diferente a la persona a quién han de aplicar el procedimiento,

pueden generar un video y subirlo al grupo cerrado de la red social. En cambio, si el aprendiz no puede realizar un role play, podría grabar su ejecución, directamente con la persona que necesita la intervención, y llevar a cabo el procedimiento; para luego subirlo al grupo cerrado. En cualquiera de las dos opciones el aprendiz genera un video mostrando su empeño; para luego enviarlo al instructor a evaluar. Cuando los aprendices demuestren la ejecución de la habilidad correctamente en tres videos consecutivos, el instructor puede enseñar una destreza nueva.

Básicamente, este procedimiento requiere la publicación de videos y la entrega de retroalimentación, por esta razón, tanto el debate síncrono o asíncrono entre miembros del grupo pueden resultar útiles. En este procedimiento, la inmediatez no es un factor imprescindible, no obstante, si el instructor envía el feedback inmediatamente, el ritmo de aprendizaje será más fluido.

4.2.7 Estrategias de autogestión

Las estrategias de autogestión pretenden dotar a los estudiantes de un conjunto de habilidades que fomenten mayor independencia en ámbitos educativos, sociales y vocacionales. Están compuestas por cuatro componentes: automonitorización, autoevaluación, identificación y definición de conductas específicas y autorreforzamiento. Todos estos, requieren que el propio estudiante ejerza un rol activo. La automonitorización se produce cuando el estudiante observa y registra su propia conducta, las habilidades de autoevaluación permiten reajustar sus propias conductas para alcanzar los objetivos, previamente seleccionados. Finalmente, después de completar o cumplir con los criterios, el estudiante se recompensa a sí mismo.

La transferencia de la enseñanza de todas las tácticas a los estudiantes es fácilmente adaptable a una red social que permita crear un grupo cerrado con los estudiantes en cuestión y que disponga la opción de subir videos y entablar conversaciones con los usuarios del grupo. El instructor puede enseñar todos los componentes utilizando el procedimiento de videomodelado, comentado anteriormente. Para ello, la plataforma de red social ha de permitir la subida de videos, tanto del educador como de los estudiantes. El instructor ha de mostrar el modelo exacto de la habilidad a ejecutar por el alumno y éstos han de generar un video mostrando su empeño para que el educador lo evalúe. También, la integración de esta intervención en plataformas de redes sociales puede beneficiarse de las herramientas de etiquetado, para avisar a uno o un conjunto de miembros que revisen una publicación concreta; ya sea textual (retroalimentación) o visual. Uno de los factores que caracteriza la eficacia de esta intervención es la retroalimentación frecuente sobre el desempeño de cada alumno. Así pues, el instructor tiene la opción de comentar

cualquier publicación que contenga un video de un estudiante ofreciendo retroalimentación correctiva. Esta interacción puede suceder en tiempo real o de forma asíncrona sin afectar la eficacia de la intervención.

4.2.8 Instrucción sobre equivalencia

La instrucción basada en la equivalencia mediante el uso de softwares educativos ha tenido éxito en el ámbito académico, tal y como se ha comentado en el capítulo anterior (ver pág. X). Las redes sociales tienden a promover la colaboración y participación de un grupo de personas con intereses afines; posiblemente debido a la automaticidad de algunas de las herramientas. No obstante, la respuesta específica que emite un usuario de una red social, no acostumbra a recibir una consecuencia inmediata relacionada con la respuesta en sí. Por ejemplo, si un usuario escribe un comentario en una publicación, es poco probable que reciba inmediatamente, un nuevo comentario sobre su respuesta. Podría recibir una consecuencia, pero sin necesidad de estar relacionada con la respuesta anterior. Esta descripción de la contingencia concreta (respuesta de un miembro-consecuencia inmediata relacionada con la respuesta) es posible en softwares educativos y no en redes sociales. Para la aplicación eficaz de una intervención basada en la equivalencia, dónde se enseñan emparejamientos a la muestra y relaciones condicionales es imprescindible una “retroalimentación inmediata” al estudiante para que se produzca el fenómeno de aprendizaje. Por lo tanto, no es posible adaptar esta intervención en plataformas de redes sociales.

4.2.9 Retroalimentación

La retroalimentación inmediata favorece la adquisición, tal y como se ha argumentado en el capítulo anterior y que comprobaremos en los estudios experimentales-aplicados que se presentan a lo largo de la tesis. Se considera un factor crucial en la enseñanza, es por este motivo que otras tácticas de aprendizaje incorporan esta estrategia como elemento activo. Por ejemplo, en el entrenamiento en habilidades conductuales, en la aplicación de estrategias autogestión, de videomodelado, economía de fichas, aprendizaje cooperativo, entre muchas otras. Ofrecer retroalimentación mediante una plataforma de red social puede ser sencillo debido a la automaticidad que disponen algunas acciones que generan notificaciones automáticas. En el caso de utilizar la red social de Facebook, el etiquetado, la creación de publicaciones y el envío de comentarios, generan notificaciones instantáneas en forma de alerta, para avisar al resto de miembros que existe una novedad. En general, cualquier red social que permita crear una conversación en tiempo real (modalidad síncrona) podría ser un canal útil para proporcionar retroalimentación inmediata.

5. CONSIDERACIONES FINALES

En este capítulo resaltamos la importancia de integrar las TIC y los recursos de la Web 2.0 en la educación, sobre todo en la educación terciaria. Por ejemplo, el aprendizaje en línea se ha convertido en una necesidad en tiempos de catástrofes naturales (Dabner, 2012) o dónde el distanciamiento social ha sido una obligación, como la pandemia mundial de la COVID-19. El inicio del capítulo expone los tipos de softwares educativos y las características que han de poseer para garantizar una instrucción asistida por ordenador activa y eficaz. También, introduce el uso de la telesalud en el área clínica como canal de enseñanza a profesionales e incluso padres que no tienen acceso a una intervención conductual de calidad debido a barreras como la ubicación geográfica, la falta de recursos o el acceso limitado de profesionales locales. Cabe resaltar que la mayor parte de la investigación sobre telesalud está focalizada en la formación de adultos que están a cargo de niños o jóvenes con discapacidad intelectual o TEA (Ellison et al., 2021). No hemos encontrado literatura analítico-conductual que emplee procedimientos de formación a estudiantes de educación superior. Más adelante, se introducen las redes sociales como fenómeno global integrado en nuestra sociedad y educación; ampliamente utilizado por el sector educativo terciario. Debido a su uso generalizado por los estudiantes, existe numerosas publicaciones que tratan de integrar las redes sociales en la educación posuniversitaria, siendo la plataforma de Facebook, la más común. El capítulo finaliza con una sección que detalla las funcionalidades generales de las redes sociales (por ejemplo, etiquetado, inmediatez, compartir recursos, posibilidad de debatir a tiempo real, etc.) y las características principales de las intervenciones conductuales focalizadas e integrales, revisadas en el capítulo anterior. Su finalidad es aportar una aproximación práctica de cómo compatibilizar los procedimientos derivados de los principios conductuales en las redes sociales.

Tabla 4.2 Funcionalidades Comunes de Redes Sociales y su Valor Práctico en Modelos de Educación Conductual

	Privacidad	Etiquetado	Inmediatez	Iguales	Recursos	Debate
Intervenciones educativas integrales						
Enseñanza de precisión	Grupo reducido de estudiantes	Alertar a los estudiantes	Feedback frecuente de rendimiento	NA	Imágenes	Debate síncrono
Instrucción directa	Grupo reducido de estudiantes (15 máximo)	Alertar a los estudiantes	Feedback frecuente de rendimiento	Respuesta coral (textual)	Material para las lecciones	Debate síncrono
Sistema personalizado de instrucción	Estudiantes que comparten el mismo nivel académico	Agilizar la interacción social con los supervisores	Feedback frecuente de rendimiento	Algunos iguales ejercen el rol de supervisor	Guía de estudio con los objetivos del curso, lecturas complementarias y enlaces a videos.	Debate síncrono y asíncrono
CABAS	Grupo reducido de estudiantes	Alertar a los estudiantes	Feedback frecuente de rendimiento	Respuesta coral (textual)	Imágenes o textos	Debate síncrono
Interteaching	Grupo reducido de estudiantes	Alertar a las parejas de estudio para iniciar el debate	Feedback frecuente de rendimiento	Tutoría con iguales	Publicar la guía de preparación y el formulario de retroalimentación	Debate síncrono
Intervenciones educativas focalizadas						
Gestión del aula	Estudiantes que comparten el mismo nivel académico	La respuesta coral puede ser por escrito	Feedback frecuente de rendimiento	Respuesta coral (textual)	Ofrecer opciones de tareas	Debate síncrono
Aprendizaje Cooperativo	Contingencia grupal interdependiente	Interacción mutua	Fomentar la comunicación	Contingencia grupal interdependiente	Intercambiar recursos	Debate síncrono o asíncrono
Tutoría con iguales	Tutorías entre estudiante-estudiante (rol de tutor)	Respuesta activa del estudiante	Feedback frecuente en cada unidad de aprendizaje	El rol de tutor lo adquieren los propios estudiantes	Crear oportunidades de respuestas	Debate síncrono
Economía de fichas	Grupo de estudiantes limitado	Cualquier estudiante alerta a otro para solicitar ayuda o compartir información	Cantidad de tiempo que un estudiante responde a otro (TER: 2 min)	Fomentar la motivación	Los estudiantes comparten información o recursos	Debate síncrono o asíncrono

	Privacidad	Etiquetado	Inmediatez	Iguales	Recursos	Debate
Video-modelado (VM)	Grupo cerrado	NA	NA	NA	NA	NA
Entrenamiento de Habilidades Conductuales	Grupo cerrado	Alertar de una nueva publicación relacionada con componentes del EHC	NA	NA	Cuatro componentes del EHC (instrucción, modelado, ensayo y retroalimentación)	Debate síncrono o asíncrono
Estrategias de auto-gestión	Grupo cerrado	Alertar a cualquier miembro del grupo	Feedback frecuente de rendimiento	NA	VM para enseñar auto-monitorización, autoevaluación, identificación y definición de conductas específicas y autorreforzamiento	Debate síncrono o asíncrono
Instrucción basada en la equivalencia	NA	NA	NA	NA	NA	NA
Retroalimentación	No es necesario pertenecer a un grupo privado	Facilita el feedback	Facilita el feedback	NA	Facilita el feedback	Debate síncrono

Notas. DA = Debate asíncrono; DS = Debate síncrono; EHC = Entrenamiento en habilidades conductuales; NA = No aplicable; TER = Tiempo entre respuestas; VM = Videomodelado.

TERCERA PARTE
BASES METODOLÓGICAS

CHAPTER 5. Methodological Basis of Behavior-Analytic Interventions in Education

The intent of the current chapter is to provide an overview of the key methodological standards of behavioral education interventions in general, and the methodological constraints brought about by the social media environment, in particular. The chapter is also intended to provide an introduction and the basic rationale for the methodological analyses, whether conceptual or empirical, that would follow in subsequent chapters of this section of the dissertation.

Numerous sources of variation had been identified in the analysis of changes in academic performance. These may be considered standards to be weighed in the design of experimental studies in the subfield of behavioral interventions in tertiary education. Below, I review some of these standards, which had been critical in the design of the intervention studies included in this dissertation.

1. CHOOSING A MEANINGFUL BEHAVIORAL OUTCOME

Behavior-analytic outcomes in educational interventions have been often criticized for not being amenable to longitudinal analyses (see a recent discussion in Zamzow, 2022). Behavior analysis has a long tradition of directly observable behavioral outcomes, which are often the direct target of intervention. This approach has the potential for limiting the generalizability of findings into complex behavioral domains or broadly-defined performance criteria including academic performance. For example, Fienup et al. (2010) taught a group of college students' stimulus-stimulus relations associated with statistical hypothesis testing. The authors implemented a conditional discrimination training paradigm in order to efficiently teach a complex collection of equivalence relations. Their main outcome focused on accuracy (i.e., percentage of correct responses) in a series of discrimination trials conducted before and after training. While narrowly-focused outcomes have the advantage of establishing the interval validity of the study (as close factual connections can be established between the educational manipulation and student behavior), the results may fall short in terms of their impact in wider behavioral repertoires. In the example of

Fineup et al., it remained unclear what was the impact of the intervention on the students' statistical competence or on their ability to design or interpret a study that utilized inferential statistics.

A relatively recent trend in the field is to combine narrowly defined behavioral targets for the purposes of internal validity and monitoring treatment evaluation, while implementing in parallel more mainstream outcomes, including standardized (e.g., Reynell Developmental Language Scales) or criterion outcomes (e.g., academic achievement, cumulative number of learning targets achieved) (see for example Hanley, 2017). For example, Virues-Ortega et al. (2013) utilized the number of total behavioral targets in a longitudinal analysis of early intensive behavioral intervention for children with autism spectrum disorder. A similar strategy may be utilized for behavior-analytic interventions in tertiary education.

Given the applied nature of interventions in education, we will focus on both online behaviors generating permanent products that can be easily harvested through the social media platform (e.g., reactions, comments, posts, views), and criterion outcomes that would have direct social significance, singularly, academic achievement. The selection of social media generated outcomes will be discussed in detail in Chapter 8 as part of the methodological section of the dissertation. We briefly discuss approaches to the assessment of academic achievement in this context in the following paragraphs.

Content testing via *ad hoc* multiple-choice questionnaires or tests is a common strategy to assess academic performance in educational interventions (Haladyna, 2004). This approach has the advantage of being integrated in instructional programs being a readily available ecologically valid intervention outcome. Potential challenges are the representativeness and relevance of the material (Shin et al., 2019). For example, practical applied psychology courses may be more challenging to assess via content testing. A particular concern in this context is that the items (discriminative stimuli), should focus on real-life scenarios that can effectively emulate the ecological antecedent conditions that are expected to gain control over the "information" or complex verbal responses acquired (Reese, 2017).

In Table 1 and Table 2 we present a content analysis of test items and the complexity of the discriminations that are needed depending on whether the stem includes or not ecological details (context-level information) and distractors, whether content-level distractors or context-level distractors. This cursory analysis reveals some of the discriminative dimensions of multiple-choice items. For example, applied scenarios often include a larger number of key

discriminations, some of which are contextual (e.g., “A boy can take a soft drink from a vending machine”) or require complex inductive reasoning (i.e., stimulus generalization) on the part of the test taker (e.g., “Operating a vending machine is reinforced by access to a favorite soft drink”). As shown in this example, it might be possible to rank the complexity of the final discriminated response to a multiple-choice item weighting the number of content-related discriminations, contextual discriminations, and inductive propositions. Based on this conceptual analysis, and for the purposes of developing multiple-choice items as a means to monitor academic achievement through content testing, we will utilize the following corollaries or standards.

1. A relatively high prevalence of items including content and context information and at least some content level distractors and inductive propositions would minimize the probability of ceiling effects.
2. Items including content and context information and at least some content-level distractors and inductive propositions may be relatively more accurate indicators of actual performance when teaching practical content.
3. Correct responding in multiple-choice items may be a function of the total number of context- and content-level discriminations and inductive propositions embedded within the multiple-choice item text.

2. INTEROBSERVER AGREEMENT

A long-standing standard in behavior-analytic intervention is the assessment of agreement for any outcome recorded by a human observer. The social media channel presents itself with unique challenges and opportunities in this respect. First, digital activity is often recorded as a permanent product, that is, a permanent change in the environment that can be recorded retrospectively by the observer (Kelly, 1976). The quantitative dimensions of social media posts, comments, and reactions (e.g., latency, inter-response time, extension, frequency) leave a permanent digital product that can be subsequently recorded. Ideally, researchers could make use of purposely developed *application programming interfaces* (APIs) that would collect critical behavioral outcomes seamlessly. While this technology is developed (we have not been able to find publicly available functionalities at least for the major social media platforms), low-resource studies may still use traditional behavioral observation considering digital inputs simply as behavior products, which can then be the subject of traditional interobserver agreement analysis.

3. PROCEDURAL INTEGRITY

It is critical for conceptually systematic educational interventions to be delivered exactly as planned. Behavioral procedures often prescribe specific sequence of discrete events (e.g., when A occur, present B within 5 s of A) repeated over arbitrary time units (e.g., blocks of trials, sessions, days, etc.) during intervention protocols. The disruption of the prescribed sequence of events that is the intervention would render meaningless the findings of any treatment evaluation. Therefore, monitoring procedural integrity is critical in behavioral interventions including in the field of education.

Procedural integrity may be defined as the extent to which the independent variable is delivered as intended over the course of the intervention (Fiske, 2008). Unlike other methodological standards that are mostly prescriptive (c.f., range of usable values of interobserver agreement), the diversity of intervention procedures means that procedural integrity strategies are customized for every study. It is up to the researcher to determine which critical aspects of the intervention are to be sampled in order to document procedural integrity, while document all aspects is not often feasible. For example, functional analysis procedures often document the timely delivery of social stimuli (i.e., prompts, social attention, demands) as the elements of its procedural integrity evaluation (Virues-Ortega et al., 2022). For intervention procedures that are common in the applied field (such as functional analysis), researchers often choose to omit these details from the publications, whereas novel interventions (and interventions that rely on procedures delivered by third parties, like teachers and parents) often include a detailed account of procedural integrity (see for example Najdowski et al., 2008). When procedural integrity indicators are recorded through direct behavioral observation, usual interobserver agreement standards apply.

Procedural integrity has only gradually been established as a widely accepted standard in experimental treatment evaluations in applied behavior analysis. However, the surveys available indicate that most behavior-analytic studies do not report procedural integrity even today (Cymbal et al., 2021; Wheeler et al., 2009). This is certainly the case for the wider field of educational psychology, and cognitive educational psychology in particular, where procedural integrity is rarely documented. The common dismissal of procedural integrity in spite its potentially critical impact on internal validity may be an understudied factor in the replication crisis in psychology (Tackett et al., 2019). For example, in a skill acquisition study, Greenberg (2010) showed that a decrease of 50% in procedural integrity in the context of a skill acquisition single-subject study

resulted in a 32% increase in trials to criterion. Its effects in other behavior dimensions have been rarely described.

The social media context imposes various restriction to observation that ought to be considered in the design of procedural integrity strategies. For example, self-management skills training may rely on video posts (whether synchronous or asynchronous). Therefore, operationally defined behavior targets should be monitored over the duration of this broadcasts to verify procedural integrity standards. To this end, the use of specific paper-and-pencil and computer-based observations systems should be validated in this particular setting. The goal of Chapter 7 within the methodological section of the dissertation was intended to respond specifically to this need.

4. INTENSITY OF THE INSTRUCTIONAL ACTION

A critical consideration in the interpretation of experimental findings in the behavioral education subfield is whether the effect of the intervention is to be attributed to the intervention or simply to the social input of the instruction regardless of any other consideration. Attributing a particular behavioral effect to an instructional intervention, when several are being compared, requires equating, to the extent possible, some of the quantitative dimensions of the instruction across experimental groups (for group-based studies) or conditions (for within- and single-subject analyses). In the context of instructional interventions delivered through the social media channel, these might be operationalized as units of instructional action including number of instructor-led posts, duration of video-based posts, extension of textual responses, and the like. When comparing various interventions or when designing a control group or a baseline condition, it bears some importance to document some of these quantitative dimensions to establish the experimental argument that the experimental conditions are generally comparable except for the intended manipulation of the independent variable. Therefore, a void intervention control group or baseline condition may be a suboptimal approach in this context, whereas a reference condition with comparable instructional inputs strengthens internal validity. Therefore, for the purposes of the interventional studies that would follow, we would endeavor to characterize in detail the background educational and social media contexts intended to be constant across experimental conditions.

5. COMBINATION OF GROUP-BASED AND SINGLE-SUBJECT APPROACHES

Randomized control trials have become the currency of treatment evaluation in the health sciences and psychology, and to a lesser extent in evidence-based education practices (Styles & Torgerson, 2018). By contrast, behavior-analytic interventions have rarely utilized RCTs for treatment evaluation until relatively recently, owing to a decades-long tradition of single-subject designs in behavioral interventions. This methodological tradition is in part justified by the time-dependent (continuous) nature of behavioral processes as they interact with the environment (Skinner, 1938). It also allows to demonstrate socially important effects in the single individual and modify the intervention parameters accordingly if the intended effect is not forthcoming, which has obvious clinical advantages. However, from an outcomes research perspective, single-subject designs have been criticized for not being amenable to large-*n* longitudinal evaluations and standardized outcomes, while being subjected to sample attrition and selection biases (see Hagopian, 2020 for a recent approach intended to minimize bias in single-subject designs). Moreover, health decision-makers are generally unaware of single-subject research and in spite of some recent efforts for developing consensual methodological standards (e.g., What Works Clearinghouse, 2017). The single-case experimental design (SCED) literature is often ignored in the development of health and educational policies. This trend is shifting only very gradually. For example, 12 of the 17 (71%) randomized controlled trials and randomized clinical trials published in the *Journal of Applied Behavior Analysis* have been published after 2013.

A distinct challenge of randomization regimes in behavior-analytic interventions is due to their ethical ramifications. Behavioral interventions are often longitudinal and it is not possible to withhold a potentially effective treatment for methodological reasons during critical months or years of a child's development (Narita, 2021). This challenge is less of a concern in educational contexts. However, the time dimension of the intervention in educational settings is still significant. An RCT approach in this context may be ethically possible if the intervention is intended to optimize and already existing educational procedure. Therefore, void baseline and control conditions may not be practically possible (and may not even be desirable as discussed in the section on the *Intensity of the Instructional Procedure*).

We had been inspired by the work of Nathan H. Azrin (1930-2013) in regards to the potential for combining single-subject analysis and RCTs. Azrin appreciated both the wide acceptance of group-based designs beyond the behavior analysis field and the clinical relevance

of SCEDs. He conducted several studies that capitalized on the compatibility of both methodological traditions by combining wait-list controlled RCTs and multiple-baseline designs (see a discussion in Virues-Ortega et al., 2021). Indeed, it is possible to merge RCT and SCEDs given that the intervention delay imposed by the multiple-baseline design (MBL) does not induce any discernable trends that could challenge the researcher’s ability to compare the arms of the RCT. The latter is hardly a practical challenge in that MBL designs rarely imposes long baselines (see recent examples of combining RCTs and single-subject designs in Sivaraman et al., 2022, and Fuentes et al., 2022). Figure 5.1 presents a diagram that summarizes the decision-making process for combining RCTs and SCEDs.

The proposed multi-design approach has two distinctive advantages that may be beneficial to both the single-subject and the group-based evaluations: (1) the control group provides an opportunity to include a set of participants in a “continuous baseline” as part of the MBL design, (2) the MBL, when applied to large groups can utilize a randomized treatment initiation time-point (as opposed to a large effect treatment transition criterion, Bailey & Burch, 2017), (3) transitioning part or all control group participants to a cross-over phase provides an opportunity for delayed baselines that can be integrated as part of the MBL, (4) multiple-arm RCTs can provide the opportunity for component analyses that are not subject to sequence effects, which are difficult to minimize in add-in and drop-out component analyses in SCEDs (Ward-Horner & Sturmey, 2010), and (5) combining SCED and RCT studies provide the opportunity to produce mainstream metrics of treatment effect in parallel to a detailed characterization of subgroup or idiosyncratic treatment effects.

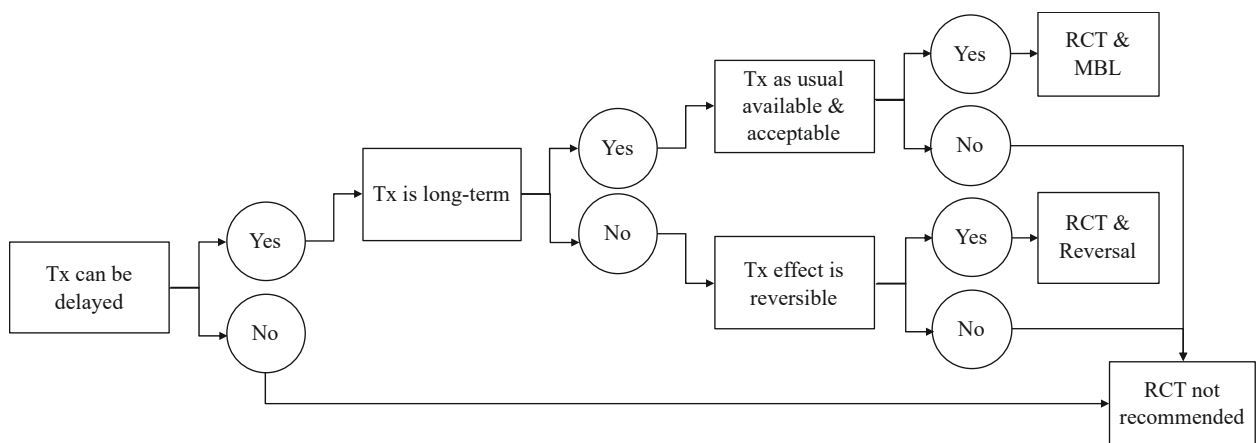


Figure 5.1 A decision tree for the combination a randomized controlled trial (RCT) and a single-case experimental design (SCED).

Notes. MBL = Multiple-baseline design; Tx = Treatment.

6. STATISTICAL APPROACHES TO ANALYZING COMBINED GROUP-BASED AND SINGLE-SUBJECT DESIGNS

The combination of the RCT and SCED approaches to experimentation in applied behavioral education studies may require a degree of statistical sophistication. Traditionally, single-subject analyses have relied on straightforward visual analysis strategies to ascertain treatment effects owing to the large intervention effects focus of applied behavior analysis (Baer et al, 1968). Over the last decades numerous statistical approaches to the analysis of single-subject data have been developed that, in practicality, have offset the historical reservations of the field to statistical analysis (see for example Michael, 1974). For example, hierarchical linear models (HLM) can help to account for types of variability that are specific to single-subject designs such as cross-baseline correlation in MBLs or cross-condition auto-correlation in multi-element designs. For the purposes of fine-tuning our approach to the statistical analysis of the applied studies of the dissertation we have undertaken a thorough review of the existing approaches to the statistical analysis of single-subject datasets (Chapter 6). This analysis will provide the basis for utilizing visual analysis strategies and SCED-specific effect size metric in our first intervention study.

7. METHODOLOGICAL COROLLARIES AND THE DESIGN OF STUDIES WITH SCIENTIFIC RIGOR AND APPLIED FOCUS

The methodological standards briefly described above will serve as guidance for the studies presented in the subsequent chapters, including those with a methodological focus. These standards can be summarized as follows:

1. Carefully designed multiple-choice items may be an ecologically-valid approach to content testing that allows for frequent measurement of the outcome (time-series analysis). The latter is a critical advantage to using multiple-choice items to monitor academic achievement in SCEDs owing to the need for repeated measurements of the outcome.
2. Time-stamped social media outcomes (e.g., reactions, comments, posts) may provide an opportunity to analyze important mediating variables when analyzing the effects of behavioral education interventions on academic achievement. This standard will be elaborated upon in Chapter 8 where we will undertake a systematic review of quan-

titative social media specific behaviors that may be used as outcomes or mediating factors within behavioral intervention studies.

3. Interobserver agreement and procedural integrity standards are highly relevant to behavioral education interventions. In Chapter 7 we will take a closer look to approaches to validating behavioral observation systems for video-based posts that are integral to common procedural integrity strategies in our specific research setting. We will use some of the systems validated in Chapter 7 as part of our behavioral observation procedures in our last intervention study. Moreover, equating the intensity of the instructional activity across groups may be an important standard in behavioral education interventions that should be an integral part of procedural integrity protocols. We will incorporate this latter standard also in our last intervention study.
4. There are distinct advantages to combining SCED and RCT studies according to the scheme proposed in this chapter. In our first intervention study we will illustrate this combination by utilizing both an RCT and a multiple-baseline design.
5. The combination of SCED and RCT strategies call for the utilization of statistical techniques that are specific to either design. As the statistical tools for SCEDs have gained in complexity over the last decades, we present a detailed review of these procedures on Chapter 6. This analysis will provide the basis for our choice of statistical analyses in our first intervention study.
6. A multi-arm RCT provides the opportunity to optimize a multi-component intervention by discerning the relative contribution of specific elements of the intervention. This methodological corollary will be illustrated in our second intervention study.

Table 5.1 *Discrimination Complexity of Various Multiple-Choice Item Types.*

	Discrimination complexity	Stimulus induction
Content-level information only without distractors	Low	Low
Content-level information with content-level distractors	Mid	Low
Content- and context-level information without distractors	Mid	High
Content- and context-level information with content-level distractors	High	High
Content- and context-level information with context-level distractors	High	High
Content- and context-level information with content and context-level distractors	High	High

Table 5.2 Multiple-Choice Items by Type and Selected Underlying Propositions.

Type	Underlying Propositions
Content-level information with content-level distractors	It is true that behaviorism argues that public and private events are subject to the same laws and principles.
<i>Which of the following statements is true regarding radical behaviorism?</i>	It is not true that behaviorism was superseded by methodological behaviorism.
a. <i>It argues that public and private events are subject to the same laws and principles.</i>	It is true that behaviorism argues that private events play an important causal role in human behavior.
b. <i>It was rapidly superseded by methodological behaviorism.</i>	First statement is truer than the third
c. <i>It argues that private events play an important causal role in human behavior.</i>	It is not true that behaviorism limits the scope of science to what can be publicly observed.
d. <i>Limits the scope of a behavioral science to those manifestations that can be publicly observed.</i>	
Content- and context-level information with content-level distractors	A boy can take a soft drink from a vending machine.
<i>A boy wants to take his favorite soft drink out of a vending machine. However, when he gets to the machine, he notices that there is a sign that says "out of order". As a result, the boy leaves without his soft drink. What function does the sign have on the behavior of depositing a coin in the vending machine?</i>	"Out of order" is a stimulus.
a. <i>A discriminative function.</i>	Leaving the vending machine is discriminated by "Out of order"
b. <i>An abolishing operation function.</i>	Approaching the vending machine is discriminated by the vending machine.
c. <i>A punishing function.</i>	Operating a vending machine is reinforced by access to a favorite soft drink.
d. <i>An extinction function.</i>	A vending machine can be operated by depositing a coin.
	"Out of order" is a sign.
	"Out of order" is not a contingent stimulus, therefore, it does not have a punishing function.
	"Out of order" does not alter the reinforcing potency of a soft drink, therefore, it does not have an abolishing function.
	"Out of order" is not a reinforcer that was withdrawn; therefore, it does not induce extinction.

CHAPTER 6. Quantifying Outcomes in Applied Behavior Analysis through Visual and Statistical Analyses: A Synthesis⁴⁶

ABSTRACT

Applied behavior analysis employs visual analysis as the primary means of establishing treatment effects in single-subject experimental designs. In addition, there are several strategies that can be used to evaluate intervention effectiveness such non-overlap indices, mean-based effect sizes, analyses resulting in a p value as their main output, and regression-based analyses. This chapter will review strategies for the visual and statistical analysis of single-case data in applied behavior analysis.

Key-words: applied behavior analysis, ABA, visual analysis, single-case experimental design, effect size, hierarchical linear modeling.

INTRODUCTION

Five adults with intellectual disabilities who attended a day-training center were prescribed with physical exercise. Krentz et al. (2016) aimed to increase walking in these adults. The authors began with a baseline phase in which they recorded the number of laps completed in one hour by each adult. During the treatment phase, the adults earned tokens for each lap completed, and could exchange these tokens for a preferred item at the end of the one-hour session. The researchers observed an increase in the number of laps by each adult during the token reinforcement phase relative to baseline. After a few treatment sessions, baseline conditions were reinstated, i.e., the adults stopped receiving tokens. The authors observed that the laps completed by each adult de-

46 El presente capítulo ha sido publicado en la revista SpringerNature y se presenta en el idioma y el formato en el que se publicó.

Virues-Ortega, J., Moeyaert, M., Sivaraman, M., Tarifa-Rodriguez, & Fernandez Castilla, B. (2022). Quantifying Outcomes in Applied Behavior Analysis through Visual and Statistical Analyses: A Synthesis. In J. L. Matson, *Applied Behavior Analysis: A Comprehensive Handbook*. SpringerNature.

creased until their frequency was comparable to the one observed during the initial baseline. The study was completed with a return to the treatment phase in which participants again received tokens for walking laps. The results showed that walking was low during the baseline phase, increased during the treatment phase, decreased during the return to baseline, and increased once again during the final treatment phase for all participants. The authors concluded that token reinforcement was effective in increasing walking among their adult participants.

The study described above exemplifies a reversal design involving repeated and frequent measures of behavior (i.e., number of laps walked), the experimental manipulation of an independent variable (i.e., introduction and withdrawal of token reinforcement for walking) and replications within and across subjects. Single-case experimental designs (SCED) allow the practitioner to evaluate the effects of an intervention using repeated behavioral measurements across sessions and conditions. This is in sharp contrast to group designs that often focus on the aggregated effect of a treatment on an entire group. Through an SCED, the practitioner can monitor individual performance on a session-by-session basis, potentially drawing cause-effect links between experimental manipulation (i.e., treatment) and behavior (What Works Clearinghouse, 2020). Single-subject designs are particularly well-suited to practitioners and clinicians as they engage in practice-oriented research. The design can accommodate large between-subject variations commonly encountered in practice (e.g., level of baseline performance, time availability, celeration, etc.). In addition, intervention procedures and SCED may be modified on an on-going basis in the event of limited or no improvements in performance (Kratowill & Levin, 2014; Kratowill et al., 2010).

Single-subject designs have been criticized for being subject to selection bias, relying on visual analysis for treatment outcome evaluation, and not accommodating to mainstream inferential statistics.

Historically, there has been no methodological requirement for reporting on all participants initially screened or selected for participation in behavior-analytic studies. Some have criticized that this practice may have led to publication bias, whereby positive results are more likely to be reported and published. This threat was confirmed by Sham and Smith (2014), who observed

that the effect size in a selection of applied studies was on average 20% greater in published relative to unpublished studies. Recent suggestions for extending methodological standards borrowed from the medical sciences, including consecutive client admission (Hagopian, 2020) and intention-to-treat analysis (see for example, Taylor et al., 2019), provide the tools to overcoming patient selection bias in SCEDs.

Relying on visual analysis has also been criticized as an unreliable and pre-quantitative approach to data analysis (e.g., DeProspero & Cohen, 1979). A major justification for visual analysis lies on the effective dimension of the field of applied behavior analysis (Baer et al., 1968). In the hopes of prioritizing socially valid interventions, the field has traditionally focused on interventions producing large-effects that can be evident from the visual inspection of the time series and are less prone to type I errors (false-positives). Moreover, statistics were seen at the time as a distracting armamentarium due to the frequent mismatch between clinical and social significance (see for example Michael, 1974). These factors led to a decades-long rejection of mainstream inferential statistics in applied behavior analysis.

These concerns are now gradually coming to the fore as behavior analysis produces service models that ought to be weighted and prioritized by health and policy decision makers causing a need for quantitative and widely intelligible summaries of scientific evidence. Visual analysis allows a good understanding of emerging patterns in the data, whereas the statistical analysis can help to quantify the magnitude of effects, study co-variables, and synthesize evidence from multiple individuals and studies. In addition, structural visual analyses and visual aids minimize subjectivity and Type-1 errors while visually analyzing SCED datasets.

In this chapter we will briefly review the various approaches to quantifying outcomes in SCED and how new developments in statistics can support the process of *mainstreaming* SCEDs into the outcome research literature by combining visual and statistical analyses. In the next sections we will describe techniques to evaluate SCEDs using visual and statistical analyses, and also outline the potential synergies between the two for the purposes of quantifying, comparing, and pooling treatment effects. We recommend that the reader is familiarized with Chapters 24

(Single-Case Designs), 29 (Treatment Integrity), and 30 (Social Validation). Concepts from these sections will be used throughout the current chapter.

VISUAL ANALYSIS

How Does it Work?

Session-by-session SCEDs graphs provide the most common basis for conducting a visual analysis. While there are numerous graph formats in applied behavior analysis (e.g., line graphs, bar graphs, cumulative record, scatter plot), the line graph is predominant in applied behavior analysis for displaying non-cumulative behavior recordings over time. In a line graph the y-axis represents the dependent variable (i.e., the behavior of interest) and the x-axis typically represents time (i.e., sessions, days). Letters in alphabetical order are often used to denote the various conditions in an SCED and its graphic display. For example, an ABAB design represents two A conditions and two B conditions. A typically refers to the baseline condition, whereas B, C... denote successive treatment conditions. A transition from one condition to another is represented using a phase change line on a graph (see a set of guidelines for graph construction to minimize bias in Cooper et al., 2019; Kubina et al., 2017; see also Chapter 24).

Data graphing and visual analysis go hand-in-hand throughout a behavior change project and they are considered essential processes for both on-going clinical decision-making and evaluation of the intervention effect. Specifically, a major aim of a visual analysis is to identify whether there was a change in the behavior of interest and whether that change was the reliable and consistent result of the intervention. Intervention effects in SCEDs are primarily evaluated for their clinical significance⁴⁷ (i.e., focus on large effect sizes à la Baer et al., 1968). A visual analysis involves the evaluation of data patterns both within and across phases or conditions.⁴⁸ In the following sections we will revise the main dimensions of within- and across-phases visual analysis.

47 This is in sharp contrast to group designs, which often focus on statistical significance and abstract effect size metrics.

48 We will use the terms phase and condition interchangeably. Phase may be more appropriate for independent variable manipulations comprising a continuous time series, whereas condition may be more appropriate for multi-element manipulations where the time series for a particular independent variable is not a continuous sequence.

Within-Phase Visual Analysis

Treatment effects are assessed relative to baseline. As a first step, the visual analyst would evaluate *steady-state responding* in the initial baseline phase. During steady-state responding the dependent variable is expected to present low variability and no discernable trends. The presence of high variability and/or trend within baseline threatens the baseline logic and may result in the researcher extending the baseline phase (until steady-state is attained) or modifying the study procedures in order to ensure baseline stability. In order to allow for timely phase change decisions, it is recommended that visual analyses are conducted on a continuous basis following the data collection process as closely as possible.

A visual analyst focuses on three major dimensions of the visual display of the data within a phase or condition: (a) *level*, the average amount of behavior; (b) *trend*, directional movement of the data over time; and (c) *variability*, data variation across the time series that cannot be attributed to trends. In effect, the visual analysis will provide information about how much behavior occurs (level), how the behavior changes over time (trend), and how much change do we see in the behavior that cannot be clearly attributed to the independent variable or to existing trends. The What Works Clearinghouse standards (2020) suggest that a minimum of five data points in each phase or condition are needed before introducing a new phase in order to properly appreciate level, trend, and variability (see also Maggin et al., 2013).

Level

The level is the overall or average amount of behavior within a condition with reference to the y-axis (Barton et al., 2018). *Floor* (zero), *low*, and *high* are common qualifiers of the level of behavior (see Figure 1a). The level within an SCED phase is often estimated using the mean or median (Figure 1b). The mean is highly sensitive to outliers. Therefore, in the presence of outliers or significant variability, the median is recommended. However, computing phase level may be informative only to the extent that the data has no trend and relatively low variability. Behavior analysts are mostly concerned with changes in level that may be revealed over the course of an intervention, whereas the mean (or median) change in level may have little practical relevance.

Trend

Trend refers to the directional movement of data over time. Trend may be *accelerating*, *decelerating*, *increasing*, *decreasing*, *therapeutic*, *counter-therapeutic*, or *null* if absent. Visually detecting

trends within a condition can offer information about the progressive improvement (or worsening) of the target behavior, and it may be the basis for phase change decisions. For example, (a) a therapeutic trend during baseline (i.e., a trend in the expected direction of the intervention) may indicate that the introduction of treatment is premature or unnecessary, (b) a counter-therapeutic trend during baseline (i.e., a trend opposing the expected direction of the intervention) or no trend may warrant the introduction of treatment, and (c) a therapeutic

trend during treatment may indicate that the treatment is having the intended effect and that termination of treatment before its effect becomes asymptotic may be premature.

Figure 2 (upper panel) presents a decreasing trend of vocal requests during baseline followed by an increasing trend after the onset of the treatment phase. While data stability is considered optimal before making a phase-change decision (i.e., steady-state responding), a baseline phase composed of a counter-therapeutic trend may be acceptable on occasions when it is not practically possible to wait for steady-state responding. The inception of an opposing trend immediately after treatment implementation is a form of effect demonstration—even though comparing mean levels would suggest no effect. Steep localized trends observed upon the implementation of treatment (or the reversal to baseline conditions), as the ones shown in Figure 2 (first B phases), are sometimes called *transition trends* and should not affect the ascertainment of behavior level once the trend is replaced with steady-state responding.

Discontinuing a study phase with an ongoing trend, while compatible with effect demonstration, may limit the researcher's ability to assess the intervention maximum effect and the intervention effect stability. Phase change or cessation decisions may be pre-determined by socially valid mastery criteria. For example, educational interventions focusing on behavior accuracy may require two successive sessions or blocks of trials with 80% of correct responses or more as a mastery criterion. This criterion can be portrayed in the graph as a horizontal mastery criterion line aiding visual analysis.

Trend can be appreciated visually or estimated through trend lines. A variety of techniques can be used to calculate trend lines including the split-middle line method or a simple regression line. Most graphing software packages (e.g., Prism GraphPad, SigmaPlot, SPSS) can compute regression trend lines seamlessly. The split-middle method (Cooper et al., 2019, p. 149-150) relies on the median and is therefore less sensible to outliers than linear regression lines.

Variability

In the context of a visual analysis, variability may be defined as the extent to which the data points within a condition vary randomly. It may also be defined as the fluctuation of data around the trend or level line (Barton et al., 2018). It can be referred to as *stable*, or *variable*.⁴⁹ Figure 1 presents examples of stable and variable baselines. As is the case with level and trend, variability can be ascertained by visually inspecting the data. Crying duration in Figure 1c seems fairly stable upon visual inspection. Variability may be augmented or minimized by the scale of the y-axis. In order to prevent a false appreciation of variability, it is recommended that the graph scale reflects the range of socially relevant variations of the dependent variable. In addition to visual analysis, some authors recommend computing a stability envelope to visually portray variability. According to Ledford et al. (2018), on occasions when 80% of data points or more are within $\pm 25\%$ of the median, the data may be described as stable. For example, for the time series in Figure 1d ($Me = 8$), the stability envelope would range from 6 and 10. There are only five values within the envelope (33%); therefore, the time series may be considered variable. For time series with an increasing or decreasing trend, the stability envelope may be calculated around the trend line, instead of the median. Similarly, Hagopian et al. (1997) suggested using criterion lines below and above the control condition as the basis for identifying intervention effects (see *Visual analysis: A case study* later in this chapter).

Within-phase variability is indicative of weak experimental control (i.e., weak functional relationship between the treatment and the target behavior). Therefore, variability in the data will hamper our ability to draw conclusions from it. Variability may also be interpreted differently in different experimental designs. For example, a greater level of variability is to be expected in a study with few data points per condition, where steady-state responding is unlikely. Similarly, multielement designs typically present greater variability than study designs where the independent variable is not quickly alternated, and are, therefore, less vulnerable to carryover effects. Whenever possible, it is recommended to identify the sources of unexplained variability, for example, by using different data paths for sessions conducted at home vs. sessions conducted in the community, or sessions conducted by different therapists, etc.

A visual analysis of variability often weights more observations conducted relatively later in a particular phase (i.e., towards the right of the graph). Namely, variability occurring be-

fore steady-state responding is attained can, to some extent, be ignored if it can be reasonably attributed to the gradual effect of the baseline or intervention conditions (e.g., gradual effect of discrimination, reinforcement, habituation, and extinction processes). By contrast, continued unexplained variability over an extended period of time suggests that key sources of variation remain unidentified, which could take the researchers back to the drawing board. The research setting may also affect the visual analyst judgement of variability. For example, we would expect more unexplained variability in studies conducted in ecological settings (e.g., classroom, home environment), where the opportunities to produce the target behavior may vary greatly (see for example Valbuena et al., 2015).

Across-Phase Visual Analysis

Across-phases visual analysis can help to establish (a) whether there was a change in the target behavior across conditions, and (b) whether this change was caused by the intervention. The What Works Clearinghouse standards (2020) offer specific criteria to evaluate a treatment effect. For example, the standards require at least three demonstrations of the intervention effect by considering changes in level, trend, and variability across adjacent phases in a reversal designs (or across participants, behaviors, or contexts in a multiple-baseline design). In the following sections we will consider immediacy, consistency, and overlap as specific dimensions of across-phase visual analysis.

Immediacy

Immediacy may be defined as the amount of time elapsed since the implementation of treatment (or the reversal to baseline conditions) to an identifiable change in behavior. Immediacy may be defined with reference to either the initial change of the dependent variable and/or the maximum change of the dependent variable. Visual analysts focus on the first few sessions after the implementation of treatment to assess immediacy. For example, in multielement designs, lack of immediacy could result in frequent dips in the intervention effect, or worse, undifferentiated data paths across conditions.

Abrupt changes in behavior upon the introduction of a new condition may seem ideal. However, abrupt or gradual changes in behavior are both acceptable from an internal validity perspective assuming the delay in behavior change is minimal (see several examples in the phase transitions of Figure 2). Moreover, abrupt changes in a behavioral time series may be partially

attributable to the degree of data aggregation. For example, computing responses per minute in 5-min sessions can more easily result in a gradual phase transition than computing responses per hour in 12-hour sessions (Fahmie & Hanley, 2008). Also, the nature of the underlying behavioral process may be key to obtaining gradual or abrupt changes in behavior (e.g., discriminated change in behavior allocation vs. shaping).

Overlap

Overlap refers to the extent to which the level and range of behavior is comparable across adjacent conditions. The greater the overlap, the lesser the confidence in a functional relation between treatment implementation and changes in behavior. Presence of variability during baseline and treatment phases greatly increases the likelihood of overlap. Overlap can be computed quantitatively by way of the percentage of non-overlapping data points (PND) and similar indices (see section on non-overlap indices later in this chapter). On occasions when overlap is present, it is important to evaluate whether it changes along local transition trends or whether overlap affects the entirety of the phases under analysis. In the latter scenario, our confidence in identifying a functional relation will diminish. For example, the reversal design in Figure 2 (lower graph) presents a failed baseline reversal (i.e., the second baseline phase fails to verify the level of behavior found in the initial baseline). The overlap between the treatment phases and the second baseline implies lack of experimental control. However, the level of overlap in this particular scenario does not suggest an absence of effect, but rather a failed replication or *non-reversibility* event. Experimental control can be restored by accompanying non-reversible effects with between-subject replications (cf., multiple-baseline design across participants, see Chapter 24).

Consistency

Consistency refers to the extent to which the level of behavior in a particular phase can be verified in subsequent phases featuring the same treatment, and the extent to which treatment effects can be replicated through multiple within-subject and between-subject replications. For example, in an ABAB reversal design, the transition between baseline and treatment is repeated. Therefore, the ABAB design provides an opportunity to verify baseline and treatment levels and to replicate the effect of treatment. Specifically, the ABAB design allows for the evaluation of the effectiveness of the intervention at three different moments in time: transition from first baseline to first intervention, from first intervention to second baseline and from second baseline to second intervention. The more similar the treatment effect across replications, the greater the

consistency. Highly consistent intervention effects enhance our ability to identify a functional relation. For example, the treatment replications in the upper graph of Figure 2 are reasonably consistent, whereas the replication was unsuccessful in the lower graph.

Non-reversible behavior changes, such as language acquisition, are not amenable to within-subject replications (e.g., reversal design). Yet, it is still possible to assess treatment consistency across participants. For example, the consistency across participants is high in the first baseline-to-treatment transition for both datasets in Figure 2, even though the treatment effect was not reversible for the lower dataset.

Reliability of Visual Analysis

While visual analysis is the most common method for analyzing SCEDs in applied behavior analysis, studies on its reliability and validity have been limited. Ninci et al. (2015) conducted a meta-analysis and a moderator analysis of SCED studies reporting the inter-rater reliability of visual analysts. Overall, they found a moderate level of interrater reliability (.76). Their moderator analysis suggested that the visual analyst's level of expertise did not facilitate interrater reliability, whereas providing consistent interpretation instructions, visual aids, and consistent training facilitated interrater agreement. In addition, Kahng et al. (2010) found consistent interpretations of SCED line graphs among well-trained visual analysts. However, they utilized relatively simple low-variability line graphs in their study. Other studies, are less optimistic about the reliability of visual analysis (see for example Wolfe et al., 2016).

Consistently with Ninci et al. (2015), the use of structured guidelines for visual analysis may increase its reliability and reduced the potential for Type I error. For example, Hagopian et al. (1997) proposed a set of guidelines for visually analyzing multielement designs used in the functional analysis of problem behavior. Their guidelines have been used in several reviews allowing visual analysts to attain over 90% inter-rater agreement (see for example Cox & Virues-Ortega, 2016). Moreover, Fisher, Kelley, and Lomas (2003) developed the dual-criteria (DC) and conservative dual-criteria (CDC) method to aid visual inspection of SCEDs. Broadly defined, the DC method involves calculating the mean and trend lines for baseline data and extending these lines over the subsequent phases. Next, the number of points in the successive data path(s) that fall above or below both lines are counted and compared to a cut-off value based on a binomial distribution. The CDC is a more stringent variation of the DC method, wherein the mean and trend lines are increased or decreased by one fourth of a standard deviation (Fisher et al., 2003). The

DC and CDC methods are intended to provide a more objective and reliable basis to conduct visual analyses. In yet another example, J. M. Ferron and others (2017) have proposed *masked visual analysis* as a means to minimize the potential of Type I errors in visual analysis (i.e., probability of finding an effect when there is none). We discussed masked visual analysis in the section on statistical analyses that produce a p value as the main outcomes. In addition to using structured criteria, it is possible that using automated tools to ascertain level, trend, and variability may also enhance the reliability of visual analysis (see for example De et al., 2020 and Manolov, 2020). More research is needed to validate the various approaches that have been proposed to enhance the reliability of visual analysis including supplementing visual analysis with statistical analyses.

Visual Analysis: A Case Study

We will illustrate the use of structured criteria for visual analysis with the guidelines proposed by Hagopian et al. (1997) for functional analysis of problem behavior. These criteria have been revised by Cox and Virues-Ortega (2016). We will apply these revised criteria to four multielement SCEDs reported by Conners et al. (2000) (a fully re-graphed version of the data is available in Figure 3). Before implementing these criteria, the visual analyst would draw in the functional analysis graphs the upper criterion line (UCL), defined as the mean plus the standard deviation of the play (control) condition; and the lower criterion line (LCL), defined as the mean minus the standard deviation of the play condition (Figure 3). The revised Hagopian et al. criteria are summarized in Table 1. Table 2 presents a synthesis of the revised criteria as used with the participants in Conners et al. (2000). Below we combine the structured visual analysis with some commentary on the within-phase (level, trend, variability) and across-phase (immediacy, overlap, consistency) dimensions of visual analysis.

The only test condition in Annette's functional analysis that met the criterion for differentiation was attention (*D1: observations above the UCL minus observations below the LCL equal at least half of the observation in the time series*). In addition, attention met the criteria for an upward differentiated trend (*T2: all observations above the UCL are in the second half of the assessment*). Therefore, the outcome of the visual analysis is attention. The visual inspection shows that alone, play, and demand, all present steady-state responding at zero or near-zero levels. The variability and overlap of the effect of attention (and its lack of immediacy and consistency) may be almost completely attributed to a local transition trend, possibly following the discriminative learning process present in a functional analysis.

For Bob's functional analysis, both alone and attention met the D1 criterion for differentiation. Yet, alone also met the criterion for a downward undifferentiated trend (*T1: among the observation above the UCL, fewer than 2 data points [or 20% of the time-series] occur in the second half of the assessment*). Therefore, alone should be discarded as a differentiated test condition. The outcome of the assessment is attention. In sharp contrast with Annette's functional analysis, Bob's presents a U-shaped trend that is identifiable to various extents in all conditions. This trend may explain some of the variability in the assessment in addition to the partial overlap of the fourth attention session. Variability is also evident by the relatively wide range defined by the LCL and UCL. Otherwise, the effect of attention is immediate and consistent.

On occasions when the UCL is below 0.5, the UCL is arbitrarily set at 0.5 (D3). This is the case for Jed and Max. For both of these participants only demand met the criterion for differentiation (D1), making this condition the outcome of the functional analysis. Attention, demand, and play present steady-state responding at near-zero levels for both datasets. The effect of demand is more consistent and demonstrates less overlap for Max. By contrast, Jed's demand time series presents two *dips* that overlap with some of the undifferentiated conditions. This lack of consistency may be attributed to the effect of demand being less immediate for Jed—to explore this possibility further, the effect of longer demand sessions or the presence of within-session trends could be evaluated. Even though the effect of demand is differentiated in both participants the relatively low rate of problem behavior could have easily become undifferentiated in the presence of higher response variability in the other conditions of the assessment.

SYNERGY BETWEEN STATISTICAL AND VISUAL ANALYSES

Arguments For and Against

Applied behavior analysts moved away from inferential statistics in part due to the treatment of within- and between-subject variation as residuals (error). By contrast, variation is often meaningful in applied behavior analysis. Specifically, the continuous analysis of a time-series at the individual level is critical to deliver services in applied behavior analysis. For example, one participant may respond to a reinforcement-based intervention (e.g., token reinforcement) within one session, whereas another may require 10 sessions for the intervention to reach its peak, yet a third participant may show no effect whatsoever prompting a swift change in course on the part of the interventionist (e.g., conducting new preference or reinforcer assessments, changing the

ratio requirement of a token economy, etc.). An idiographic approach requires a continuous feedback loop between intervention effect monitoring and procedural adjustments. This continuous process may not be possible solely by evaluating aggregated effects after the intervention.

Another argument against statistics in applied behavior analysis poses that statistical significance is heavily influenced by sample size, data variability and data distribution. These three aspects made traditional inferential statistics ineligible to small-*n* experimental designs. Specifically, the modest number of participants in single-subject experimental designs negates significant results even for hugely effective interventions. Moreover, analyzing trends as well as variability as a form of random variation or error can penalize bona fide effects of behavior-analytic interventions, where variability can be attributed to the gradual effect of the independent variable (i.e., trends) or to procedural fine-tuning (i.e., ideographic approach).

Finally, single-subject datasets cannot accommodate to traditional data distributions including the normal distribution due to the limited number of observations often reported in SCEDs. Fortunately, statistical developments over the last few decades can help to address some of these concerns.

Statistics are needed to enter the high-stakes outcome research literature upon which decisions for treatment policy and treatment prioritization are often made (see for example Keenan & Dillenburger, 2011 for a discussion). Effect size metrics and inferential statistics are sorely needed in applied behavior analysis, if not to compete, at least to achieve a common language with randomized controlled trials (RCT) and other group-based designs. However, statistical analyses should acknowledge the ideographic, hierarchical and longitudinal nature of behavior-analytic interventions.

STATISTICAL TECHNIQUES

Statistical techniques in SCED remain an active area of study where widely accepted methodological recommendations are lacking. In the following section we evaluate a selection of the most promising (and a few of the most popular) statistical techniques that have been developed for SCEDs. In the interest of brevity, we have classified the analytical techniques and effect size metrics for SCEDs in four groups: non-overlap and other non-parametric indices, mean-based effect sizes, analyses producing a *p* value as the main outcome, and regression-based analyses (a more detailed classification has been proposed by Manolov & Moeyaert, 2017). Below we pres-

ent samples of the most commonly used analyses within these groups. Systematic comparative summaries of these techniques can be found in Tables 3 and 4.

Non-Overlap and other Non-Parametric Indices

Non-overlap indices focus on a key feature of across-phase visual analysis by quantifying the extent to which the range of variation in baseline and treatment phases overlap. The greater the overlap, the less confident one can be on the treatment effect. These indices are easy to compute, require minimal training and have been widely used in the literature. In addition, they do not impose data distribution assumptions (e.g., normality). Yet, they are subject to numerous sources of bias. First, some of these indices are computed on the basis of a selection of the observations. For example, the percentage of non-overlapping data points (PND) is obtained by computing the percentage of data points during treatment that are above the highest baseline data point. Second, with few exceptions, the value of a non-overlap index is not indicative of the magnitude of the effect, but simply of the level of overlap. For example, an intervention increasing the level of baseline responding by 100 can produce a relatively lower non-overlap index compared to another intervention increasing responding by 1.5, assuming that the former incurs on greater overlap. Moreover, outliers can easily bias these indices as they often cause overlap. Additional limitations to non-overlap indices include their limited use in meta-analysis or moderator analysis (other than by averaging). They are not amenable to inferential statistics, or the analysis of co-variables, and they do not produce *p* values, although it is possible to obtain confidence intervals for some of them (Parker & Vannest, 2009). Moreover, there has been limited research on what represents a small, medium or large non-overlap index. Parker and Vannest (2009) obtained percentiles of common non-overlap indices providing some basis to ascertain effect magnitude (Table 3). Some non-overlap indices are less biased than others. For example, the NAP (non-overlap of all pairs), the PAND (percentage of all nonoverlapping data points) and the Tau-U include all the values in the time-series in the calculation process and are less vulnerable to the effect of outliers.

The log-response ratio is a non-parametric effect size index based on proportional change of rates of behavior across contiguous study phases (Pustejovsky, 2015). Log-response ratio effect sizes do not rely on often unverified parametric assumptions setting it apart from mean-based effect sizes. Unlike non-overlap indices it has been proposed to be fully amenable to meta-analysis (Pustejovsky, 2018).

Mean-Based Effect Sizes

The standardized mean difference (SMD) is the difference between the intervention level and baseline level in standardized units. In other words, the difference between the intervention and baseline mean is divided by the pooled within-case standard deviation (or the standard deviation of baseline). This is similar to Cohen's *d* developed for group studies (Cohen, 1992). Cohen's *d* can be corrected for small sample size bias (the standard deviation for each group is weighted by the sample size), which is Hedges' *g*. A version of Hedges' *g* was developed for use with SCEDs, and is known as the *between-case standardized mean difference* (BC-SMD; Pustejovsky et al., 2014). BC-SMD can also be characterized as a regression-based metric, as it is estimated using a two-level hierarchical linear model; level 1 is a within-case regression model and level 2 is between-case variation in regression coefficients (Pustejovsky et al., 2014). Because the difference between the baseline and intervention data is standardized using both the within- and between-case standard deviation, BC-SMD is assumed to be on the same scale as group designs, therefore, Cohen's *d* scale can be used for interpretation. Two other SCED metrics that relate to standardized mean difference is the mean phase difference (MPD; Manolov & Solanas, 2013) and mean baseline reduction (MBLR; Campbell, 2004). Both formulas for MPD and MBLR are similar to Cohen's *d*, but without the standard deviation denominator. Thus, they are unstandardized mean difference metrics.

Analyses Producing a *p* Value as the Main Outcome

We have selected the non-parametric randomization test and masked visual analysis to illustrate this category. The randomization test estimates the probability of the observed sequence in level changes across phases, relative to the null hypothesis (i.e., no changes in level across phases). Levin et al. (2012) provides a practical example:

“[S]uppose that there are 8 alternating A and B phases (i.e., an ABABABAB design), each consisting of one or more outcome observations per phase. Further suppose that it had been predicted that the intervention (B) phases would yield generally higher outcome observations than the baseline (A) phases. With 8 total phases, 4 A and 4 B, there are $8!/4!4! = 70$ possible assignments of 4 observations (or means based on multiple observations per phase) apiece to two groups, A and B. For each possible assignment, the mean difference between the B and A observations is calculated, yielding a permutation distribution of 70 mean differences. Now suppose that the actual mean difference

(i.e., the mean difference produced by the study's 4 actual B phase observations and 4 actual A phase observations) turned out to be the third largest of all 70 mean differences. Assuming that the 70 mean differences are all equally likely, the probability of obtaining a mean difference as large as or larger than the one actually observed is $3/70 = .043$, with a pre-designated one-tailed Type I error probability (α) of .05, would be a statistically significant result" (p. 604).

Ease of calculation, fitness to all major SCEDs (including the multielement design), and lack of data distribution assumptions are distinct advantages of randomization tests. However, the power of randomization tests is diminished in multi-element designs using a fixed condition alternation order (Levin et al., 2012), which is a common approach to functional analysis (e.g., Hammond et al., 2013). In addition, a randomization test is fully compatible with other statistical techniques including non-overlap effect sizes, mean-based effect sizes, and regression-based analyses (see a practical illustration of this combined approach in Heyvaert & Onghena, 2013).

Ferron and Jones (2006) have proposed using randomization strategies as part of a masked visual analysis process with the aim of reducing Type I errors (see also Ferron et al., 2017). For example, in a multiple-baseline design with four participants the experimenter would hide phase transitions in the graph and randomize the order of the participants within the graph. The experimenter would then ask an independent visual analyst to guess the time points at which the intervention started for each participant. In this example, the probability of correctly guessing the order in which the intervention started for each participant by chance equals $1/n$, where n is the number of participants in the multiple baseline design. Therefore, if a visual analyst correctly identifies the order of the intervention sequence, this would ensure that the Type I error is below .05. Moeyaert et al. (2020) have developed a user-friendly mobile app to conduct masked visual analysis.

Regression-Based Analyses

Different types of regression-based approaches have been used for SCEDs such as ordinary least squares (OLS; Huitema & McKean, 2000), generalized least squares (GLS; Swaminathan et al., 2014), and multilevel models (Noorgate & Onghena, 2003). OLS is a parametric approach that can quantify the change in level between baseline and intervention phase, as well as the change in trend (Huitema & McKean, 2000). Additional parameters can be added to the model depending on the research question (e.g., it can be used to estimate the baseline level at the start

of data collection, the trend during baseline, and the changes in trend and/or level). However, OLS makes several assumptions, including the assumption of normality, homoscedasticity (equality of variances across conditions), and independence of errors. In instances where these assumptions may be inappropriate to make (e.g., if heteroscedasticity or autocorrelation is present), the GLS approach is a more viable and appropriate alternative. GLS is similar to OLS, except that it can reflect count data (Declercq et al., 2019) and account for autocorrelation by adjusting the residuals (Swaminathan et al., 2014). Like OLS, GLS is able to reflect both change in slope and change in level together in one model and provide separate estimates of these regression parameters.

Hierarchical linear modeling is an extension of single-level regression analysis, and accounts for the nesting of measurement occasions within participants. This modeling approach can be used to estimate change in level and/or slope for individual participants and across participants. As such, HLM is able to estimate between-case differences in intervention effectiveness. If a decent amount of between-case variance is found, moderators can be added in an effort to explain variability in intervention effectiveness between participants.

Dataset produced in behavior-analytic studies can be characterized as having a hierarchical or multilevel structure. Levels can be construed as key aspect of the data source that may determine different aspects of the variability. For example, in a study including several multiple baseline designs across behaviors for three participants, some of the differences observed in the behavior may be explained by differences within participants across sessions (i.e., variability across time points) and by differences between participants characteristics (variability across subjects). In addition, subject personal characteristics (e.g., diagnostic group, age, standardized outcomes) may also be included as covariates within the multilevel analysis.

An exciting aspect of multi-level analysis is its flexibility. It is possible to add as many levels to the analysis as it would be conceptually and statistically appropriate⁵⁰. It is also possible to add a study level in order to generalize intervention effectiveness across similar studies, thereby conducting a meta-analysis (Moeyaert et al., 2014a). It is also possible to compute the effect size of key predictors in HLM (e.g., treatment status) and to compare the goodness of fit of various models (Lorah, 2018).

50 Adding a specific level might have conceptual sense, but if that factor has few units, then the estimates could be biased and might be preferable to omit that level.

Statistical Analysis: A Case Study with a Two-Level Hierarchical Model

Among the various statistical approaches described to evaluate treatment effects in single-subject data, multilevel analysis or hierarchical linear modeling (HLM) has various distinct advantages. First, acknowledging the hierarchical nature of SCED data can, potentially, minimize unexplained variance and effect underreporting (Baek et al., 2014; Moeyaert et al., 2014a, 2017). In addition, the ability to add predictors can account for characteristics of SCED, which have been ignored in alternative approaches. For example, few of the traditional effect size metrics discussed here have been adapted to multielement designs; one of the most common research designs in applied behavior analysis (Virues-Ortega et al., 2016). The case study below illustrates the use of multilevel analysis in the functional analyses by Connors et al. (2000) and exemplifies the synergistic use of visual and statistical analyses in SCEDs (see visual analysis case study earlier in this chapter).

We will consider two hierarchical levels: measurement occasions (level 1) are nested within subjects (level 2). Behavior refers to the amounts of behavior measured in responses per minute. Subject identifies each of the four participants. The key predictor for the behavior level is treatment status. For the purposes of this analysis the differentiated conditions of the functional analysis (attention for Annette and Bob, and demand for Jed and Max) are considered as target treatment conditions, whereas undifferentiated conditions (i.e., alone, play, and demand for Annette and Bob; and alone, attention, and play for Jed and Max) will be considered non-target treatment conditions. Therefore, we created the variable *treatment status* (*tts*) as a dummy variable, coded as 1 for the sessions of differentiated conditions and 0 for the sessions of undifferentiated conditions. *Session* refers to the ordinal session number for each participant. The behavior level, Y_{ij} , is the behavior level of participant (j) at measurement occasion (i). The independent variable is a dummy variable, indicating the treatment status of participant (j) at measurement occasion (i). Therefore, β_{0j} indicates the expected behavior level for participant j during the undifferentiated condition and β_{1j} indicates the expected change in behavior level between the undifferentiated and differentiated condition for participant j . The within-participant errors (e_{ij}) are assumed to be homogeneous and normally distributed, with a lag1 autocorrelation. A lag1 autocorrelation indicates that two consecutive errors are correlated, which is likely in repeated measures design. The level 1 model is displayed in Equation 1:

$$\text{Level 1 (within participants): } Y_{ij} = \beta_{0j} + \beta_{1j}D_{ij} + e_{ij} \text{ with } e_{ij} \sim N(0, \sigma_e^2) \quad (1)$$

The level 1 parameters (at the right side of the equation sign in Equation 1) are allowed to vary at the second level as it is to be expected that behavior levels during the undifferentiated treatment (β_{0j} 's), as well as changes in level between the undifferentiated and differentiated treatments (β_{1j} 's), vary between participants. Therefore, these parameters are a function of an overall average effect across participants (reflected by the θ 's) and individual differences (reflected by the u 's). This is reflected in the level 2 model, displayed in Equation 2:

$$\text{Level 2 (across participants): } \begin{cases} \beta_{0j} = \theta_0 + u_{0j} \\ \beta_{1j} = \theta_1 + u_{1j} \end{cases}$$

$$\text{with } \begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{u_0}^2 & \\ & \sigma_{u_1}^2 \end{bmatrix} \right), \text{ and } e_{ij} \sim N(0, \sigma_e^2) \quad (2)$$

θ_0 refers to the overall behavior level during the undifferentiated condition across the J participants. The deviation of participant j from the overall level θ_0 is indicated by u_{0j} . These deviations are assumed to be normally distributed around 0 with a variance of $\sigma_{u_0}^2$. As such, $\sigma_{u_0}^2$ represents the between-participant variance of the behavior level during the undifferentiated conditions of the functional analysis. Similarly, θ_1 refers to the overall change in behavior level between the undifferentiated and differentiated conditions across the J participants. u_{1j} refers to the deviation of participant j from the difference in level. These deviations are assumed to be normally distributed with an average of 0 and variance $\sigma_{u_1}^2$. Thus, $\sigma_{u_1}^2$ represents the between-participant variance of the change in level. Likewise, the within-participant errors, e_{ij} are assumed to be normally distributed around 0 with a variance of σ_e^2 . Therefore, σ_e^2 indicates the within-participant error variance. The within-participant errors (i.e., e_{ij} 's) are assumed to be correlated. By substituting the level 2 equations into the level 1 equation, the combined two-level hierarchical linear model can be written as:

$$y_{ij} = \theta_0 + u_{0j} + (\theta_1 + u_{1j})D_{ij} + e_{ij} \text{ with } \begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N(0, \Sigma_v), e_{ij} \sim N(0, \sigma_e^2), \quad (3)$$

The basic two-level HLM approach introduced in Equation 3 is promising and its appropriateness in summarizing SCED data has been empirically validated through large-scale Monte Carlo simulation studies (e.g., Ferron et al., 2010).

We evaluated the current multi-level model (presented in Equation 3) using the statistical package *R* for Mac (Urbanek et al., 2020). Table 5 presents the coding of the design matrix for one randomly selected participant (i.e., Bob). For a detailed discussion on how to appropriately

setting up the design matrix, see Moeyaert et al. (2014b). The R code for the proposed analyses is displayed and described in Figure 4. The results of the two-level HLM analysis are presented and described in Figure 5. The output indicated that the difference in behavior level between the target and non-target conditions is estimated to be 5.86, and is statistically significant, $\hat{\theta}_0 = 5.86$, $t(103) = 2.07$, $p = .04$. This analysis provides a quantitative indication of functional analysis differentiation in this context. However, the same analysis could be used to quantify treatment effects. We encourage the interested reader to check Becraft et al. (2020) for a tutorial on using multi-level models and Rodabaugh and Moeyaert (2017) for additional information.

CONCLUSION

Some researchers have argued that the use of visual inspection alone as a standard for demonstrating functional relations may inhibit communication with researchers from other fields unaccustomed to this methodology (Falligant et al., 2020). The modest interrater agreement indices reported in the literature (Ninci et al., 2015) in addition to other sources of bias present in SCED call for more stringent methodological standards for visual analysis in SCED as applied behavior analysis enter the high-stakes outcome research arena.

We expect that the status of visual analysis and SCED as mainstream methods of treatment evaluation and data analysis will be established more firmly as progress continues in the following fronts: (a) widespread evaluation of methodological quality in SCED (as it is customary for RCTs), (b) widespread evaluation of the interrater reliability of visual analysts (as it is the case for interobserver agreement or treatment fidelity), (c) widespread use of intention-to-treat analysis and/or consecutive case reporting, (d) widespread use of structural guidelines for visual analysis (e.g., Hagopian et al., 1997; Fisher et al., 2003), and (e) widespread use of statistical analyses and meta-analysis-compatible effect size metrics *in conjunction with* visual analysis.

Progress in these various fronts would require a level of consensus in the field that is yet to be achieved. Yet, the basic elements to attain these standards are already well understood. A challenge of particular importance for incorporating statistical analysis in SCED results from the limited sensitivity of the available statistical techniques to the logic of SCEDs. For example, practically no statistical techniques properly model steady-state responding, weights within-subject replications, or provides estimates of carryover effect (Table 4). These SCED-specific features ought to be modelled in a unified theory of SCED data analysis. In this respect, hierarchical mod-

els offer the greatest flexibility to researcher and practitioners to assess these various effects. The practical conjunction of statistical and visual analyses may depend on the theoretical consistency of these two approaches to data analysis and interpretation.

REFERENCES

- Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park: Sage.
- Baek, E., Moeyaert, M., Petit-Bois, M., Van den Noortgate, W., Beretvas, S., & Ferron, J. (2014). The use of multilevel analysis for integrating single-case experimental design results within a study and across studies. *Neuropsychological Rehabilitation*, 24(3/4), 590-606. <https://doi.org/10.1080/09602011.2013.835740>
- Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*, 1(1), 91–97. <https://doi.org/10.1901/jaba.1968.1-91>
- Bailey, J. S., & Burch, M. R. (2017). *Research methods in applied behavior analysis* (2nd ed.). Routledge.
- Barton, E. E., Lloyd, B. P., Spriggs, A. D., Gast, D. L. (2018). Visual analysis of graphic data. In J. R. Ledford & D. L. Gast (Eds.), *Single-case research methodology: Applications in special education and behavioral sciences* (pp. 179–213). Routledge.
- Becraft, J. L., Borrero, J. C., Sun, S. & McKenzie, A. A. (2020). A primer for using multilevel models to meta-analyze single case design data with AB phases. *Journal of Applied Behavior Analysis*, 53: 1799-1821. <https://doi.org/10.1002/jaba.698>
- Brossart, D. F., Laird, V. C., & Armstrong, T. W. (2018). Interpreting Kendall's Tau and Tau-U for single-case experimental designs. *Cogent Psychology*, 5(1). <https://doi.org/10.1080/23311908.2018.1518687>
- Campbell J. M. (2004). Statistical comparison of four effect sizes for single-subject designs. *Behavior Modification*, 28(2), 234–246. <https://doi.org/10.1177/0145445503259264>

- Center, B. A., Skiba, R. J., & Casey, A. (1985). A methodology for the quantitative synthesis of intra-subject design research. *The Journal of Special Education*, 19(4), 387–400.
- Cohen J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155–159. <https://doi.org/10.1037//0033-2909.112.1.155>
- Conners, J., Iwata, B. A., Kahng, S. W., Hanley, G. P., Worsdell, A. S., & Thompson, R. H. (2000). Differential responding in the presence and absence of discriminative stimuli during multiple functional analyses. *Journal of Applied Behavior Analysis*, 33(3), 299–308. <https://doi.org/10.1901/jaba.2000.33-299>
- Cooper, J. O., Heron, T. E., & Heward, W. L. (2019). *Applied Behavior Analysis (3rd Edition)*. Pearson Education.
- Cox, A. D., & Virues-Ortega, J. (2016). Interactions between behavior function and psychotropic medication. *Journal of Applied Behavior Analysis*, 49(1), 85–104. <https://doi.org/10.1002/jaba.247>
- De, T. K., Michiels, B., Vlaeyen, J. W. S., & Onghena, P. (2020). Shiny SCDA [Computer software]. <https://ppw.kuleuven.be/mesrg/software-and-apps/shiny-scda>
- Declercq, L., Jamshidi, L., Fernández-Castilla, B., Beretvas, S., Moeyaert, M., Ferron, J., & Van den Noortgate, W. (2019). Analysis of single-case experimental count data using the linear mixed effects model: A simulation study. *Behavior Research Methods*, 51(6), 2477–2497. <https://doi.org/10.3758/s13428-018-1091-y>
- DeProspero, A., & Cohen, S. (1979). Inconsistent visual analysis of intrasubject data. *Journal of Applied Behavior Analysis*, 12, 285–296. <https://doi.org/10.1901/jaba.1979.12-573>
- Edgington, E. S. (1967). Statistical inference from N=1 experiments. *The Journal of Psychology*, 65, 195–199.
- Fahmie, T. A., & Hanley, G. P. (2008). Progressing toward data intimacy: a review of within-session data analysis. *Journal of Applied Behavior Analysis*, 41(3), 319–331. <https://doi.org/10.1901/jaba.2008.41-319>

- Falligant, J. M., McNulty, M. K., Hausman, N. L., & Rooker, G. W. (2020). Using dual-criteria methods to supplement visual inspection: Replication and extension. *Journal of Applied Behavior Analysis*, 53(3), 1789-1798. <https://doi.org/10.1002/jaba.665>.
- Ferron, J. M., Farmer, J. L., & Owens, C. M. (2010). Estimating individual treatment effects from multiple-baseline data: a Monte Carlo study of multilevel-modeling approaches. *Behavior Research Methods*, 42(4), 930-943. <https://doi.org/10.3758/BRM.42.4.930>
- Ferron, J., & Jones, P. K. (2006). Tests for the visual analysis of response-guided multiple-baseline data. *Journal of Experimental Education*, 75, 66-81. <https://doi.org/10.3200/JEXE.75.1.66-81>
- Ferron, J. M., Joo, S. H., & Levin, J. R. (2017). A Monte-Carlo evaluation of masked-visual analysis in response-guided versus fixed criteria multiple-baseline designs. *Journal of Applied Behavior Analysis*, 52(4), 701-716. <https://doi.org/10.1002/jaba.410>
- Fisher, W. W., Kelley, M. E., & Lomas, J. E. (2003). Visual aids and structured criteria for improving visual inspection and interpretation of single-case designs. *Journal of Applied Behavior Analysis*, 36(3), 387-406. <https://doi.org/10.1901/jaba.2003.36-387>
- Hagopian, L. P., Fisher, W. W., Thompson, R. H., Owen-DeSchryver, J., Iwata, B. A., & Wacker, D. P. (1997). Toward the development of structured criteria for interpretation of functional analysis data. *Journal of Applied Behavior Analysis*, 30(2), 313-326. <https://doi.org/10.1901/jaba.1997.30-313>
- Hagopian L. P. (2020). The consecutive controlled case series: Design, data-analytics, and reporting methods supporting the study of generality. *Journal of Applied Behavior Analysis*, 53(2), 596-619. <https://doi.org/10.1002/jaba.691>
- Hammond, J. L., Iwata, B. A., Rooker, G. W., Fritz, J. N., & Bloom, S. E. (2013). Effects of fixed versus random condition sequencing during multielement functional analyses. *Journal of Applied Behavior Analysis*, 46(1), 22-30. <https://doi.org/10.1002/jaba.7>
- Hedges, L. V., Pustejovsky, J. E., & Shadish, W. R. (2012). A standardized mean difference effect size for single case designs. *Research Synthesis Methods*, 3(3), 224-239. <https://doi.org/10.1002/jrsm.1052>

- Hedges, L. V., Pustejovsky, J. E., & Shadish, W. R. (2013). A standardized mean difference effect size for multiple baseline designs across individuals. *Research Synthesis Methods*, 4(4), 324–341. <https://doi.org/10.1002/jrsm.1086>
- Heyvaert, M., & Onghena, P. (2013). Analysis of single-case data: Randomisation tests for measures of effect size. *Neuropsychological Rehabilitation*, 24(3-4), 507–527. <https://doi.org/10.1080/09602011.2013.818564>
- Heyvaert, M., & Onghena, P. (2014). Randomization tests for single-case experiments: State of the art, state of the science, and state of the application. *Journal of Contextual Behavioral Science*, 3(1), 51–64. <https://doi.org/10.1016/j.jcbs.2013.10.002>
- Huitema, B. E., & McKean, J. W. (2000). A simple and powerful test for autocorrelated errors in OLS intervention models. *Psychological Reports*, 87(1), 3–20. <https://doi.org/10.2466/PRO.87.5.3-20>
- Kahng, S., Chung, K. M., Gutshall, K., Pitts, S. C., Kao, J., & Girolami, K. (2010). Consistent visual analyses of intrasubject data. *Journal of Applied Behavior Analysis*, 43(1), 35–45. <https://doi.org/10.1901/jaba.2010.43-35>
- Keenan, M., & Dillenburger, K. (2011). When all you have is a hammer ...: RCTs and hegemony in science. *Research in Autism Spectrum Disorders*, 5(1), 1–13. doi:10.1016/j.rasd.2010.02.003
- Kratochwill, T. R., & Levin, J. R. (2014). Introduction: An overview of single-case intervention research. In T. R. Kratochwill & J. R. Levin (Eds.) *Single-Case Intervention Research: Methodological and Statistical Advances*. American Psychological Association.
- Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). Single-case design technical documentation. http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf.
- Krentz, H., Miltenberger, R., & Valbuena, D. (2016). Using token reinforcement to increase walking for adults with intellectual disabilities. *Journal of Applied Behavior Analysis*, 49(4):745-750. <https://doi.org/10.1002/jaba.326>.

- Kubina, R. M., Kostewicz, D. E., Brennan, K. M., & King, S. A. (2017). A critical review of line graphs in behavior analytic journals. *Educational Psychology Review*, 29, 583-598. <https://doi.org/10.1007/s10648-015-9339-x>
- Ledford, J. R., Lane, J. D., & Severini, K. E. (2018). Systematic use of visual analysis for assessing outcomes in single case design studies. *Brain Impairment*, 19(1), 4–17. <https://doi.org/10.1017/Brlmp.2017.16>
- Levin, J. R., Ferron, J. M., & Kratochwill, T. R. (2012). Nonparametric statistical tests for single-case systematic and randomized ABAB...AB and alternating treatment intervention designs: New developments, new directions. *Journal of School Psychology*, 50(5), 599–624. <https://doi.org/10.1016/j.jsp.2012.05.001>
- Lorah, J. (2018). Effect size measures for multilevel models: definition, interpretation, and TIMSS example. *Large-Scale Assessment in Education*, 6, 8. <https://doi.org/10.1186/s40536-018-0061-2>
- Ma, H. H. (2006). An alternative method for quantitative synthesis of single-subject research: Percentage of data points exceeding the median. *Behavior Modification*, 30(5), 598–617.
- Maggin, D. M., Briesch, A. M., & Chafouleas, S. M. (2013). An application of the What Works Clearinghouse standards for evaluating single-subject research: Synthesis of the self-management literature base. *Remedial and Special Education*, 34(1),44-58. <https://doi.org/10.1177/0741932511435176>
- Manolov, R. (2020). *Visual aids and nonoverlap indices* [online app]. <https://manolov.shinyapps.io/Overlap/>
- Manolov, R., & Moeyaert, M. (2017). Recommendations for choosing single-case data analytical techniques. *Behavior Therapy*, 48(1), 97–114. <https://doi.org/10.1016/j.beth.2016.04.008>
- Manolov, R., & Solanas, A. (2013). A comparison of mean phase difference and generalized least squares for analyzing single-case data. *Journal of School Psychology*, 51(2), 201–215. <https://doi.org/10.1016/j.jsp.2012.12.005>

- Michael J. (1974). Statistical inference for individual organism research: some reactions to a suggestion by Gentile, Roden, and Klein. *Journal of Applied Behavior Analysis*, 7(4), 627–628. <https://doi.org/10.1901/jaba.1974.7-627>
- Moeyaert, M., Akhmedjanova, D., Ferron, J., Beretvas, S., & Van Den Noortgate, W. (2020). Effect size estimation for combined single-case experimental designs. *Evidence-based Communication Assessment and Intervention*. 10.1080/17489539.2020.1747146
- Moeyaert, M., Bursali, S., & Ferron, J. (2020). SCD-MVA: A mobile application for conducting single-case experimental design research during the pandemic. *Human Behavior and Emerging Technologies*. <https://doi.org/10.1002/hbe2.223>
- Moeyaert, M., Ferron, J., Beretvas, S., & Van den Noortgate, W. (2014a). From a single-level analysis to a multilevel analysis of single-subject experimental data. *Journal of School Psychology*, 52(2), 191-211. <https://doi.org/10.1016/j.jsp.2013.11.003>
- Moeyaert, M., Rindskopf, D., Onghena, P., & Van den Noortgate, W. (2017). Multilevel modeling of single-case data: A comparison of Maximum Likelihood and Bayesian estimation. *Psychological Methods*, 22(4), 760-778. <https://doi.org/10.1037/met0000136>
- Moeyaert, M., Ugille, M., Ferron, J., Beretvas, S., & Van den Noortgate, W. (2014b). The influence of the design matrix on treatment effect estimates in the quantitative analyses of single-case experimental design research. *Behavior Modification*, 38(5), 665-704. <https://doi.org/10.1177/0145445514535243>
- Ninci, J., Vannest, K. J., Willson, V., & Zhang N. (2015). Interrater agreement between visual analysts of single-case data: A meta-analysis. *Behavior Modification*, 39(4), 510-541. <https://doi.org/10.1177/0145445515581327>
- Parker, R. I., Hagan-Burke, S., & Vannest, K. (2007). Percentage of All Non-Overlapping Data (PAND). *The Journal of Special Education*, 40(4), 194–204. <https://doi.org/10.1177/00224669070400040101>
- Parker, R. I., & Vannest, K. J. (2009). An improved effect size for single case research: Non-overlap of all pairs (NAP). *Behavior Therapy*, 40, 357-367. <https://doi.org/10.1016/j.beth.2008.10.006>

- Parker, R. I., Vannest, K. J., & Brown, L. (2009). The improvement rate difference for single-case research. *Exceptional Children*, 75(2), 135–150.
- Parker, R. I., Vannest, K. J., Davis, J. L., & Sauber, S. B. (2011). Combining nonoverlap and trend for single-case research: Tau-U. *Behavior therapy*, 42(2), 284–299. <https://doi.org/10.1016/j.beth.2010.08.006>
- Pustejovsky, J. E. (2015). Measurement-comparable effect sizes for single-case studies of free-operant behavior. *Psychological Methods*, 20(3), 342–359. <https://doi.org/10.1037/met0000019>
- Pustejovsky, J. E. (2018). Using response ratios for meta-analyzing single-case designs with behavioral outcomes. *Journal of School Psychology*, 16, 99–112. <https://doi.org/10.1016/j.jsp.2018.02.003>
- Pustejovsky, J. E., Hedges, L. V., & Shadish, W. R. (2014). Design-Comparable Effect Sizes in Multiple Baseline Designs. *Journal of Educational and Behavioral Statistics*, 39(5), 368–393. doi:10.3102/1076998614547577
- Rodabaugh, E., & Moeyaert, M. (2017). Multilevel modeling of single-case data: An Introduction and tutorial for the applied researcher. *NERA Conference Proceedings 2017*, 8. <https://opencommons.uconn.edu/nera-2017/8>
- Scruggs, T. E., & Mastropieri, M. A. (1998). Summarizing Single-Subject Research. *Behavior Modification*, 22(3), 221–242. <https://doi.org/10.1177/01454455980223001>
- Sham, E., & Smith, T. (2014). Publication bias in studies of an applied behavior-analytic intervention: An initial analysis. *Journal of Applied Behavior Analysis*, 47(3), 663–678. <https://doi.org/10.1002/jaba.146>
- Swaminathan, H., Rogers, H. J., & Horner, R. H. (2014). An effect size measure and Bayesian analysis of single-case designs. *Journal of School Psychology*, 52(2), 213–230. <https://doi.org/10.1016/j.jsp.2013.12.002>
- Swan, D. M., & Pustejovsky, J. E. (2018). A Gradual Effects Model for Single-Case Designs. *Multivariate Behavioral Research*, 53(4), 574–593. <https://doi.org/10.1080/00273171.2018.1466681>

- Tarlow, K. R. (2017a). An improved rank correlation effect size statistic for single-case designs: Baseline corrected Tau. *Behavior Modification*, 41, 427-467. <https://doi.org/10.1177/0145445516676750>
- Taylor, S. L., Purdy, S., Jackson, B., Phillips, K., & Virues-Ortega, J. (2019). Evaluation of a home-based Behavioral treatment model for children with tube dependency. *Journal of Pediatric Psychology*, 44, 656-668 <https://doi.org/10.1093/jpepsy/jsz014>
- Urbanek, S., Bibiko, H.-J., & Iacus, S. M. (2020). *R for Mac* (Version 1.73). [Desktop computer software]. <https://cran.r-project.org/bin/macosx>
- Valbuena, D., Miltenberger, R., & Solley, E. (2015). Evaluating an Internet-based program and a behavioral coach for increasing physical activity. *Behavior Analysis: Research and Practice*, 15(2), 122-138. <http://dx.doi.org/10.1037/bar0000013>
- Van den Noortgate, W., & Onghena, P. (2003). Combining single-case experimental studies using hierarchical linear models. *School Psychology Quarterly*, 18(3), 325-346. <https://doi.org/10.1521/scpq.18.3.325.22577>
- What Works Clearinghouse (2020). *Standards handbook* (version 4.1). Washington, DC: Institute of Education Sciences.
- Virues-Ortega, J., Pear, J. J., Boris, A. L., Cox, A. D., Hu, L., Julio, F. M., Michalyszyn, C. A., & Wightman, J. K. (2016). Single-subject research design. In H. L. Miller, *The SAGE Encyclopedia of Theory in Psychology* (pp. 859-862). New York: Sage. <http://doi.org/10.4135/9781483346274.n293>
- Wolfe, K., Barton, E.E. & Meadan, H. (2019). Systematic protocols for the visual analysis of single-case research data. *Behavior Analysis in Practice*, 12, 491-502. <https://doi.org/10.1007/s40617-019-00336-7>
- Wolfe, K., Seaman, M. A., & Drasgow, E. (2016). Interrater agreement on the visual analysis of individual tiers and functional relations in multiple baseline designs. *Behavior Modification*, 40(6), 852-873. <https://doi.org/10.1177/0145445516644699>
- Xu, R. (2003). Measuring Explained Variation in Linear Mixed Effects Models. *Statistical Methods*, 22, 3527-3541. <https://doi.org/10.1002/sim.1572> <https://doi.org/10.1002/sim.1572>

Table 6.1 Structured criteria for the visual analysis of a multielement design functional analysis (according to Hagopian et al., 1997, and Cox & Virues-Ortega, 2016).

<i>General criteria</i>	
D1	Criterion for differentiation: points above the UCL line minus points below the LCL ≥ 5
D2	If play LCL is 0 or lower, zeros counted as below the lower line.
D3	If the play's UCL is below 0.5, draw the upper criterion line at 0.5.
D4	If no play/control condition and the highest condition is attention, demand or tangible (or any two-term combination of these) set LCL and UCL to 0 and 0.5, respectively.
D5	For % of intervals or % of session duration, the maximum UCL is 100%, any 100% data point will be considered above the UCL.
D6	If there are less or more than 10 data points per condition use criteria proportionally, e.g., 50% of data points above UCL vs. 5 data points above UCL. Invoke D6 only if there are at least 2 data points per condition. (D1 is relevant to A3, T1, and T2).
<i>Automatic reinforcement</i>	
A1	Alone is the highest and is significantly higher than play (is the highest condition in most series).
A2	Behavior is higher in conditions with lower external stimuli (one or more of alone, attention, tangible) relative to those with higher stimuli (one or more of play, demand).
A3	All conditions are high and relatively stable with no overall trends (mean of all conditions is above 1.5 per min [or 15% intervals] and less than 5 zero points).
<i>Trends</i>	
T1	Downward trend (undifferentiated): If less than two data points above the UCL occur in the second half of the assessment. This rule does not apply to demand and tangible if responding adapts to an efficient rate (e.g., every 30 s in demand).
T2	Upward trend (differentiated): All 5 data points above the UCL are in the second half of the assessment, ignore points below the LCL.
T3	Overall trend (differentiated): condition that is consistently higher than play.
<i>Low rate or magnitude</i>	
LR	1. Most of the data points are low across all conditions AND 2. More than half of the high sessions occur in a test condition AND 3. More than half of the behaviors occur in the same condition as #2 AND 4. At least one of the high points in the condition identified in #2 should occur in the second half of the assessment.
LM	Condition meets criterion for differentiation (D1) by a small amount, then raise UCL by 20%. LM1 can be used only if T2 is not applicable.
<i>Multiple control</i>	
M1	Multiple meet criterion for differentiation and alone does not have the highest mean.
M2	M1 applies to all test conditions, then ignore alone.
M3	Two differentiated conditions and alone is the lowest (compute means); outcome: multiply controlled (i.e., automatic plus a social contingency).

Notes. Series = every alone-attention-play-demand sequence; LCL = lower criterion line (play condition mean minus its standard deviation); UCL = upper criterion line (mean of the play condition plus its standard deviation).

Table 6.2 Structured visual analysis for Annette, Bob, Jed, and Max.

	Revised Hagopian criteria		Outcome
	General	Specific	
Annette	D6		Attention
Alone			
Attention		D1, T2	
Demand			
Bob	D6		Attention
Alone		T1	
Attention		D1	
Demand			
Jed	D2, D3, D6		Demand
Alone			
Attention			
Demand		D1	
Max	D2, D3, D6		Demand
Alone			
Attention			
Demand		D1	

Table 6.3 Key characteristics of selected single-subject effect size metrics and analytic techniques.

	Concept	Range of usable values	Key references
<i>Non-overlap and other non-parametric indices</i>			
NAP	Non-overlap of all pairs, percentage of data that improve from the baseline phase to the intervention phase.	P10 = 50%; P25 = 69%; P50 = 84%; P75 = 98%; P90 = 100%	Parker & Vannest (2009)
PND	Percentage of non-overlapping data, ratio of number of Phase B data points above the lowest/highest data point in Phase A divided by Phase B total data points.	Range, 0% to 100%; >70% effective intervention; 50% to 70%, questionable effectiveness; <50% no observed effect.	Parker & Vannest, (2009), Scruggs & Mastropieri (1998)

	Concept	Range of usable values	Key references
PAND	Percentage of all non-overlapping data, minimum number of data points removed from Phase A and/or Phase B to eliminate all overlap between phases.	P10 = 60%; P25 = 69%; P50 = 82%; P75 = 93%; P90 = 100%	Parker et al. (2007), Parker & Vannest, (2009)
Tau-U	Non-overlap index that accounts for changes in level and trend in Phase B. It has the ability to correct for baseline trend.	P25 = .56; P50 = .84; P75 = 1.00 (with baseline trend correction)	Brossart et al., (2017), Parker et al. (2011), Tarlow (2017)
Log-response ratio	Metric based on proportionate change.	Not available	Pustejovsky (2015)
<i>Mean-based effect sizes</i>			
WC-SMD	<i>Within-case standardized mean difference</i> , difference between the baseline and treatment means divided by the baseline standard deviation.	Range, $-\infty$ to $+\infty$; very small, 0.01 to 0.19; small, 0.20 to 0.49; medium, 0.50 to 0.79; large, ≥ 0.80	Busk & Serlin (1992)
BC-SMD	<i>Between-case standardized mean difference</i> , difference between treatment and baseline means divided by the square root of the sum of the variance of observations within-cases plus the variance of observations between-cases.	Range, $-\infty$ to $+\infty$; very small, 0.01 to 0.19; small, 0.20 to 0.49; medium, 0.50 to 0.79; large, ≥ 0.80	Hedges et al. (2012), Pustejovsky et al. (2014)
<i>Analyses producing a p value as the main outcome</i>			
Randomization test	Probability of the rank of the actual baseline and treatment observations relative to all possible permutations of ranks.	Same as a one-tailed Type I error probability. If $p \leq 0.05$, then the distribution of higher (or lower) treatment observations is statistically significant.	Heyvaert, M., & Onghena, P. (2014).
Masked visual analysis	Probability that an independent visual analyst correctly guesses the true intervention sequence order.	Same as a one-tailed Type I error probability. If $p \leq 0.05$, then the distribution of higher (or lower) treatment observations is statistically significant.	Ferron and Jones (2006)
<i>Regression-based analyses</i>			
Hierarchical linear modeling and Cohen f^2 effect size	Variance explained by purposely selected hierarchical factors (e.g., behavior, subject, treatment). The Cohen f^2 effect size can be computed by comparing a target model with a reference model.	Range, $-\infty$ to $+\infty$; small, 0.02 to 0.15; medium, 0.15 to 0.34; large, ≥ 0.35	Lorah (2018), Noorgate & Onghena, (2003)

Table 6.4 Standards of selected single-subject analytic techniques.

	NAP	PAND	Tau-U	LRR	WC-SMD	BC-SMD	MVA RT	HLM
<i>Feasibility & acceptability</i>								
Calculation software available	●	●	●	●	●	●	●	●
Interpretation guidelines available	●	●	●	●	●	●	●	
Minimal training needed	●	●	●		●		●	
Value proportional to magnitude	●		●	●	●	●		●
Widely used		●			●			
<i>Sensitivity to single-case datasets</i>								
All data-points computed	●			●	●	●	●	●
Amenable to zero/near-zero baselines	●	●	●	●		●	●	(●)
Amenable to all major single-subject designs	●	●	●				●	●
Amenable to short time-series	●	●	●	●				●
Amenable to varying length time-series		●	●	●		●	●	●
Some robustness against autocorrelation						●		●
Some robustness against local trends				●				(●)
Some robustness against within-phase trends				●				●
Some robustness against the effect of outliers			●	●	●			●
Weights no. of between-subject replications						●	●	●
Weights no. of within-subject replications						●		●
Weights steady-state responding								(●)
<i>Statistical standards</i>								
Allows for the assessment of co-variables						●	●	●
Amenable to inferential statistics						●	●	●
Meta-analysis/moderator analysis possible	●			●	●	●	●	●
Does not assume normality	●	●	●	●	●		●	●
Does not rely on central-tendency	●	●	●	●	●	●	●	●
Provides effect size variance				●	●	●	●	●
Minimal floor/ceiling effects			●	●	●	●		●
Confidence intervals/ <i>p</i> values provided	●		●	●	●	●	●	●

Notes. Major single-subject designs include reversal, multielement, multiple-baseline, and changing-criterion designs. Parenthesis indicate that the standard can be met if modelled. Local trends denote carryover effects, multiple-treatment interference, and phase transition trends. BC-SMD = Between-case standardized mean difference; HLM = Hierarchical linear modeling; LRR = Log-response ratio; MVA = Masked visual analysis; NAP = Non-overlap of all pairs; PAND = Percentage of all non-overlapping data; PND = Percentage of non-overlapping data; RT = Randomization test; WC-SMD = Within-case standardized mean difference.

Table 6.5 Design matrix for one participant.

Name	Participant	Treatment status (tts)	Session	Behavior (beh)
Bob	1	0	1	7.21
Bob	1	1	2	21.32
Bob	1	0	3	8.52
Bob	1	0	4	6.53
Bob	1	0	5	7.26
Bob	1	1	6	12.01
Bob	1	0	7	3.17
Bob	1	0	8	9.10
Bob	1	0	9	8.91
Bob	1	1	10	11.44
Bob	1	0	11	2.14
Bob	1	0	12	4.52
Bob	1	0	13	1.39
Bob	1	1	14	6.24
Bob	1	0	15	3.06
Bob	1	0	16	1.43
Bob	1	0	17	3.55
Bob	1	1	18	14.72
Bob	1	0	29	4.70
Bob	1	0	20	7.65

Figure 6.1 Some graphical examples of common visual analysis scenarios.

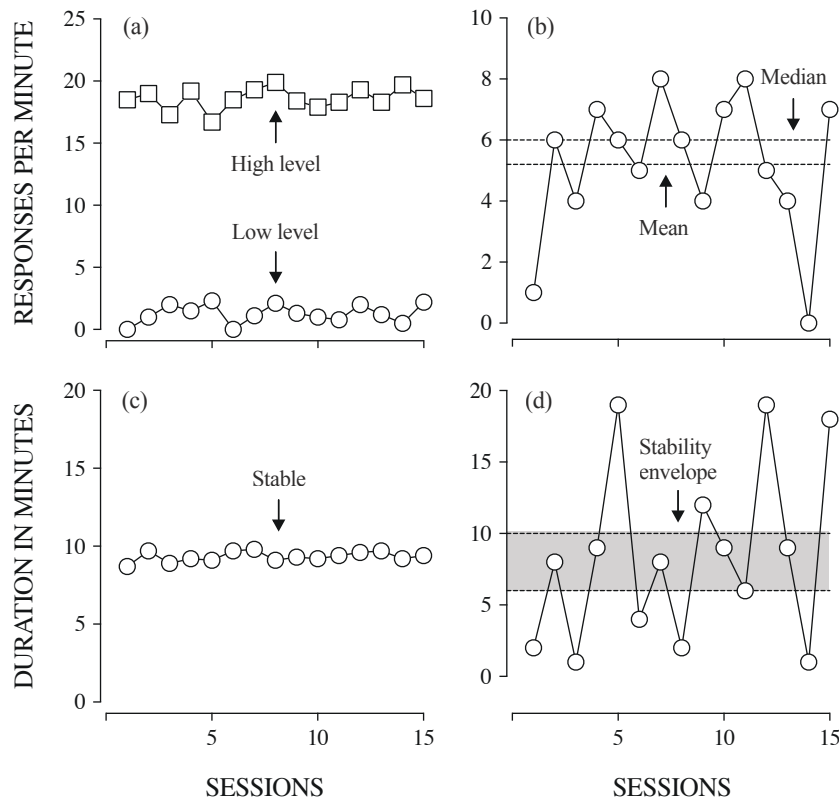


Figure 6.2 Local trend (1st B phase), and non-reversibility (2nd A phase, lower graph).

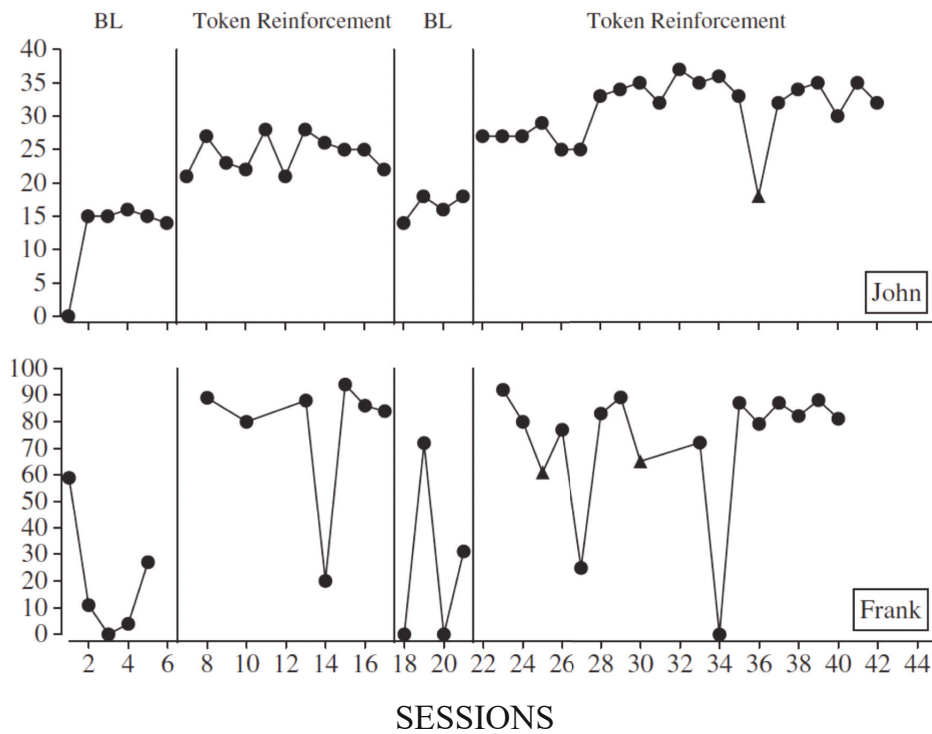
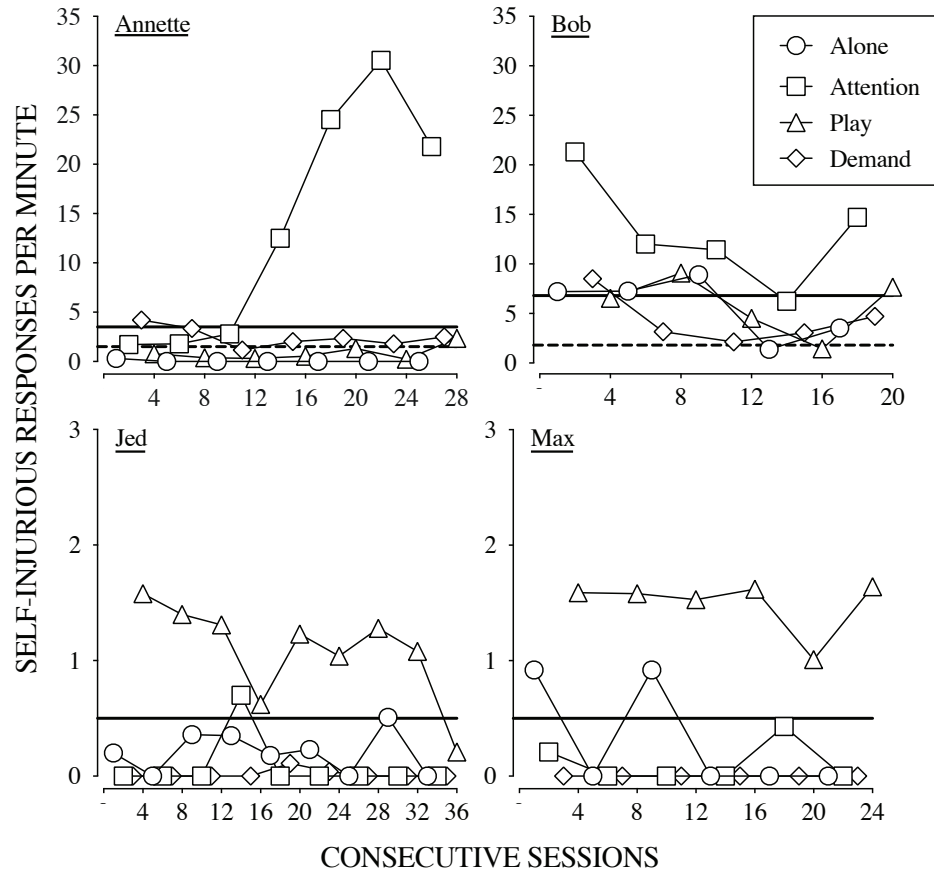


Figure 6.3 Cases of attention- (Bob and Annette) and escape-maintained (Max and Grey) self-injury.



Notes. Solid and broken horizontal lines denote the upper and lower criterion lines for the play condition (Hagopian et al., 1997). Originally re-graphed data (Source: Conners et al., 2000).

Figure 6.4 R code for a two-level hierarchical linear model with lag1 autocorrelation and heterogeneity of variances.

```
install.packages ("nlme")
library("nlme")
reg1_AR_HE<-lme (beh ~ tts, random=~tts |name,data=raw, na.action="na.omit",
                correlation=corAR1(form=~ 1|name),weights = varIdent(form = ~1 | tts),
                control=list(opt = "optim",
                             optctrl = list(method = "REML ")))
summary (reg1_AR_HE)
```

In order to run the two-level hierarchical linear model, the R package “nlme” needs to be installed [`install.packages("nlme")`] and loaded [`library("nlme")`]. The package “nlme” allows running a linear mixed effects model. The function “lme()” is called and between the brackets the specific model is defined, together with some optional arguments. The behavior outcome score (“beh”) is regressed on the dummy coded treatment variable (“tts”, or the dummy coded variable indicating the undifferentiated versus differentiated condition). The intercept and “tts” are allowed to vary between participants (indicated by `random =~tts/names`). The argument “data=raw” indicates that we are using the data saved in the dataframe “raw”. The next argument “na.action=” is needed in case there are variables with a missing score (na = not applicable). The “correlation=corar1” argument indicates that we are modelling a lag1 autocorrelation per participant (“|name”). The “weights” defines that the within-participant variance is heterogeneous (i.e., a different amount of variance in the treatment versus non-treatment condition is modeled). Lastly, the control and optCtrl define that the restricted maximum likelihood estimation approach is used (which is recommended when working with a small number of observations). The obtained output is saved in “reg1_AR_HE” (but any other name can be chosen) and a display of this output is obtained by running “summary(reg1_AR_HE)”. The obtained output is included in Figure 5.

Figure 6.5 R output for a two-level hierarchical linear model with lag1 autocorrelation and heterogeneity of variances.

```

Linear mixed-effects model fit by REML
Data: raw
AIC          BIC          logLik
479.0524     500.3599     -231.5262
Random effects:
Formula: ~tts | name
Structure: General positive-definite, Log-cholesky parametrization

              StdDev          Corr
(Intercept)      2.443743      (intr)
tts              5.029188      0.579
Residual         1.347879
Correlation Structure: AR(1)
Formula: ~1 | name
Parameter estimate(s):
Phi
0.2855608
Variance function:
Structure: Different standard deviations per stratum
Formula: ~1 | tts
Parameter estimates:
0          1
1.000000   5.131597
Fixed effects: beh ~ tts
              Value          Std.Error    DF          t-value          p-value
(Intercept)      1.672828      1.236072    103      1.353341      0.1789
tts              5.858234      2.825021    103      2.073696      0.0406
Correlation:
(Intr)
tts 0.576
Standardized within-Group Residuals:
Min          Q1          Med          Q3          Max
-2.84792960   -0.21736258 -0.09689296   0.08955786   2.87216994
Number of Observations: 108
Number of Groups: 4

```

The first part of the output provides goodness of fit indices (i.e., AIC, BIC and the Log Likelihood). These indices can help to choose between multiple competing nested models. The smaller the values for the fit indices, the better the model fit. The hierarchical liner model is very flexible. For instance, different functional forms for the trend (e.g., linear trend, quadratic trend, etc.) and/or autocorrelation can be specified. The model can also be expanded by including participant-specific covariates (e.g. gender, age, etc.). Competing models can be compared using these fit indices. The next part of the output indicates the random components estimates, i.e., the estimate of the between-participant variance of behavior level in the non-differentiated condition (i.e., $\sigma_{u_0}^2 = 2.44^2$), the between-participant variance of changes in behavior level (i.e., $\sigma_{u_1}^2 = 5.03^2$), the covariance between the non-differentiated and differentiated condition ($\sigma_{u_0u_1} = 2.44 \times 5.03 \times 0.60$), and the within-participant variance. The within-participant variance is estimated separately for the non-differentiated (i.e., $\sigma_e^2 = 1.35^2$) and the differentiated treatment condition [i.e., $\sigma_e^2 = (5.13 \times 1.35)^2$]. This indicates that most variance is due to difference in behavior level between-participants, and within the treatment condition. The next lines in the output provide an estimate of the lag1 autocorrelation (Phi = 0.29). This value indicates that there is a decent amount of correlation between two consecutive errors. Next, the fixed effect parameter estimates are provided. The estimated behavior level during the non-differentiating condition is 1.67 and not statistically significantly different from 0 [$\hat{\theta}_0 = 1.67$, $t(103) = 1.35$, $p = .18$]. The difference in behavior level between the non-differentiating and differentiating condition is estimated to be 5.86, and is statistically significant at the .05 level [$\hat{\theta}_1 = 5.86$, $t(103) = 2.07$, $p = .04$]. The within-participant standardized residuals indicate that the median error is close to 0, and that there are no significant outliers (as the standardized errors range from -2.85 to +2.87). The last line of code indicates that there is a total of 108 observations, nested within 4 participants.

CHAPTER 7. Accuracy of Paper-and-Pencil Systematic Observation versus Computer-aided Systems⁵¹

ABSTRACT

Computer-aided behavior observation is gradually supplanting paper-and-pencil approaches to behavior observation, but there is a dearth of evidence on the relative accuracy of paper-and-pencil versus computer-aided behavior observation formats in the literature. The current study evaluated the accuracy resulting from paper-and-pencil observation and from two computer-aided behavior observation methods: The Observer XT® desktop software and the Big Eye Observer® smartphone application. Twelve postgraduate students without behavior observation experience underwent a behavior observation training protocol. As part of a multi-element design, participants recorded 60 real clinical sessions randomly assigned to one of the three observation methods. All three methods produced high levels of accuracy (paper-and-pencil, $.88 \pm .01$; The Observer XT, $.84 \pm .01$; Big Eye Observer, $.84 \pm .01$). A mixed linear model analysis indicated that paper-and-pencil observation produced marginally superior accuracy values, whereas the accuracy produced by The Observer XT and Big Eye Observer did not differ. The analysis suggests that accuracy of recording was mediated by the number of recordable events in the observation videos. The implications of these findings for research and practice are discussed.

Keywords: Accuracy, Behavior observation, Computer-aided observation

51 El presente estudio ha sido publicado en la revista *Behavior Research Methods* y se presenta en el idioma y el formato en el que se publicó.

Virues-Ortega, J., Casas, C. D., Martin, N., Tarifa-Rodriguez, A., Hidalgo, A., Cox, A. D., & Navarro Guzmán, J. I. (2022). Accuracy of paper-and-pencil systematic observation versus computer-aided systems. *Behavior Research Methods*. <https://doi.org/10.3758/s13428-022-01861-0>

INTRODUCTION

Focus on direct observation of behavior is a key method for experimental psychology, clinical psychology, education, comparative psychology, ethology, behavior analysis and numerous other fields (Behavior Analyst Certification Board, 2017; Bonnie, 2017; Haynes & O'Brien, 2020; Hintze, 2004; Yasukawa & Bonnie, 2017). Examples of the range of behaviors that are commonly observed by researchers include classroom teaching (Gargani & Strong, 2014), parent-child interactions (Hudson & Rapee, 2001), patient-doctor interactions (Hayward et al., 2015), and animals across a wide range of settings and species (e.g., Bailey et al., 2004). Systematic behavior observation is paramount to the assessment and treatment of clinically important behavior such as sleep problems (Roth et al., 1976), feeding disorders (Piazza, 2008), and problem behavior of children, adolescents and adults (Hanley et al., 2013). Access to mobile and computer-based behavior observation applications has made the use of such technology near-universal yet research on its ability to enhance the accuracy of behavior observation has remained rare.

Numerous software-based applications for animal and human behavior observation have become available over the last few decades: Big Eye Observer[®] (ABA España, 2019), CATOS (Oh & Fitch, 2017), Countee (Gavran & Hernandez, 2020), EthoVision XT[®] (Noldus Information Technology, 2021), JWatcher (Blumstein et al., 2006), ObsWin (Martin et al., 1999), Solomon Coder (Péter, 2019), The Observer XT[®] (Noldus Information Technology, 2019), ZoneMinder (Farrimond et al., 2009), to mention just a few examples (also see early applications in Bass, 1987, and Kahng & Iwata, 1998). In the early 2000s, studies began to report the use of desktop and hand-held computer applications for conducting systematic observation. Jackson and Dixon (2007) used an application created with Microsoft Visual Basic for recording data during direct observation of functional analysis sessions using a hand-held computer. The software produced frequency- and interval-based observations with output files that could be imported into statistical and graphing software packages. Crowley-Koch and Van Houten (2013) described a variety of technological solutions aimed at facilitating the collection of data via direct observation, including video-synced software applications and internet-based applications. Software, such as The Observer XT (Noldus Information Technology, 2019), has been used widely in studies involving systematic observation in the areas of animal learning and experimental psychology (e.g., Franchi et al., 2016), organizational and consumer behavior (e.g., Allen et al., 2015), clinical interaction in psychotherapy (e.g., Virues-Ortega et al., 2011; Pardo-Cebrian et al., 2021), intervention studies in children and adults with and without developmental and intellectual disability (Hutman et al.,

2012; Meirsschaut et al., 2011; Mossman, 2011; Naber et al. 2008), and studies in dementia (Moyle et al., 2014) among numerous other applications. These studies show the range of applications of such systems, and they have all consistently reported high levels of interobserver agreement for computer-aided observations.

Several authors have noted the advantages of computer-aided observation over more traditional methods, highlighting the convenience of electronic data storage and analysis, usability of computer interfaces, and discreteness of hand-held devices (e.g., Tarbox et al., 2010). In spite of the apparent advantages of computer-aided data collection systems, very little research has empirically evaluated their impact on the accuracy of the data collection process (Kahng & Iwata, 1998; Jackson & Dixon, 2007; Tarbox et al., 2010).

Wessel (2015) has described the advantages of computer-assisted direct-observation apps, relative to desktop computer-based systems, which may allow for accurate real-time coding of events *in vivo*. Real-time observation apps are highly accessible via smartphones and simplify the ethical concerns derived from video-based observation. They also provide the observer with the opportunity to record contextual information that could be easily missed in video-based observations. A key consideration when comparing *in vivo* observation with recorded retrospective video-based coding is the ability, with the latter, to use video playback functions (i.e., pausing, rewinding, use of slow motion, advancing frame-by-frame) as these may have the potential to optimize accuracy, particularly during eventful sessions. Yet, the potential impact of session busyness on accuracy in terms of *in vivo* versus video playback observation modes have not been evaluated in the literature.

Very few studies have compared the differences between paper-and-pencil and computer-aided systematic observation. In a notable exception, Tarbox et al. (2010) compared these in the context of services being provided to children diagnosed with autism spectrum disorder. A total of four participants used either paper-and-pencil or a hand-held computer to observe across a range of sessions. Interestingly, the results suggested that computer-aided data collection was less time-efficient than paper-and-pencil methods. The accuracy of paper-and-pencil and computer-aided observation ranged from 98-100% and from 84-95% respectively, suggesting slight superiority of paper-and-pencil observation. The authors suggested that in this particular arrangement observers required less time to record their observation with the paper-and-pencil format, whereas the computer-aided format required the observer to sequentially select multiple

keys and screens. While data collection with the hand-held computer application was somewhat slower, it did allow for automated graphing. The findings from Tarbox et al. (2010) require replication with alternative computer-aided observation systems, including those with simplified interfaces. In addition, participants in Tarbox et al. (2010) recorded live sessions. Therefore, it was not possible to discard other factors that may have had an effect on the accuracy of recording such as the number of events recorded and the emission rates of those events.

Given the scarcity of studies devoted to the evaluation of data collection systems for behavior observation, a systematic comparison of some of the different formats of observation would be highly informative. In the current study we evaluated the relative accuracy of using the paper-and-pencil method, desktop computer software that allowed retrospective video analysis (The Observer XT), and a mobile app (Big Eye Observer). The current selection of systems intended to capture the range of platforms and capabilities of existing systems (see Table A in the Supplementary Online Material for a comparison of representative systems). The study adds to the literature by using a comparable set of observation sessions across participants, controlling for the participants' behavior observation training and experience, using a relatively larger sample size than has been reported in previous studies, and accounting for the number of recordable events in a session and the observers' use of video playback functions.

MATERIAL AND METHODS

Participants and Setting

Twelve female students (mean age 23.2 years, range 20-41) without past experience in behavior observation and enrolled in a postgraduate course in applied psychology in New Zealand participated in the study. Students enrolled in the course ($n = 16$) received an email inviting them to participate (recruitment period, April 1, 2019 through April 30, 2019). Observation sessions took place in a quiet laboratory space on university premises and at the participants' homes. To prevent observer bias, participants were blind to the goals of the study and did not receive performance-related feedback over the course of the study (Lerman et al., 2010). One individual declined to continue to participate shortly after the study started (P7). One additional participant did not follow the expected order of sessions resulting in their data being excluded from further analysis (P10). We conducted Monte Carlo simulations to study *a priori* the power of the intended mixed model analysis and sample size (Gelman & Hill, 2008). The achieved power was above

.90 for both a sample size of 10 and 15 subjects, assuming a .05 mean difference in accuracy across the levels of the main fixed-effect factor (observation method) and 60 successive participant-nested measurements. Sample size was established *a priori* and was not subsequently modified.

The study protocol was approved by the social sciences ethics committee of the University of Manitoba (Canada), the human ethics committee of The University of Auckland (New Zealand), and the ethics committee of the University of Cadiz (Spain). All participants signed an informed consent form. The current report adheres to the TREND statement (Jarlais et al., 2004). Dataset files are available from Virues-Ortega, Delgado Casas, et al. (2022).

Materials

Observation videos. Participants observed 60 distinct 5-min videos over the course of the study. All videos had a resolution of 720 x 480 pixels and a frame rate of 25 frames per second. The video collection was originally obtained as part the study by Cox and Virues-Ortega (2021). The videos portrayed actual demand sessions from experimental functional analyses and featured a range of problem behaviors and clients. The functional analysis sessions presented in the videos followed the procedures described by Iwata et al. (1982/1994) with the procedural adaptations specified by Cox and Virues-Ortega (2021). Demand sessions were chosen because there were a reasonably high number of recordable events in all sessions – observation sessions with a relatively low number of events can produce artificially high levels of agreement/accuracy when using the block-by-block method of agreement (cf. Page & Iwata, 1986; Mudford et al., 2009) that was used in this study. Target events included *compliance, praise, demand*, and each video featured one of a number of distinct problem behavior topographies (e.g., self-injurious, destructive, aggressive behaviors) – definitions of all target events are available consistent with the taxonomy by Ray et al. (2011) are available on request from the corresponding author. The mean number of recordable events across each 5-min video was 62.5 (range 20-134). Data from each of the participants based on the observation of four specific videos were excluded from the analysis after the study was completed – the camera angle of these particular videos made some of the target events visually ambiguous.

Paper-and-pencil observation. Participants used observation data sheets with 30 separate rows for each of the 10-s intervals required for recording a 5-min video. Each datasheet had columns indicating the interval number, interval start and end, and separate columns for each

target event. Participants used VLC Media Player, version 3.0.6 (VideoLAN Association, 2019) to play the videos to be recorded using paper-and-pencil. The VLC Media Player was chosen because time elapsed and time remaining counters are clearly visible during playback and also to standardize the way in which the videos were displayed. Additionally, the media player had the specific video playback functions that could be used.

The Observer XT. Participants used The Observer XT version 14.2 observation software. The observation module displays the session's video, a video timeline, keys for each observation code (i.e., target events) and video playback functions. The research team provided participants with instructions to create an event observation coding scheme that would include all relevant target events to be recorded as discrete events (point events). The research team subsequently verified that the observation coding schemes were set up correctly. Output files in text format were produced for further analysis.

Big Eye Observer. The Big Eye Observer iPhone®/iPad® application is a single-screen systematic observation app that allows the recording of up to 12 distinct events concurrently using a variety of methods (i.e., frequency/count, partial interval, total interval, duration). The research team provided participants with instructions to create an event observation template that would include all relevant target events. The research team subsequently verified that the observation templates were set up correctly. All target events were recorded as discrete events (frequency events). Output files were subsequently emailed in text format for further analysis.

Criterion Reference for Videos

All videos were observed and data collected on the four target events in each video independently by two trained observers with over 10 years of experience in systematic observation. Observers used a Microsoft Excel spreadsheet to input their observations. After completing the observations, the observers checked their agreement in each of the videos on an interval-by-interval basis with the videos divided up into 10-s intervals. Any disagreements on the number of events for each of the four target events was discussed and consensus reached. Most disagreements had to do with events occurring around the start or end of a 10-s interval, events occurring in quick succession, or events that were difficult to discriminate (visually or audibly) and which often required frame-by-frame replay until consensus could be reached. The resulting criterion reference recordings provided the agreed number of recordable events for each interval within each video and were used to assess the accuracy of the data collected by participants from

all observation sessions using one of the three different observational recording methods (paper-and-pencil, The Observer XT, Big Eye Observer).

Accuracy

Accuracy was calculated using the block-by-block method of analysis. The 5-min videos were divided into 30 10-s intervals and each 10-s interval was scored by dividing the smaller number of events recorded in each interval by the larger number of events, or scored as 1 if the number of events were identical for the participant and reference recording. This process was repeated for each of the four target events recorded. An accuracy index value for each session was then calculated by summing the scores for each of the 30 intervals for each of the four target events and dividing by 120 (the total number of assessed intervals).

Design

Participants conducted 20 observation sessions for each of the observation methods (paper-and-pencil, The Observer XT, Big Eye Observer) totaling 60 sessions and these were randomly alternated as part of a multi-element experimental design embedded in a within-subjects design (Kazdin, 2011). In order to evaluate whether observation methods would yield differential results under (simulated) *in vivo* conditions or when able to use the video playback functions, we used the ability to use video playback functions as a secondary independent variable. Specifically, participants were told either that they could use video playback functions (*video playback phases*) or to refrain from using the playback functions (*in vivo proxy phases*) across four successive 15-session study phases. Each phase included five sessions for each observation method randomly sequenced within the 15-session phase. Phases were alternated as part of an ABAB reversal design. In order to control for the effects of phase order, participants 1, 3, 5, 9 and 11 (P1, P3, P5, P9, P11) initiated the reversal design with the video playback phase, whereas P2, P4, P6, P8, and P12 started with the *in vivo proxy phase*.

Procedure

Participants had no past experience with observation systems. All participants attended a session with the principal investigator where the materials of the observation training protocol by Dempsey et al. (2012) were presented. This protocol contains six 10-min videos showing a range of problem behavior and environmental events (instructions, praise, demand, attention). Each video has an increasing number of events and behavior codes relative to the preceding one.

For example, video #1 has a single target behavior and a total of 30 recordable events, whereas video #6 has six distinct target behaviors and a total of 178 recordable events. Participants had to record the target behaviors from each video but were instructed to advance to the next video only if they had reached an accuracy index of .90 or higher for all target behaviors. Lower accuracy would result in the participant repeating the observation and recording for that video. Participants had a mean accuracy of .97 (range .94-.99) in their last attempt in all training videos, required a mean 8.6 sessions to attain the criterion for the six videos (range 6-13), and required a mean training time of 83.8 min (range 60-130). Participants took between 1 and 4 days to complete the training. All data collection occurred using printed observation data sheets.

After completing the behavior observation training protocol, participants received written instructions, a one-hour video-tutorial, and a two-hour workshop led by the principal investigator. The written instructions primarily covered the process of accessing the study materials, how to use the relevant observational tools, and the operational definitions of the target events. The video-tutorial was on the use of VLC Media Player, The Observer XT and Big Eye Observer for the purposes of the study. The workshop was intended to provide hands-on training for using the three systems and to troubleshoot any technical or logistical difficulties. Participants were able to contact the research team throughout the study via phone or email to have their questions or concerns addressed.

Participants were allowed to use playback functions (i.e., pausing, rewinding, use of slow motion, advancing frame-by-frame) during video playback phases. This was also true for the Big Eye Observer app, which had an option to “pause recording.” While the app was paused, participants would have been able to use video playback controls in the video player. The only caveat of using this approach with the app was that participants would need to sync the video time counter with that of the app when resuming the observation. The video playback functions were integrated in The Observer XT interphase.

Participants conducting the study at the university had access to a desktop computer with VLC Media Player and The Observer XT installed, and a 9.7-inch iPad with the Big Eye Observer installed. In addition, participants conducting part of the observation at home received the temporary loan of a laptop computer with VLC Media Player and The Observer XT installed, and an activation code for the Big Eye Observer app to run on their iPad or iPhone. Participants con-

ducting any observation from home that did not own an iPad or iPhone were loaned a sixth-generation 9.7-inch iPad with the Big Eye Observer installed.

Participants were instructed to use the VLC Media Player during both their paper-and-pencil and Big Eye Observer observation sessions and asked to keep the relative size of the video window on their screen constant when using all three observation methods. Participants were instructed to conduct observations in a quiet environment and to keep a handout with the operational definitions of the target behaviors within their reach for easy reference. The observation data sheets used during paper-and-pencil observation sessions were subsequently scanned and the participants received a personalized secure link to upload the scanned files and The Observer XT and Big Eye Observer output files.

Participants conducted the observation in bouts that typically included 10 to 15 observation videos with minimal breaks in between. Participants engaged in these observation bouts across several days until the study protocol was completed. The mean number of successive calendar days to complete the study protocol was 9.4 (range 1-21).

Procedural integrity

To evaluate whether study procedures were followed as intended, participants were asked to fill out a form as they progressed through the study noting the sequence, start time, and end time of behavior observation sessions. The percentage of observation sessions conducted in the expected sequence was calculated, the percentage of missing sessions, and the percentage of duplicate sessions. The percentage of observation sessions completed using the expected sequence was 100% for all participants. The mean percentage of missing sessions across participants was only 2.7% (range 0-16.7%). Only two duplicate sessions were identified in the dataset, both produced by P3, and only the session recorded first was included in the analysis and the duplicate sessions were discarded. Neither an iPad nor iPhone could be secured for P5 and so only paper-and-pencil and Observer XT sessions were completed. P11 completed all sessions but failed to properly upload the data files for 10 of the Observer XT sessions.

Because using the video playback functions during those phases was dependent on the participants following the instructions provided, session start and end times were extracted from The Observer XT and Big Eye Observer data outputs and from the paper-and-pencil data sheets such that the total session time spent recording each video could be calculated. This helped to

indirectly verify whether the participants used any playback functions and whether playback functions were only used during the expected phases (i.e., session lengths during the video playback phases should have been consistently longer than the length of the 5-min video). Figure 1 presents the mean observation time for the *in vivo* proxy and video playback phases across observation methods. The mean observation time during *in vivo* proxy phases for each observation method suggests that participants did not use the video playback functions while using Big Eye Observer (302.00 ± 0.78 s), but did use them (albeit sparingly) during the paper-and-pencil (328.48 ± 6.52 s) and Observer XT sessions (327.95 ± 7.22 s). Interestingly, a similar pattern was observed in The Observer XT (336.16 ± 9.58 s) and Big Eye Observer sessions (304.91 ± 3.66 s) during the actual video playback phases. The observation time data seems to suggest that participants only used the intended video playback functions comprehensively for the paper-and-pencil sessions (436.04 ± 19.78 s).

Data Processing

The final dataset included approximately 600 datasheets in three different formats. All data files were reviewed manually for any missing or duplicate sessions. All data were either tabulated or imported into Microsoft® Excel spreadsheets. An Excel Visual Basic for Applications script was developed to transform all files into a single format, import all into one spreadsheet, extract the relevant data, compute the number of events for each of the four target events for each 10-s interval for each 5-min session, and conduct the accuracy analyses (i.e., compare the participants' data with the criterion reference for that session).

Statistical Analysis

The time-series nature of the data, lack of a normal distribution, and the presence of missing values supported using mixed linear models with the current dataset. Participant number was used as the subject variable, session number as the time-based variable, and accuracy as dependent variable. A first-order autoregressive covariance structure rendered the best goodness of fit values during the model development process ($\rho = -0.15 \pm 0.07$). Observation method (paper and pencil, The Observer XT, Big Eye Observer), number of recordable events, video mode experimental phase (video playback, *in vivo* proxy), and their interaction were added as fixed-effect factors. Maximum Likelihood Estimation was used to determine the model's parameters and pairwise comparisons were computed across observation methods and video mode experimental phase. Pairwise coefficients were expressed using the same metric as the dependent variable

(range 0-1). Factors and co-variables failing to improve goodness of fit by a minimum of two units using Akaike's Information Criterion (AIC) were excluded from the model (Akaike, 1974; Burhan & Anderson, 2002). This resulted in phase order, reversal design order, and observation time not being included in the final mixed model analysis as either factors or co-variables. The model goodness of fit was optimized further (AIC, -1133 vs. -908) by adding method and video mode experimental phase as participant-nested factors, and recordable events nested by observation video (i.e., video number). All analyses were conducted with SPSS® IBM® Statistics, version 27 (IBM Corporation, 2021). A p value of .05 was used throughout with Bonferroni adjustments for multiple comparisons.

RESULTS

Figure 2 shows the mean accuracy data across participants disaggregated by phase and observation method. A preliminary visual analysis of these data suggested relatively high levels of accuracy across all three methods (paper and pencil, $.88 \pm .01$; The Observer XT, $.84 \pm .01$; Big Eye Observer, $.84 \pm .01$) although slightly better accuracy for the paper-and-pencil method. This effect was also apparent when the data are aggregated across participants on a session-by-session basis (Figure 3). However, due to high levels of variability across participants it was not possible to ascertain consistently different levels of accuracy across specific observation methods or experimental phases through the visual analysis of individual participants' data – session-by-session graphs for each participant are available on request from the corresponding author.

The final mixed model analysis (Table 1) confirmed a significant effect of observation method, $F(2, 543) = 16.022, p < .001$. Pairwise comparisons indicated that both The Observer XT and Big Eye Observer produced slightly lower levels of accuracy, relative to paper-and-pencil observations. Paper and pencil vs. The Observer XT, coefficient = $.035 \pm .007, p < .001$; paper-and-pencil vs. Big Eye Observer, coefficient = $.039 \pm .007, p < .001$. Thus, accuracy decreased by approximately .04 when switching from paper-and-pencil observation to computer-aided observation systems. There were also significant fixed effects for the number of recordable events, $F(45, 526) = 5.308, p < .001$, and the interaction of this with observation method, $F(88, 527) = 1.844, p < .001$. Figure 4 shows a scatterplot with Spearman-rank correlation analysis suggesting that the mediating effect of the number of recordable events was observed primarily in The Observer XT ($R_s = -0.15, p = .021$) and Big Eye Observer ($R_s = -0.24, p = .001$) observation methods.

Increased observation time in only the paper and pencil sessions during the (expected) video playback phases (Figure 1) is a potential confound to the slightly superior assessed accuracy of paper-and-pencil sessions. However, the video mode experimental phase (video playback and *in vivo* proxy) was not established as a significant fixed-effect factor in the mixed model analysis, $F(1, 263) = 3.016, p = .084$. However, it reached statistical significance when included as a participant-nested factor, $F(9, 256) = 3.068, p = .002$, suggesting that video mode may have a mediating role over accuracy for some individuals. The pairwise comparison indicated a trend toward increased accuracy of the video playback mode, coefficient = $.011 \pm .006, p = .065$. Finally, adding observation time as a random-effects co-variable had deleterious effects in the model's fitness.

DISCUSSION

Assessment, treatment planning, implementation and evaluation depend heavily on the ability to be able to collect meaningful and robust data through behavior observation, and on the accuracy of those data. Therefore, it is important to maximize the efficiency of behavior observation methods while collecting data reliably and accurately. The rationale for assessing the reliability of data collected in practice is three-fold: (a) to check the consistency of observations; (b) to minimize bias; and (c) to verify the adequacy of response definitions (Kazdin, 2011). In the current study, we compared the accuracy of data collected using traditional paper-and-pencil methods with two computer-aided methods. Block-by-block agreement, one method used to assess the reliability of behavior observation data, was utilized to assess the accuracy of the different systematic observation formats. Specifically, we used the same analysis to compare the data recorded by the participants using different methods against a pre-determined criterion reference for each dataset. Our findings suggest that formally-trained observers can reach high levels of accuracy with a range of behavior observation methods with minimal accuracy loss that could be attributed to observation difficulty.

The current study suggests that the use of paper-and-pencil data collection still results in marginally higher levels of accuracy and appears to replicate the findings of Tarbox et al. (2010). The question remains as to whether this marginal difference is of any clinical importance. It would be unwarranted to establish a criterion of clinical significance without consideration of the specific behaviors of interest and their baseline rates. For example, a .04 difference in accuracy may be clinically important for low-frequency aggressive behavior, whereas it might not be for

high-frequency stereotypy. In order to provide additional context, Figure 5 presents the distribution of inter-observer agreement (IOA) values (used here as an analogue for accuracy) obtained from a selection of the functional analysis literature as reported in Virues-Ortega, Clayton, et al. (2022). This ancillary analysis suggests that a range of .04 accounts for one standard deviation and a non-trivial one fifth of the range of usable values reported in this sample of the literature (0.8-1.0). Moreover, IOA values ranging .04 from one another have a cumulative probability of being found in the literature of up to 43%, again underlying that a .04 difference in accuracy may not be trivial. However, given the fact that all methods yielded accuracy scores greater than .80, where the general convention is to accept agreement/accuracy scores of .80 or greater (Cooper et al., 2019), the difference may be anecdotal for practical purposes.

The current findings result from an analysis of a dataset derived from the observational data collected from a diverse range of video sessions that include multiple target behavior topographies (self-injurious, destructive, aggressive behaviors) and a range in terms of the number of occurring events. It is surprising that the data also suggest that paper-and-pencil observation is still likely to be more accurate when the number of recordable events is relatively high (a proxy metric for behavior observation difficulty in this context). Future analyses should evaluate alternative indicators of behavior observation difficulty such as number of concurrent target behaviors and operational definition complexity. We controlled for participant observation experience and training, which are known confounds of observation performance (Mash & McElwee, 1974; Wildman et al., 1975). However, it is unclear whether our findings would be typical of more experienced observers, or whether behavior observation practice could effectively mediate the difficulty-accuracy relation. The potential impact of observation difficulty should be evaluated experimentally and not just as a *post hoc* correlational analysis, and such findings should be replicated with more varied and extensive samples of behavior.

One factor that may have influenced the apparent superiority of paper-and-pencil observation may have been the familiarity and simplicity of using a basic tool relative to the complexities involved in learning how to use a computer-based system. It should be noted that participants used the paper-and-pencil method during the behavior observation training protocol and received no performance-based training for the other two methods. We could start to address this concern by examining the aggregated session-by-session performance for each of the three methods over time. Specifically, Figure 3 does not reveal an apparent ascending trend (which

would be suggestive of a learning effect) for any of the three methods (see a trend-stationarity test in the Supplementary Online Material, Table B, for further details).

It is also possible that computer-aided observation required a larger number of in-session responses to complete key routines including recording an event, deleting an incorrectly recorded event, and re-playing a specific video section. Engaging in longer chains of responses increases the potential for error (see Podofillini et al., 2013 for an empirical analysis of the linear relation between task complexity and operator error probability). In addition, increased effort may have influenced compliance with study procedures, particularly the intended use of video playback functions (see Hinz et al., 2014 for an applied demonstration of the relation between response effort and compliance in the context of behavior observation). Additional research may try and standardize the response effort and complexity of using any system in order to truly evaluate its utility and rule out such a confound.

Overall, without evidence that all computerized devices and the software systems, and their human-computer interfaces, are equivalent, which seems extraordinarily unlikely, the potential impact of the present work may be seen as restricted. In order to fully address this concern, it would have been necessary to conduct a detailed analysis of all existing systems or a component analysis to isolate the effect of common functionalities and interphase components. Either approach would have been impractical. Instead, we selected two considerably diverse systems such that, would convergent findings be obtained, a modest indication of generality would seem plausible.

Despite the apparent but marginal superiority of the paper-and-pencil method of collecting observational data suggested by this study (albeit that all methods yielded accuracy scores within accepted limits), there remain some ostensibly clear advantages to the use of computer-based systems that should be systematically evaluated, both in terms of practical application as well as outcome data. For example, the ability to automatically generate output files, produce descriptive summary statistics and graph the data may well offset the marginal loss in apparent accuracy. By contrast, the use of paper-and-pencil observation remains a well-established approach requiring minimal staff training and resources, which may be critical advantages for low-resource communities.

As the complexity of computer-based systems for observational data collection increases, as seems inevitable with technological advances occurring exponentially and ubiquitously, there

is a trade-off between the various features offered by a product and the basic ability to observe and collect observational data. Further research could evaluate these and other factors by, for example, controlling for complexity of use (e.g., the length of time and/or the number of key presses used for recording). In this context it is interesting to note that the overall observer accuracy for the two software systems utilized here (The Observer XT, Big Eye Observer) was almost identical in spite of the numerous differences in their respective user interphases. Specifically, synced video functionality, number of discrete actions needed to complete key routines (e.g., score, delete, playback), and number of attention shifts required differed across the two observation systems.

Technological developments over the last few decades have provided scientists with a diverse set of observation tools with a concomitant impact on efficiency of use and accuracy of recording (Hall et al., 2014; Sarkar et al., 2006). Behavior observation capabilities have been improved by way of bespoke software for handheld and desktop devices (e.g., McKerchar & Abby, 2012; Virues-Ortega et al., 2011), sometimes incorporating the facility for video processing and retrospective coding (e.g., Hall et al., 2014). Studies using computer-aided observation often report high levels of interobserver agreement and/or accuracy but direct comparisons are lacking. In addition, behavior-analytic studies rarely report the technology supporting the behavior observation process. Therefore, more systematic replications are needed to evaluate any favorable or deleterious effects that may be caused by computer-aided observation.

CONCLUSIONS

The current analysis makes it possible to draw some tentative conclusions on the various methods utilized. First, our study suggests that it is possible to generate high accuracy of recording from newly trained observers with no previous experience in systematic observation using a range of observation methods with naturalistic observation materials. Second, paper-and-pencil observation induced a marginally superior level of accuracy relative to computer-aided observation systems. Third, there was a significant mediating effect of session busyness on observation accuracy for computer-aided observation systems but not for paper-and-pencil observation. Fourth, two computer-aided observation systems with very diverse user interphase features induced near-identical observer accuracy. We believe that these findings help validate the continued use of both paper-and-pencil and computer-aided systems in behavior observation applications. They also highlight the need for additional research on key mediating factors including observer experience, observation difficulty, and the response effort involved in operating various systems.

ACKNOWLEDGEMENTS

The authors thank Alex Tefay from SDR Scientific for facilitating access to The Observer XT®. ABA España provided free activation codes for the Big Eye Observer® app. Dr. Katrina Phillips and Dr. Sarah Leadley assisted with the recruitment process. The first author received support from a research contract with The University of Auckland (project no. CON02739). Supporting organizations were not involved in the study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the article for publication.

OPEN PRACTICE STATEMENT

The dataset generated during the current study is available from Virues-Ortega, J., Delgado Casas, C., Martin, N., Tarifa-Rodriguez, A., Reina Hidalgo, A. J., Cox, A. D., & Navarro Guzmán, J. I. (2022). *Paper-and-pencil systematic observation versus computer-aided systems: A comprehensive accuracy dataset* [Data set]. Author. <https://doi.org/10.6084/m9.figshare.19555096>

REFERENCES

- ABA España (2019). *Big Eye Observer* (Version 2.0). [Mobile App]. <https://apps.apple.com/us/app/big-eye-observer/id1299214188>
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723. <https://doi.org/10.1109/TAC.1974.1100705>
- Allen, J. A., Lehmann-Willenbrock, N., & Rogelberg, S. G. (2015). Effective leader behaviors in regularly held staff meetings: Surveyed vs. videotaped and video-coded observations. In M. Hoogeboom & C. Wilderom (Eds.), *The Cambridge Handbook of Meeting Science* (pp. 381-412). New York: Cambridge University Press.
- Bailey, D. W., Keil, M .R., Rittenhouse, L. R. (2004). Research observation: Daily movement patterns of hill climbing and bottom dwelling cows. *Journal of Range Management*, 57, 20-28. [https://doi.org/10.2111/1551-5028\(2004\)057\[0020:RODMPO\]2.0.CO;2](https://doi.org/10.2111/1551-5028(2004)057[0020:RODMPO]2.0.CO;2)
- Bass, R. F. (1987). Computer-assisted observer training. *Journal of Applied Behavior Analysis*, 20, 83–88. <https://doi.org/10.1901/jaba.1987.20-83>

- Blumstein, D., Evans, C., & Daniel, J. (2006). *JWatcher* (Version 1.0). [Computer Software] <https://www.jwatcher.ucla.edu/>
- Burkardt, J. (2014). *The Truncated Normal Distribution*. Florida State University. Available from https://people.sc.fsu.edu/~jburkardt/presentations/truncated_normal.pdf
- Burnham, K. P., & Anderson, D. R. (2002). *Model Selection and Multimodel Inference* (2nd ed.). Springer.
- Cooper, J. O., Heron, T. E., & Heward, W. L. (2019). Improving and Assessing the Quality of Behavioral Measurement. In J. O. Cooper, J. O., T. E. Heron, & W. L. Heward (Eds.), *Applied Behavior Analysis* (3rd ed., pp. 101-122). Pearson.
- Cox, A. D., & Virues-Ortega, J. (2021). Long-term functional stability of problem behavior exposed to psychotropic medications. *Journal of Applied Behavior Analysis*, 55(1), 214-229, <https://doi.org/10.1002/jaba.873>
- Crowley-Koch, B. J., & Van Houten, R. (2013). Automated measurement in applied behavior analysis: A review. *Behavioral Interventions*, 28, 225-240. <https://doi.org/10.1002/bin.1366>
- Delgado, C., Garcia, R, Navarro, J. I., & Hinojo, E. (2012). Functional analysis of challenging behaviors in people with severe intellectual disabilities using The Observer XT 10.0 software. *Proceedings of Measuring Behavior 2012* (Utrecht, The Netherlands, August 28-31, 2012), 365-367.
- Dempsey, C. M., Iwata, B. A., Fritz, J. N., & Rolider, N. U. (2012). Observer training revisited: A comparison of in vivo and video instruction. *Journal of Applied Behavior Analysis*, 45, 827-832. <https://doi.org/10.1901/jaba.2012.45-827>.
- Des Jarlais, D. C., Lyles, C., Crepaz, N., & the Trend Group (2004). Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: The TREND statement. *American Journal of Public Health*, 94, 361-366. <https://doi.org/10.2105/ajph.94.3.361>
- Farrimond, J. A., Hill, A. J., Jones, N. A., Stephens, G. J., Whalley, B. J., & Williams, C. M. (2009). A cost-effective high-throughput digital system for observation and acquisition of animal

behavioral data. *Behavior Research Methods*, 41(2), 446–451. <https://doi.org/10.3758/BRM.41.2.446>

Franchi, V., Aleuy, O. A., & Tadich, T. A. (2016). Fur chewing and other abnormal repetitive behaviors in chinchillas (*Chinchilla lanigera*), under commercial fur-farming conditions. *Journal of Veterinary Behavior: Clinical Applications and Research*, 11, 60-64. <https://doi.org/10.1016/j.jveb.2015.10.002>

Gargani, J., & Strong, M. (2014). Can we identify a successful teacher better, faster and cheaper? Evidence for innovating teacher observation systems. *Journal of Teacher Education*. <https://doi.org/10.1177/0022487114542519>.

Gavran, D. P., & Hernandez, V. (2020). Countee (Version 2.2.1). [Mobile App]. <https://apps.apple.com/nz/app/countee/id982547332>

Gelman, A., & Hill, J. (2008). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.

Hall, S. S., Hustyi, K. M., Chui, C., & Hammond, J. L. (2014). Experimental functional analysis of severe skin-picking behavior in Prader-Willi syndrome. *Research in Developmental Disabilities* 35, 2284-2292. <https://doi.org/10.1016/j.ridd.2014.05.025>

Hanley, G. P., Iwata, B. A., & McCord, B. E. (2013). Functional analysis of problem behavior: A review. *Journal of Applied Behavior Analysis*, 36, 147-185. <https://doi.org/10.1901/jaba.2003.36-147>

Haynes, S. N., & O'Brien, W. H. (2000). *Principles of Behavioral Assessment*. Springer.

Hayward, J., Buckingham, S., Thomson, F., Milne, H., Sheikh, A., Fernando, B., Cresswell, K., Williams, R., & Pinnock, H. (2015). How long does it take? A mixed methods evaluation of computer-related work in GP consultations. *Journal of Innovation in Health Informatics*, 22, 409-425. <https://doi.org/10.14236/jhi.v22i4.95>

Hintze, J. M. (2004). Behavior observation in schools. In C. D. Spielberger (Ed.), *Encyclopedia of Applied Psychology* (pp. 283-287). <https://doi.org/10.1016/B0-12-657410-3/00761-3Get>

- Hinz, K. L., McGee, H. M., Huitema, B. E., Dickinson, A. M., & Van Enk, R. A. (2014). Observer accuracy and behavior analysis: Data collection procedures on hand hygiene compliance in a neurovascular unit. *American Journal of Infection Control*, 42(10), 1067–1073. doi:10.1016/j.ajic.2014.06.017
- Hudson, J.L., & Rapee, R.M. (2001). Parent-child interactions and anxiety disorders: An observational study. *Behaviour Research and Therapy*, 39, 1411-1427. [https://doi.org/10.1016/s0005-7967\(00\)00107-8](https://doi.org/10.1016/s0005-7967(00)00107-8)
- Hutman, T., Chela, M.K., Gillespie-Lynch, K., & Sigman, M. (2012). Selective visual attention at twelve months: Signs of autism in early social interactions. *Journal of Autism and Developmental Disorders*, 42, 487-498. <https://doi.org/10.1007/s10803-011-1262-5>
- IBM Corporation. (2021). SPSS® IBM® Statistics (Version 27). [Computer Software]. <https://www.ibm.com/products/spss-statistics>
- Iwata, B., Dorsey, M., Slifer, K., Bauman, K., & Richman, G. (1994). Toward a functional analysis of self-injury. *Journal of Applied Behavior Analysis*, 27, 197-209. <https://doi.org/10.1901/jaba.1994.27-197> (Original document published 1982).
- Jackson, J., & Dixon, M. R. (2007). A mobile computing solution for collecting functional analysis data on a Pocket PC. *Journal of Applied Behavior Analysis*, 40, 359-384. <https://doi.org/10.1901/jaba.2007.46-06>
- Kahng, S., & Iwata, B. A. (1998). Computerized systems for collecting real-time observational data. *Journal of Applied Behavior Analysis*, 31, 253-261. <https://doi.org/10.1901/jaba.1998.31-253>
- Kazdin, A. E. (2011). *Single-case research designs: Methods for clinical and applied settings* (2nd ed.). New York, NY: Oxford University Press.
- Lerman, D. C., Tetreault, A., Hovanetz, A., Bellaci, E., Miller, J., Karp, H., Mahmood, A., Strobel, M., Mullen, S., Keyl, A., & Toupard, A. (2010). Applying signal-detection theory to the study of observer accuracy and bias in behavioral assessment. *Journal of Applied Behavior Analysis*, 43(2), 195–213. <https://doi.org/10.1901/jaba.2010.43-195>

- Mash, E. J., & McElwee, J. D. (1974). Situational effects on observer accuracy: Behavioral predictability, prior experience, and complexity of coding categories. *Child Development*, 45(2), 367. <https://doi.org/10.2307/1127957>
- Martin, N., Oliver, C., & Hall, S. (1999). *ObsWin: Observational Data Collection & Analysis for Windows*. *CTI Psychology Software News*, 9, 14-16.
- Mckerchar, P. M., & Abby, L. (2012). Systematic Evaluation Of Variables That Contribute To Non-compliance: A Replication And Extension. *Journal of Applied Behavior Analysis*, 45, 607-611. <https://doi.org/10.1901/jaba.2012.45-607>
- Meirsschaut, M., Roeyers, H., & Warreyn, P. (2011). The social interactive behaviour of young children with autism spectrum disorder and their mothers. *SAGE Publications and The National Autistic Society*, 15, 43-64. <https://doi.org/10.1177/1362361309353911>
- Mossman, A. (2011). A strength-based approach to parent education for children with autism. *Journal of Positive Behavior Interventions*, 13, 178-190.
- Moyle, W., Jones, C., Cooke, M., O'Dwyer, S., Sung, B., & Drummond, S. (2014). Connecting the person with dementia and family: a feasibility study of a telepresence robot. *BMC Geriatrics*, 14, 1-11. <https://doi.org/10.1186/1471-2318-14-7>
- Naber, F. B. A., Bakermans-Kranenburg, M. J., Van IJzendoorn, M. H., Swinkels, S. H. N., Buitelaar, J.K., Dietz, C., Van Daalen, E., & Van Engeland, H. (2008). Play behavior and attachment in toddlers with autism. *Journal of Autism and Developmental Disorders*, 38, 857-866. <https://doi.org/10.1007/s10803-007-0454-5>
- Nave, C. S., Feeney, M. G., & Furr, R. M. (2018). Behavior observation in the study of personality and individual differences. In V. Zeigler-Hill & T. K. Shackelford (Eds.), *The SAGE handbook of personality and individual differences: The science of personality and individual differences* (pp. 317–340). Sage Reference. <https://doi.org/10.4135/9781526451163.n15>
- Noldus Information Technology. (2021). *EthoVision XT (Version 16)*. [Computer Software]. <https://www.noldus.com/ethovision-xt>
- Noldus Information Technology. (2019). *The Observer XT (Version 14.2)*. [Computer Software]. <https://www.noldus.com/observer-xt>

- Mudford, O. C., Taylor, S. A., & Martin, N. T. (2009). Continuous recording and interobserver agreement algorithms reported in the *Journal of Applied Behavior Analysis* (1995–2005). *Journal of Applied Behavior Analysis*, 42, 165–169. <https://doi.org/10.1901/jaba.2009.42-165>
- Oh, J., & Fitch, W. T. (2017). CATOS (Computer Aided Training/Observing System): Automating animal observation and training. *Behavior Research Methods*, 49(1), 13–23. <https://doi.org/10.3758/s13428-015-0694-9>
- Page, T. J., & Iwata, B. A. (1986). Interobserver agreement: History, theory and current methods. In A. Poling & R. W. Fuqua (Eds.), *Research Methods in Applied Behavior Analysis: Issues and Advances* (pp. 99–126). Plenum.
- Pardo-Cebrian, R., Virues-Ortega, J., Calero-Elvira, A., & Guerrero-Escagedo, M. C. (2021). Toward an experimental analysis of verbal shaping in psychotherapy. *Psychotherapy Research*. <https://doi.org/10.1080/10503307.2021.1955418>
- Péter, A. (2019). *Solomon Coder* (Version 19.08.02). [Computer Software]. <https://solomon.andraspeter.com>
- Piazza, C. C. (2008). Feeding disorders and behavior: What have we learned? *Developmental Disabilities Research Reviews*, 14, 174–181. <https://doi.org/10.1002/ddrr.22>
- Podofilini, L., Park, K., & Dang, V. N. (2013). Measuring the influence of task complexity on human error probability: An empirical evaluation. *Nuclear Engineering and Technology*, 45(2), 151–164. <https://doi.org/10.5516/net.04.2013.702>
- Ray, R. D., Ray, J. M., Eckerman, D. A., Milkosky, L. M., & Gillins, L. J. (2011). Operations analysis of behavioral observation procedures: a taxonomy for modeling in an expert training system. *Behavior Research Methods*, 43(3), 616–634. <https://doi.org/10.3758/s13428-011-0140-6>
- Roth, T., Kramer, M., & Lutz, T. (1976). The nature of insomnia: A descriptive study of a sleep clinic population. *Comprehensive Psychiatry*, 17, 217–220. [https://doi.org/10.1016/0010-440X\(76\)90072-9](https://doi.org/10.1016/0010-440X(76)90072-9)

- Sarkar, A., Dutta, A., Dhingra, U., Dhingra, P., Verma, P., Juyal, R., Black, R. E., Menon, V. P., Kumar, J., & Sazawal, S. (2006). Development and use of behavior and social interaction software installed on Palm handheld for observation of a child's social interactions with the environment. *Behavior Research Methods*, 38, 407-415. <https://doi.org/10.3758/BF03192794>
- Tarbox, J., Wilke, A. E., Findel-Pyles, R. S., Bergstrom, R. M., & Granpeesheh, D. (2010). A comparison of electronic to traditional pen-and-paper data collection in discrete trial training for children with autism. *Research in Autism Spectrum Disorders*, 4, 65-75. <https://doi.org/10.1016/j.rasd.2009.07.008>
- VideoLAN Association. (2019). *VLC Media Player (Version 3.0.6)*. [Computer Software]. <https://www.videolan.org/vlc/releases/3.0.6.html>
- Virues-Ortega, J., Montaña-Fidalgo, M., Froján-Parga, M. X., & Calero-Elvira, A. (2011). Descriptive Analysis of the Verbal Behavior of a Therapist: A Known-Group Validity Analysis of the Putative Behavioral Functions Involved in Clinical Interaction. *Behavior Therapy*, 42, 547-559. <https://doi.org/10.1016/j.beth.2010.12.004>
- Yasukawa, K., & Bonnie, K. E. (2017). Observational and experimental methods in comparative psychology. In J. Call, G. M. Burghardt, I. M. Pepperberg, C. T. Snowdon, & T. Zentall (Eds.), *APA Handbook of Comparative Psychology: Basic Concepts, Methods, Neural Substrate, and Behavior* (pp. 65-86). American Psychological Association. <https://doi.org/10.1037/0000011-004>
- Virues-Ortega, J., Clayton, K., Pérez-Bustamante, A., Gaerlan, B., & Fahmie, T. A. (2022). Functional analysis patterns of automatic reinforcement: A review and component analysis of treatment effects. *Journal of Applied Behavior Analysis*, 55(2), 481-512. <https://doi.org/10.1002/jaba.900>
- Virues-Ortega, J., Delgado Casas, C., Martín, N., Tarifa-Rodríguez, A., Reina Hidalgo, A. J., Cox, A. D., & Navarro Guzmán, J. I. (2022). *Paper-and-pencil systematic observation versus computer-aided systems: A comprehensive accuracy dataset [Data set]*. Author. <https://doi.org/10.6084/m9.figshare.19555096>

Wildman, B. G., Erickson, M. T., & Kent, R. N. (1975). The effect of two training procedures on observer agreement and variability of behavior ratings. *Child Development*, 46(2), 520–524. <https://doi.org/10.2307/1128151>

Wessel, D. (2015). The Potential of Computer-Assisted Direct Observation Apps. *International Journal of Interactive Mobile Technologies*, 9, 31-40. <https://doi.org/10.3991/ijim.v9i1.4205>

Zaiontz, C. (2021). *Real Statistics Resource Pack*. [Microsoft Excel Add-In Function]. <https://www.real-statistics.com/free-download/real-statistics-resource-pack/>

Table 7.1 Linear Mixed Effects Model for Accuracy Proportion (n = 10).

Fixed effects	F	df	p
Method (A)	16.022	2, 543	< .001
Recordable events (B)	5.308	45, 526	< .001
Video mode (C)	3.016	1, 263	.084
Interaction, A X B	1.844	88, 527	< .001
Interaction, A X C	1.552	2, 537	.213
Fixed effects (nested factors)			
Method (Participant)	11.179	17, 530	< .001
Video mode (Participant)	3.068	9, 256	.002
Recordable events (Video number)	6.467	10, 535	< .001
Pairwise comparisons	Coefficient	df	p
Video mode			
Video playback vs. <i>in vivo</i> proxy	.011± .006	1, 265	.065
Method			
P&P vs. OXT	.035± .007	2, 541	< .001
P&P vs. BEO	.039± .007	2, 539	< .001
OXT vs. BEO	.003± .008	2, 542	1.000

Notes. First-order autoregressive covariance structure ($\rho = -0.15 \pm 0.07$). The coefficient of pairwise comparisons is the mean difference (first term minus second term) expressed in the metric of the dependent variable. BEO = Big Eye Observer, OXT = The Observer XT, P&P = paper-and-pencil.

Figure 7.1 Observation time across phases and observation methods.

Note. All means and standard errors.

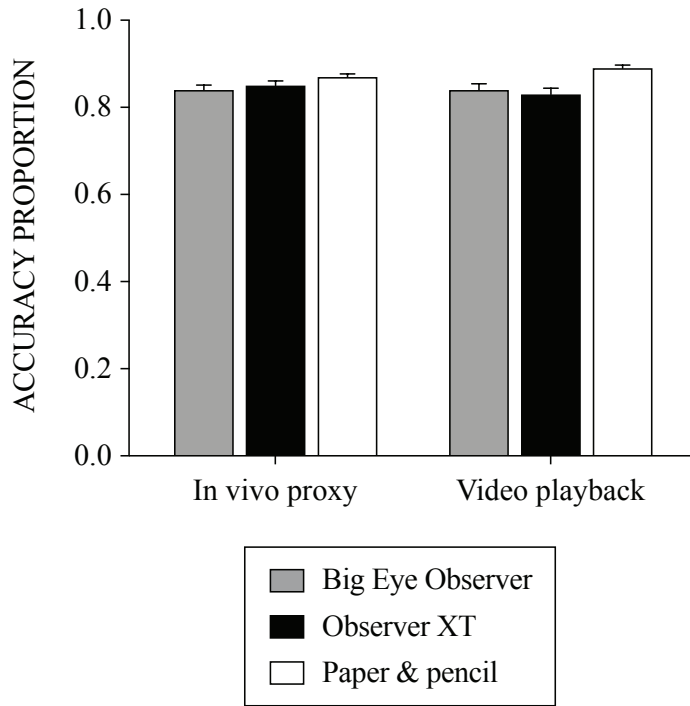


Figure 7.2 Accuracy across phases and observation methods.

Note. All means and standard errors.

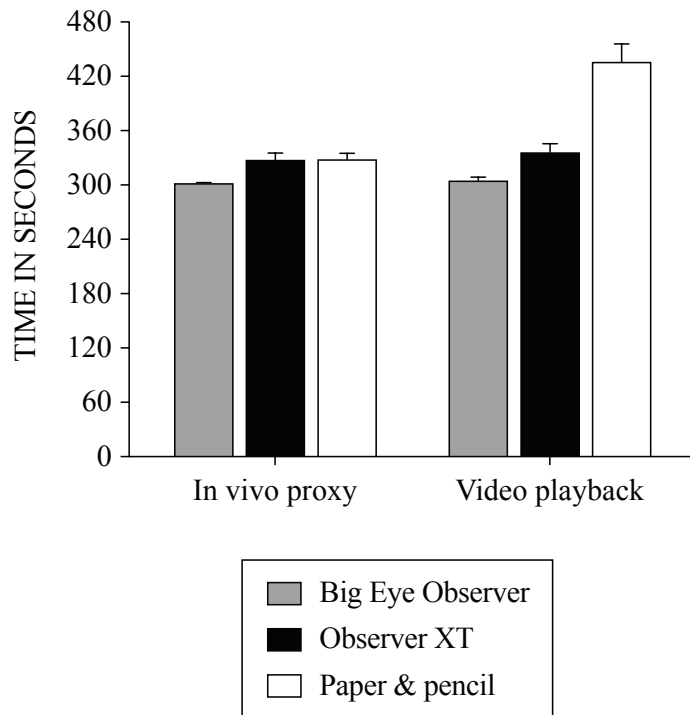
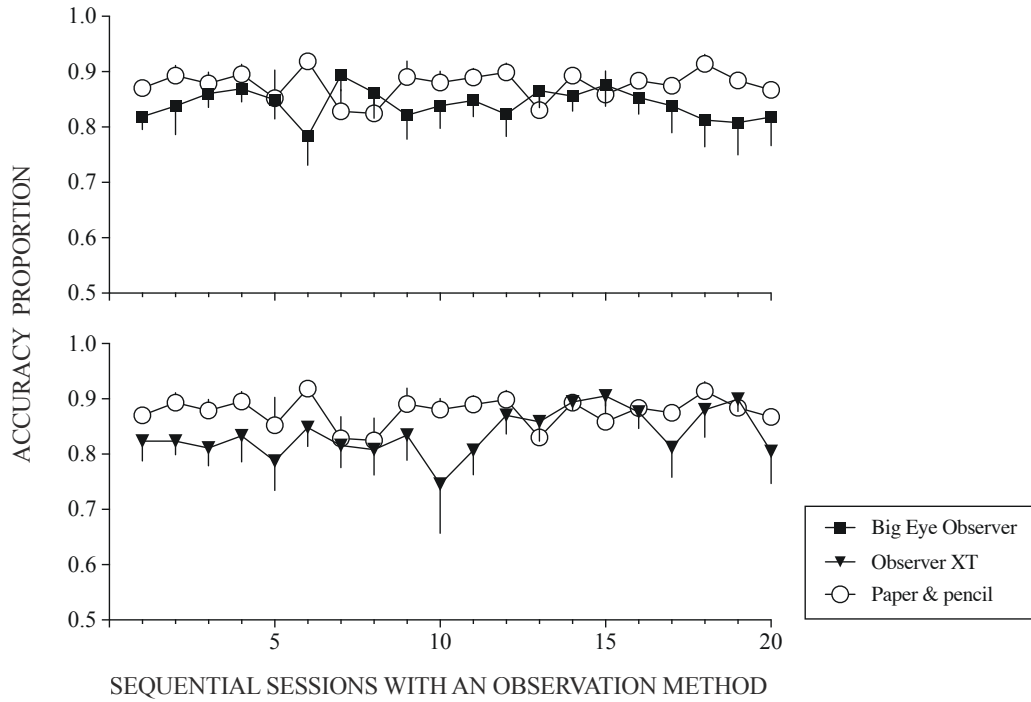


Figure 7.3 Observation Accuracy of The Observer XT® vs. Paper and Pencil (top panel) and of Big Eye Observer® vs. Paper and Pencil (Lower Panel) across Participants.



Notes. All means and standard errors. The successive order of the multi-element design has been suppressed to allow for data aggregation across participants.

Figure 7.4 Accuracy by Number of Recordable Events across All Observation Sessions.

Note. R_s = Spearman rank correlation coefficient.

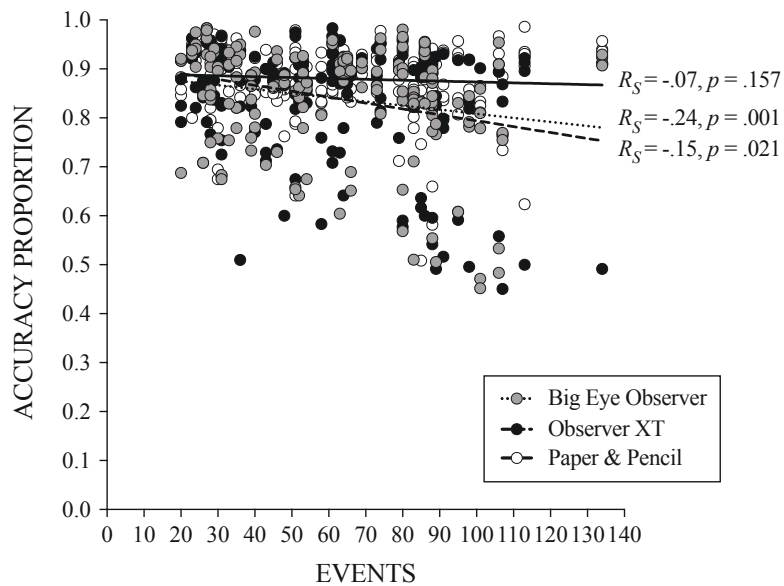
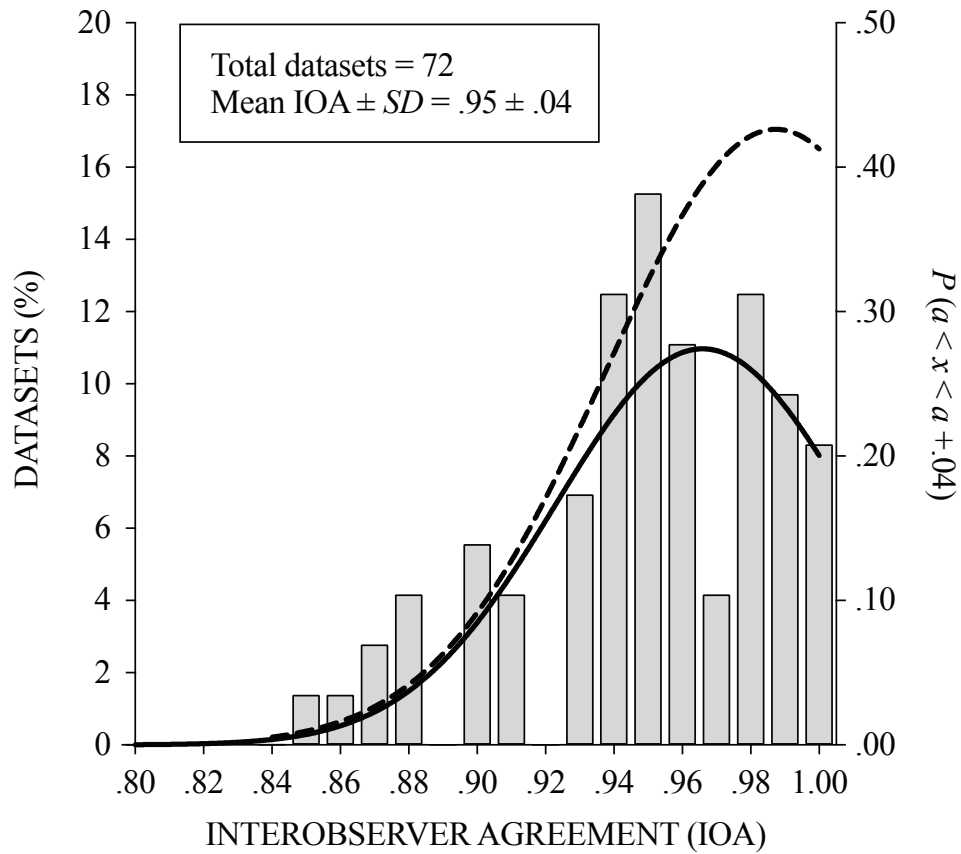


Figure 7.5 Distribution of Interobserver Agreement Values in the Literature.



Notes. The bars and solid line (left y axis) denote the empirical and estimated distribution of IOA values in the literature as reviewed by Virues-Ortega et al. (2022). The broken line (right y axis) is the cumulative probability function between any IOA value a and an IOA value $a + .04$ (according to the probability density function for a truncated normal distribution, calculated according to Burkardt, 2014 and Zaiontz, 2021).

CHAPTER 8. Quantitative Indices of Student Social Media Engagement in Tertiary Education: A Systematic Review and a Taxonomy⁵²

ABSTRACT

Recent studies have evaluated the use of social media as learning aids in tertiary education. Emerging research in this area has focused primarily on non-quantitative approaches to student social media engagement. However, quantitative engagement outcomes may be extracted from student posts, comments, likes, and views. The goal of the present review was to provide a research-informed taxonomy of quantitative and behaviorally-based metrics of student social media engagement. We selected 75 empirical studies comprising a pooled sample of 11,605 tertiary education students. Included studies used social media for educational purposes and reported student social media engagement outcomes (source databases: PsycInfo and ERIC, search date: October 28, 2020). We used independent raters and an interrater agreement process for the screening of references to mitigate bias. Over half of the studies (52%, $n = 39$) utilized ad hoc interviews and surveys to estimate student social media engagement, whereas thirty-three studies (44%) used some form of quantitative analysis of engagement. The proposed taxonomy of engagement metrics resulting from this review provides the methodological basis for the analysis of social media behavior in educational settings, particularly, for human operant and behavioral education studies. Implications for future research are discussed.

Key-words: social media, social media engagement, achievement, tertiary education, behavioral engagement, online education.

52 El presente estudio ha sido enviado a la revista *Journal of Behavior Education* y se presenta en el idioma y el formato que se ha enviado.

Tarifa-Rodríguez, A., Virues-Ortega, J., Calero-Elvira, A. (2022). Quantitative indices of student social media engagement in tertiary education: a systematic review and a taxonomy. *Journal of Behavior Education*

INTRODUCTION

According to some estimates, social media platforms will reach four billion users in 2022 (Statista, 2021), with platforms such as Facebook, Instagram, WhatsApp, YouTube, TikTok, and Twitter serving as the main communication networks. This trend is particularly prominent among university students. For example, a recent survey indicated that 82% of tertiary education students are regular Facebook users (Statista, 2022). Yet, universities have only started to use these platforms for educational purposes. Social media have the potential of supporting traditional classroom environments by adding accessible, barrier-free virtual spaces that could enhance collaborative peer- and instructor-mediated learning. In fact, there is evidence to suggest that student engagement in social media discussions moderated by an instructor may be an important indicator of course content elaboration and social learning (Parks-Stamm et al., 2017).

Studies within this emerging field often rely on indirect measures of engagement (e.g., satisfaction surveys, teacher reports), which provide limited information on the quantitative dimensions of online behavior; dimensions such as frequency, latency, and intensity (Giannikas, 2020; Slim et al., 2019). Social media platforms make it possible to quantify student engagement in a variety of ways. For example, most social media platforms log the exact time and date of posts, comments, and user reactions (e.g., likes) affording a myriad of metrics (e.g., posting frequency, comment latency). It is possible that these metrics have the potential to inform the teaching-learning process when social medial channels are used in educational contexts.

Quantitative measures of engagement allow us, not only to quantify the effectiveness of interventions directed at increasing engagement, but also to detect operant learning mechanisms such as reinforcement, extinction, and punishment, that could be influencing students' online behavior (Honig & Staddon, 2022). Recently, Lindström et al. (2021) used a computational approach to assess whether operant processes could explain engagement responses in social media. These authors used response latency (the time elapsed between two successive social media posts) as an indicator of engagement. Their results showed that users of social media platforms space out their posts according to a model of social reinforcement maximization. This finding may have implications towards the use of social media for educational purposes.

An operant model of social media interaction could provide the conceptual bases for future evidence-based strategies to foster positive and learning-enhancing interventions, for example, by using online social rewards such as offering immediate or near-immediate feedback. Moreover,

it may be possible to create educational contexts in which high rates of social reinforcement are available for appropriate engagement, which could ultimately maximize academic performance. We could also obtain evidence of operant behavior allocation by monitoring posting behavior at times when instructor responses have a shorter latency or are more relevant (e.g., specific feedback), relative to times when instructor responses are delayed or are less relevant (e.g., collective feedback). Namely, more frequent posting in the former scenario and less frequent posting in the latter would provide evidence (whether correlational or experimental) of operant behavior allocation. The analyses suggested above may have direct practical implications. In order to evaluate operant processes in the social media context, it would be necessary to establish quantitative metrics of discrete student and instructor social media responses.

While the literature on the use of social media for educational purposes has grown steadily over the last decade (Tawafak et al., 2021), most of this research is primarily qualitative and does not contain empirical data on engagement or performance (Papademetriou et al., 2022), making it difficult to capitalize on this important line of research. In addition, in order to evaluate an operant model of social media interaction in educational settings, it would be important to define and validate behaviorally-based quantitative metrics of social media interaction.

The goal of the current study is to review the literature that has evaluated social media engagement in the context of tertiary education programs with integrated social media platforms. This evidence will be used as the basis for a preliminary taxonomy of quantitative engagement metrics that could be widely used in human operant and applied behavioral education.

METHODS

Study Selection

We conducted a comprehensive literature search in the PsycInfo and ERIC databases (ProQuest search engine) on October 28, 2020. After repeated preliminary searches to test search sensitivity, the following search strategy was implemented: (“Facebook” OR “social media”) AND (“engagement,” “education,” OR “achievement”) without time or search field restrictions.

We included studies meeting the following inclusion criteria: (a) the study included college-level, undergraduate, graduate, or postgraduate students (Criterion 1), (b) the study used a social media platform for educational purposes (Criterion 2), and (c) the study included at least

one social media engagement variable (Criterion 3). We screened the abstracts of the studies identified through the initial search to assess Criteria 1 and 2. We retrieved and processed the full manuscripts of studies meeting Criteria 1 and 2 for the purposes of verification and for evaluating Criterion 3. The initial search returned 766 distinct references. We implemented inclusion criteria sequentially. Figure 1 presents a detailed record of the implementation of the inclusion criteria (see also Supplementary Online Material, Table A). Seventy-five studies met all inclusion criteria and proceeded to the data extraction phase.

Two raters participated in the study. Rater #1 applied inclusion criteria to all references originally retrieved (ATR). For the purposes of evaluating interrater agreement during the screening of references, a second rater (APG) independently applied inclusion criteria to the first 100 references. The primary and secondary raters applied the inclusion criteria to all 100 references identically, resulting in an interrater agreement of 100%. The use of multiple databases, independent raters, and an interrater agreement process was intended to minimize the risk of bias during the selection of references. The current systematic review adheres to the PRISMA statement for reporting systematic reviews (see Supplementary Online Information). The database upon which this systematic review has been made available via Figshare (undisclosed authors, 2022).

Data Extraction

The following variables were extracted from all studies meeting the inclusion criteria.

Number, age, gender, and country of participants. We recorded the total number of participants as well as their age, gender, and educational level (high school, college, master's or PhD). We also recorded the country in which the study was conducted.

Field of study. We recorded the field of study of all educational interventions. These were then classified according to an *ad hoc* category system. For example, research conducted with astronomy or physics students was categorized as "physical and life sciences", psychology and sociology students were categorized as "social sciences", second language students were included within "language and communication," and students in management of information and digital content were grouped under "computer and technology". Additional categories were used for "art," "business," and "professional courses."

Study design. Studies simply using a questionnaire or interview at the end of the course were classified as “qualitative (retrospective).” Studies that implemented their assessments before and/or during the intervention were categorized as “qualitative (prospective)”. Studies presenting correlation analyses for two or more variables at a given time point were classified as “observational.” Studies using experimental designs were divided into two categories: “intervention (within subject),” for studies where all participants were exposed at least to a control and treatment or posttreatment condition; and “intervention (between groups)” when students were assigned to control and interventions groups. Randomization was not considered part of the classification process as none of the selected studies included random assignment.

Predictive and outcome variables. We also recorded predictive and outcome variables reported in the studies reviewed, including study time, academic achievement, user satisfaction, and personality constructs. Predictive and outcome variables were further classified as quantitative (objective and standardized tests) or qualitative (interviews, surveys, ad hoc questionnaires). Additional personal outcomes (e.g., motivation, sense of community, positive feelings) were included in a miscellaneous category.

Social media platforms and engagement. Social networks are Web 2.0 online spaces that allow people to create profiles to communicate and exchange content with other users. Facebook, Telegram, WhatsApp, Twitter, Instagram, and TikTok are prominent examples. We recorded the social media platform utilized in each study included in the review. Engagement could be evaluated with ad hoc interviews or surveys (“interview or survey”), standardized tests (“standardized test”), or quantitative engagement metrics (“behavioral”). The latter included discrete outcomes such as posting, commenting, and reaction frequencies, among others.

RESULTS

General Characteristics of Studies

Table 1 summarizes the results of the data extraction process. A pooled sample of 11,605 students participated in a total of 75 selected studies (age range, 17-60, 69.1% female). Social sciences (e.g., psychology and sociology) was the most common field of study among the studies reviewed (30.7%, $n = 23$), followed by language and communication (e.g., second language learning courses) (29.3%, $n = 22$), computer and technology (16%, $n = 12$), physical and life sciences (e.g., astronomy) (9.3%, $n = 7$), and vocational programs (e.g., digital content management) (5.3%, n

= 4). The studies were geographically varied. There were missing values in some of the variables targeted for data extraction, including age of the students ($n = 52$), gender of the students ($n = 42$), course level ($n = 1$), and field of study ($n = 3$).

Overall, 53.3% of studies ($n = 40$) evaluated an intervention mediated by a social media platform. Of these, 26.7% ($n = 20$) used a between-group design, and 26.7% ($n = 20$) used a pre-post within-subject design with no control group. None of the between-group studies was a randomized controlled trial. Observational or correlational studies that did not evaluate an intervention, but conducted regression analyses of engagement and personal outcomes cross-sectionally were the second most frequent study design (24%, $n = 18$). The remaining studies were prospective qualitative studies that did not include a formal intervention or control group (21.3%, $n = 16$).

Predictive and Outcome Variables

Some studies introduced social media platforms as an intervention intended to enhance academic achievement. Among intervention studies, 22.7% ($n = 17$) used pre-post academic achievement evaluated through *ad hoc* surveys, assignment marks, objective tests, and course grades. Sixteen of these intervention studies prospectively manipulated the introduction of a social media platform as a deliberate intervention. Finally, 11 studies (14.7%) evaluated both engagement and achievement with qualitative methods.

Some qualitative studies used *ad hoc* surveys to evaluate additional outcomes of the students' educational experience while using the social media platform. Specifically, 28 studies evaluated student satisfaction using the social media platform (37.3%, $n = 19$), and five studies assessed the sense of community belonging (10.2%, $n = 5$). Additional outcomes (e.g., intrinsic motivation) were not observed more than twice in the pool of studies included in the review. The vast majority of included studies that used qualitative methods did not report any additional outcomes (see Table 1 for details).

Social Media Platforms and Engagement

Approximately half of the studies selected (52%, $n = 39$) utilized *ad hoc* interviews and surveys to estimate the degree of social media participation among students, whereas thirty-three studies (44%) used some form of quantitative analysis of engagement based on the metrics provided by the social media platform (e.g., frequency of likes, comments, and posts). Table 2 summarizes the

objective measures of social media engagement present in this literature. Specifically, Clements (2015), Gordon (2016), and Pai et al. (2017a, 2017b) computed the total number (frequency) of likes, comments, and posts from the group of students using the social network as a teaching and communication channel. In addition, eight studies used the frequency of posts and comments only (Hou et al., 2015; Lim, 2010; Luo, 2018; Miller, 2013; Naghdipour et al., 2016; Nazir & Brouwer, 2019; Peeters et al., 2020; Wu et al., 2018). Miller (2013) departed from this trend by extracting the frequency of posts and comments per student as well as exploring posting immediacy. Luo (2018) and Wu et al. (2018) extracted the total number of characters composing each social media comment, while Nazir and Brouwer (2019) conducted a systematic theme analysis using the text of Facebook comments as samples. Bacile (2013), Dougherty et al. (2014) and Tran (2016) collected the frequency of posts and likes, Shih (2011) the frequency of likes and comments, and Montoneri (2015) the frequency of views and likes per post. Finally, eight additional studies reported the frequency of posts as their single quantitative engagement outcome (Daniels et al., 2014; Ercoskun et al., 2019; Harting, 2017; Orawiwatnakul et al., 2016; Owens et al 2017; Riady, 2014; Schroeder et al., 2009; Whittaker et al., 2014; Yu, 2014)

All above mentioned studies utilized Facebook as their integrated social media platform, whereas four additional studies used Twitter. Specifically, Tur et al. (2015), Naravan et al. (2019) and Popescu et al. (2020) extracted the frequency of tweets, retweets, and replies to tweets, while Martinez-Cardama et al. (2019) computed the total number of hashtags.

Finally, Alghazo et al. (2018) used the texting social media app WhatsApp and monitored class attendance and missed assignments (social media engagement outcomes were not reported). In addition, Goktalay (2015) and Moghavvemi et al. (2018) used standardized tests as an indirect assessment of social media engagement.

DISCUSSION

We reviewed the engagement metrics reported in studies using social media platforms as a learning channel in tertiary education. Our descriptive analysis showed that most studies utilized *ad hoc* surveys to document social media engagement and satisfaction (e.g., Gregory, 2014), while a minority of studies focused on objective engagement indicators including posts, comments, and reactions (e.g., Peeters et al., 2020). Most interventional studies evaluated the addition of a social media platform on student satisfaction (Akcaoglu et al., 2018), while few focused on objective or

standardized academic achievement outcomes. For example, Dougherty et al. (2014) correlated engagement, as estimated by the frequency of Facebook posts and reactions, with objective academic performance evaluated through weekly multiple-choice tests. Interventional studies followed pre-post within-subject and between-groups designs. Control groups were either withdrawn from the possibility of interacting through a target social media platform (Alghazo et al., 2018) or were exposed to a passive course instructor (Peeters et al., 2020). Controlled studies lacked randomization (no RCTs were identified). Specifically, participants in experimental groups were often offered a choice to participate in a social media platform as part of their course, while those in the control group underwent the usual course format without a social media channel to enhance peer-to-peer or instructor-to-student interaction (e.g., Gregory, 2014).

Overall, this literature offers a limited picture of the quantitative engagement responses that could be recorded from social media platforms. Objectively-defined engagement responses involved primarily posting and reacting frequency, whereas more sophisticated time-based or event-related outcomes were rarely explored (e.g., comment latency, percentage of individuals posting). In a notable exception, Miller (2013) studied posting frequency and posting immediacy. Posting immediacy may be defined as the time elapsed from a target instructor post to a post-related student response. Textual analysis methodology, including theme detection and sentiment analysis, which have become common social media research methods (Angus, 2017; Thelwall, 2017), were rare occurrences within this literature. Nazir and Brouwer (2019) illustrate an exception to this trend by qualitatively analyzing 67 post transcripts and comments made by both students and instructors in a Facebook group over the course of eight weeks. The aim of the textual analysis was to classify posts and comments according to three categories, “social presence,” “cognitive presence,” and “teaching presence.”

In an attempt to add to the systematization of this literature, we provide a summary of behaviorally-based engagement metrics in Table 3. This summary is provided as a sample of relevant and intuitive metrics and is not intended as an exhaustive collection of all possible outcomes. Some of the proposed metrics are yet to be utilized in empirical studies. Existing metrics can be divided into three categories, those based on the frequency or count of responses (count-based), those derived from the timing of the response (time-based), and those involving the automated analyses of the response length, content, and semantics (text analysis). Some of these metrics can be applied to an individual, group, or post as unit of analysis. For example,

commenting frequency (i.e., total number of comments over a period of time), could refer to a group (i.e., total number of comments made by a group of individuals over a period of time), an individual (i.e., total number of comments made by an individual over a period of time), or a post (i.e., total number of comments by any individual responding to a particular post over a period of time).

Posting frequency provides yet another example of a count-based metric, which may be defined as the total number of posts over a period of time (e.g., week, semester) and could be applicable to a group or an individual (Table 3). For example, Pai et. al. (2017a) monitored the instructor-led interactions in a Facebook group consisting of 150 biology students over a period of eight weeks. Authors extracted the total number of monthly posts by both students and instructors. The results of their descriptive analysis indicated that spontaneous student posting increased gradually over time (for an individual-level analysis of posting frequency see for example Tran, 2016). Posting frequency is significant in that, unlike commenting and reacting, reveals a spontaneous (unprompted) engagement with the discussion topic. Posting frequency can also be computed as a relative measure (i.e., total number of posts made by an individual or subgroup of individuals as a fraction of the total number of posts in the whole group or a different subgroup). Relative posting frequency may be used to obtain valuable information such as peer-to-peer or instructor-to-peer subgroup interaction or to provide data and compare the frequency of participation between subgroups within a wider social media group. These outcomes could also inform the correlation between engagement and academic performance or the effects on participation of an instructor-led intervention directed to a particular subgroup. An example of relative posting frequency can be found in Nazir and Brouwer (2019). The authors obtained the percentage of total posts and comments, including those specifically made by students and moderators in a Facebook group. However, relative measures are rarely used in the literature.

Time-based measures require additional attention. We highlight two common behavior dimensions yet to be explored as part of a quantitative analysis of engagement responses in education: latency and inter-response time (e.g., Rohrer & Wixted, 1994). Comment latency can be obtained by calculating time elapsed between a post or comment and its response. This parameter can be obtained for an individual (e.g., mean commenting latency of an individual), group (e.g., mean commenting latency of a group) or post (e.g., mean commenting latency of comments to a post). An example of the use of comment latency on social networks can be found

in Lindström et al. (2021), who used this metric to evaluate the effects of receiving feedback in the form of “likes” from other participants. Comment latency allowed Lindström et al. to demonstrate that the posting behavior of participants has operant characteristics and is causally influenced by social rewards. Even though the study did not have an educational component, Lindström et al. pioneering quantitative analysis demonstrates the potential of engagement responses to reveal operant processes. On the other hand, inter-response time can be defined as the time that elapses between two comments (comment inter-response time) or two reactions (reaction inter-response time) made by an individual. Inter-response time describes the pace at which a behavior is performed, which may be informative in online environments where receiving positive responses may be in part a function of response omission (e.g., an instructor may be more likely to respond favorably to posts from students that have not participated recently in a social media group).

Most of the proposed metrics are widely used in applied behavior analysis, specifically when evaluating teaching procedures for students with and without developmental disability. For example, latency have been used to monitor the time elapsed from the presentation of an instruction to the initiation of a relevant response (Koegel et al., 2010). Frequency, defined as the total number of times a behavior occurs (Kubina & Lin, 2008), is also widely used in educational settings (Bishop et al., 2020). Frequency provides information on how often target responses are met being a critical outcome for self-paced and fluency-based teaching-learning paradigms (Kubina & Morrison, 2000). It is likely that these metrics that have already demonstrated their effectiveness on educational contexts will also prove useful in the rapidly expanding field of educational applications of social media. In the current review we identified only six studies that quantitatively measured engagement using frequency (i.e., total number of posts and reactions) concurrently evaluated academic performance. For example, Shih (2011) and Montoneri (2015) obtained measures before and after completing questionnaires based on a 5-point Likert scale assessing content, organization, structure, and orthographic content. By contrast, Alghazo et. al. (2018), Dougherty et al. (2014), Miller (2013), and Tran (2016) used final exams grades or course assignment marks to estimate academic performance. The evidence available still portrays a much fragmentary picture on the potential relation between social media engagement and academic performance (meta-analyses remain impractical).

Text analysis provides an additional means for processing social media engagement inputs in educational contexts. Text length (i.e., post length or comment length) is among the most elementary text analysis metrics available. For example, Pai et al. (2017) calculated the mean length of posts and comments, and found that posts generally had a larger number of characters than comments. However, authors did not assess the association between textual length and academic performance. Secondly, theme and sentiment analysis may also have heuristic value in this context. Theme analysis allow for the automated processing of natural language in order to generate a hierarchy of topics, identify patterns of interest, and interpret communicative processes (Angus, 2017). There are multiple text analysis algorithms available (see, for example, Smith, 2000; Smith and Humphreys, 2006) that can be useful for processing large volumes of text extracted from social media and obtain thematic trends that would be impossible to obtain otherwise. In addition, textual analysis metrics can provide an indication of the social validity of interventions. For example, in a recent study by Anderson et al. (2021), thematic and textual analyses were conducted to assess the social validity of Likert-type behavioral-analytic scales. The more complex textual analysis metrics (i.e., sentiment and theme analysis) are yet to be implemented in the scientific literature on educational interventions using social media as a potential learning channel.

Lastly, sentiment analysis metrics can help to understand the role of emotional and motivational factors in communication processes (e.g., Thelwall, 2017). Various algorithms can estimate the intensity and quantity of sentiments expressed in a large volume of texts to ascertain sentiment patterns over time. A study by Ortigosa et al. (2014) demonstrated the possibility of obtaining useful data from Facebook posts to obtain information about learning experiences arising from interaction and posts generated in social networks. Data extraction was evaluated, although not the usefulness of the information obtained from the interactions. Sentiment analysis can also provide instructors with critical ongoing information about student engagement and performance (see, for example, Zhou & Jun-min, 2020).

Various limitations to the current analysis should be noted. First, we found that most of the literature used qualitative surveys and ad hoc questionnaires to measure engagement often cross-sectionally. Second, the studies that did assess engagement through, often relied on ad hoc Likert-type questionnaires. Only a few studies reported behaviorally-derived direct engagement responses (e.g., frequency of posting). In addition, over one third of the studies

reviewed were conducted in the context of second language learning courses (e.g., English as second language, ESL), thereby limiting the diversity of fields of study sampled in this review. Namely, the motivation and social dynamics of vocational (e.g., ESL) versus academic courses may be fundamentally different. Forth, while most studies focused on engagement (using qualitative and quantitative approaches), only a handful of studies concurrently monitored social media engagement and achievement and other relevant outcomes. Finally, none of the studies reviewed designed a control group or control condition that equated student exposure to instructor inputs. This limitation is not trivial, as equating instructor-led inputs and exposure to course content may be a critical methodological standard to ensure the relevance of the control group in future RCTs.

CONCLUSION

The analysis of discrete engagement responses opens the door to human operant and applied behavior-analytic studies within this field. In addition, pre- and post-test measures of students' academic performance should be incorporated to confirm whether the increase in engagement may be a good predictor of academic achievement. Randomized controlled trials are also missing from this literature, which are critical to determine the effectiveness of educational interventions delivered through the social media channel. These efforts could help to establish an evidence base for the now pervasive trend of integrating social media in higher education (Aldahdouh et al., 2020). This line of research could potentially help higher education institutions and educators to adapt social media platforms in the development of educational plans and teaching strategies. Finally, a wider use of quantitative behaviorally-based metrics is needed in order to further evaluate an operant learning model of online behavior in educational settings.

REFERENCES

References included in the review are highlighted with an asterisk.

- *Abd-El-Aal, W. M. M., & Steele, A. (2017). We all need we: the effect of using Facebook and group fieldwork on students' interdependence and awareness of STSE Issues. *World Journal of Education*, 7(1), 53. <https://doi.org/10.5430/wje.v7n1p53>
- *Abney, A., Cook, L., Fox, A. and Stevens, J., (2018). Intercollegiate Social Media Education Ecosystem. *Journal of Marketing Education*, 41(3), pp.254-269. <https://doi.org/10.1177/0273475318786026>

- *Akcaoglu, M., & Lee, E. (2018). Using Facebook groups to support social presence in online learning. *Distance Education*, 39(3), 334–352. <https://doi.org/10.1080/01587919.2018.1476842>
- * Al-Azawei, A. (2019). What drives successful social media in education and e-learning? A comparative study on Facebook and moodle. *Journal of Information Technology Education: Research*, 18(June), 253–274. <https://doi.org/10.28945/4360>
- *Albayrak, D., & Yildirim, Z. (2015). Using social networking sites for teaching and learning: Students' involvement in and acceptance of Facebook® as a course management system. *Journal of Educational Computing Research*, 52(2), 155–179. <https://doi.org/10.1177/0735633115571299>
- *Alberth. (2019). Use of Facebook, students' intrinsic motivation to study writing, writing self-efficacy and writing performance. *Technology, Pedagogy and Education*, 28(1), 21–36. <https://doi.org/10.1080/1475939X.2018.1552892>
- Aldahdouh, T. Z., Nokelainen, P., & Korhonen, V. (2020). Technology and social media usage in higher education: The influence of individual innovativeness. *SAGE Open*, 10(1). <https://doi.org/10.1177/2158244019899441>
- *Alghazo, Y. M., & Nash, J. A. (2017). The effect of social media usage on course achievement and behavior. *Journal of Education and Practice*, 8(2), 161–167. <http://files.eric.ed.gov/fulltext/EJ1131722.pdf>
- AlSaleem, B. I. (2018). The effect of Facebook activities on enhancing oral communication skills for EFL learners. *International Education Studies*, 11(5), 144. <https://doi.org/10.5539/ies.v11n5p144>
- *Altunkaya, H., & Topuzkanamış, E. (2018). The effect of using facebook in writing education on writing achievement, attitude, anxiety and self-efficacy perception. *Universal Journal of Educational Research*, 6(10), 2133–2142. <https://doi.org/10.13189/ujer.2018.061010>
- * Arabacioglu, T., & Akar-Vural, R. (2014). Using facebook as a LMS? *Turkish Online Journal of Educational Technology*, 13(2), 202–215.

- Anderson, R., Taylor, S., Taylor, T., & Virues-Ortega, J. (2021). Thematic and textual analysis methods for developing social validity questionnaires in applied behavior analysis. *Behavioral Interventions*. <https://doi.org/10.1002/bin.1832>
- Angus, D. (2017). Theme detection in social media. In L. Sloan and A. Quan-Haase (Eds.), *The SAGE Handbook of Social Media Research Methods* (pp. 530-544). SAGE.
- *Bacile, T. J. (2013). The Klout Challenge: Preparing your Students for Social Media Marketing. *Marketing Education Review*, 23(1), 87–92. <https://doi.org/10.2753/mer1052-8008230114>
- *Bahati, B. (2015). Extending student' discussions beyond lecture room walls via Facebook. *Journal of Education and Practice*, 6(15), 160–172. <http://iiste.org/Journals/index.php/JEP/article/view/22710>
- *Balcikanli, C. (2015). Prospective English language teachers ' experiences in Facebook: Adoption, use and educational use in Turkish context. *International Journal of Education and Development Using Information and Communication Technology*, 11(3), 82–99.
- * Bajko, R., Hodson, J., Seaborn, K., Livingstone, P., & Fels, D. (2016). Edugamifying Media Studies: Student Engagement, Enjoyment, and Interest in Two Multimedia and Social Media Undergraduate Classrooms. *Information Systems Education Journal (ISEDJ)*, 14(6), 14.
- Bishop, S. K., Moore, J. W., Dart, E. H., Radley, K., Brewer, R., Barker, L. K., Quintero, L., Litten, S., Gilfeather, A., Newborne, B., & Toche, C. (2020). Further investigation of increasing vocalizations of children with autism with a speech-generating device. *Journal of Applied Behavior Analysis*, 53(1), 475–483. <https://doi.org/10.1002/jaba.554>
- *Carver, J. (2019). InstaFrench: An investigation of learner perceptions of social media and images to develop L2 writing. *Dimension*, 8-27.
- *Chen, Y. C. (2015). Linking Learning Styles and Learning on Mobile Facebook. *International Review of Research in Open and Distance Learning*, 16(2), 94-114. <https://doi.org/10.19173/irrodl.v16i2.2038>

- *Clements, J. C. (2015). Using facebook to enhance independent student engagement: A case study of first-year undergraduates. *Higher Education Studies*, 5(4). <https://doi.org/10.5539/hes.v5n4p131>
- *Daniels, K.N & Billingsley, K.Y (2014). "Facebook"- It's not just for pictures anymore: The impact of social media on cooperative learning. *I-Manager's Journal of Educational Technology*, 11(3), 34.<https://doi.org/10.26634/JET.11.3.3008>
- Davidovitch, N., & Belichenko, M. (2018). Facebook tools and digital learning achievements in higher education. *Journal of Education and E-Learning Research*, 5(1), 8–14. <https://doi.org/10.20448/journal.509.2018.51.8.14>
- * Delen, I. (2017). Teaching argumentation by using facebook groups. *International Journal of Instruction*, 10(1), 151–168. <https://doi.org/10.12973/iji.2017.10110a>
- *Demirbilek, M. (2015). Social media and peer feedback: What do students really think about using Wiki and Facebook as platforms for peer feedback? *Active Learning in Higher Education*, 16(3), 211–224. <https://doi.org/10.1177/1469787415589530>
- *Dizon, G., & Thanyawatpokin, B. (2018). Web 2.0 tools in the EFL classroom: Comparing the effects of Facebook and blogs on L2 writing and interaction. *EuroCALL Review*, 26(1):29, 29-42. <https://doi.org/10.4995/eurocall.2018.7947>
- *Dougherty, K. D., & Andercheck, B. (2014). Using Facebook to engage learners in a large introductory course. *Teaching Sociology*, 42(2), 95–104. <https://doi.org/10.1177/0092055X14521022>
- *Duncan, D. G., & Barczyk, C. C. (2016). Facebook's effect on learning in higher education: An empirical investigation. *Information Systems Education Journal (ISEDJ)*, 14, 14–28.
- *Ercoskun, N.C, Ozan, C. and Kincal, R.Y. (2019). Investigation of Affinity towards Social Media and Expectations for Success of University Students. *Journal of Educational Issues*, 5(2), p.73. <https://doi.org/10.5296/jei.v5i2.14703>
- *Evans, C. (2014). Twitter for teaching: Can social media be used to enhance the process of learning? *British Journal of Educational Technology*, 45(5), 902–915. <https://doi.org/10.1111/bjet.12099>

- *Gamble, C., & Wilkins, M. (2014). Student attitudes and perceptions of using Facebook for language learning. *Dimension*, 49-72
- *Giannikas, C. (2019). Facebook in tertiary education: The impact of social media in e-learning. *Journal of University Teaching and Learning Practice*, 17(1). <https://doi.org/10.53761/1.17.1.3>
- *Goktalay, S. B. (2015). The impact of Facebook in teaching practicum: Teacher trainees perspectives. *Educational Research and Reviews*, 10(17), 2489–2500. <https://doi.org/10.5897/err2015.2446>
- *Gordon, J. (2016). How is language used to craft social presence in Facebook? A case study of an undergraduate writing course. *Education and Information Technologies*, 21(5), 1033–1054. <https://doi.org/10.1007/s10639-014-9366-0>
- *Gregory, P., Gregory, K., & Eddy, E. (2014). The Instructional Network: Using Facebook to enhance undergraduate mathematics instruction. *Journal of Computers in Mathematics & Science Teaching*, 33, 5–26.
- *Gregory, Peter L, Gregory, Karen M, & Eddy, Erik R. (2016). Factors contributing to student engagement in an instructional Facebook group for undergraduate mathematics. *The Journal of Computers in Mathematics and Science Teaching*, 35(3), 249.
- *Guo, R., Shen, Y., & Li, L. (2018). Using social media to improve student-instructor communication in an online learning environment. *International Journal of Information and Communication Technology Education*, 14(1), 33–43. <https://doi.org/10.4018/IJICTE.2018010103>
- *Harting, A. (2017). Using Facebook to improve L2 German students' socio-pragmatic skills. *The EuroCALL Review*, 25(1), p.26. <https://doi.org/10.4995/eurocall.2017.7014>
- *Hennessy, C. M., Kirkpatrick, E., Smith, C. F., & Border, S. (2016). Social media and anatomy education: Using twitter to enhance the student learning experience in anatomy. *Anatomical Sciences Education*, 9(6), 505–515. <https://doi.org/10.1002/ase.1610>
- Honig, W. K., & Staddon, J. E. R. (2022). *Handbook of Operant Behavior* (1st Ed.). Routledge. <https://doi.org/10.4324/9781003256670>

- *Hou, H. T., Wang, S. M., Lin, P. C., & Chang, K. E. (2015). Exploring the learner's knowledge construction and cognitive patterns of different asynchronous platforms: comparison of an online discussion forum and Facebook. *Innovations in Education and Teaching International*, 52(6), 610–620. <https://doi.org/10.1080/14703297.2013.847381>
- *Ibarra, F. D. E. (2018). Is Facebook Beneficial for Writing Practice? Ecuadorian Polytechnic Students Speak Up!. *Teaching English with Technology*, 18(3), 3–17.
- Koegel, L. K., Singh, A. K., & Koegel, R. L. (2010). Improving motivation for academics in children with autism. *Journal of autism and developmental disorders*, 40(9), 1057–1066. <https://doi.org/10.1007/s10803-010-0962-6>
- Kubina, R. M., & Lin, F.-Y. (2008). Defining frequency: A natural scientific term. *The Behavior Analyst Today*, 9(2), 125–129. <https://doi.org/10.1037/h0100651>
- Kubina, R.M., Morrison, R.S. Fluency in Education. *Behavior and Social Issues*, 10, 83–99. <https://doi.org/10.5210/bsi.v10i0.133>
- *Lee, H. Y., & Lee, H. W. (2016). Comparing social network analysis of posts with counting of posts as a measurement of learners' participation in facebook discussions. *Turkish Online Journal of Educational Technology*, 15(1), 11–19.
- Lei, H., Cui, Y., & Zhou, W. (2018). Relationships between student engagement and academic achievement: A meta-analysis. *Social Behavior and Personality: An International Journal*, 46(3), 517-528. <https://doi.org/10.2224/sbp.7054>
- *Lim, T. (2010). The use of facebook for online discussions among distance learners. *Turkish Online Journal of Distance Education*, 11(4), 72–81. <https://doi.org/10.17718/TOJDE.17195>
- Lindström, B., Bellander, M., Schultner, D. T., Chang, A., Tobler, P. N., & Amodio, D. M. (2021). A computational reward learning account of social media engagement. *Nature Communications*, 12, 1311. <https://doi.org/10.1038/s41467-020-19607-x>
- *Luo, T. (2018). Delving into the specificity of instructional guidance in social media-supported learning environments. *Journal of Information Technology Education: Innovations in Practice*, 17(March), 37–54. <https://doi.org/10.28945/3974>

- *Martínez-Cardama, S., & Caridad-Sebastián, M. (2019). Social media and new visual literacies: Proposal based on an innovative teaching project. *Education for Information*, 35(3), 337–352. <https://doi.org/10.3233/EFI-180214>
- *Miller, S. T. (2013). Increasing student participation in online group discussions via facebook. *Astronomy Education Review*, 12(1). <https://doi.org/10.3847/AER2012031>
- *Moghavvemi, S., & Salarzadeh Janatabadi, H. (2018). Incremental impact of time on students' use of E-learning via Facebook. *British Journal of Educational Technology*, 49(3), 560–573. <https://doi.org/10.1111/bjet.12545>
- *Montoneri, B. (2015). Impact of Students' Participation to a Facebook Group on their Motivation and Scores and on Teacher's Evaluation. *IAFOR Journal of Education*, 3(1), 61–74. <https://doi.org/10.22492/ije.3.1.04>
- *Montoneri, B. (2017). Facebook Posts as Complementary Teaching Material for a French University Course in Taiwan. *IAFOR Journal of Education*, 5(1), 141–162. <https://doi.org/10.22492/ije.5.1.08>
- *Naghdipour, B., & Eldridge, N. H. (2016). Incorporating social networking sites into traditional pedagogy: a case of Facebook. *TechTrends*, 60(6), 591–597. <https://doi.org/10.1007/s11528-016-0118-4>
- * Narayan, V., Herrington, J., & Cochrane, T. (2019). Design principles for heutagogical learning: Implementing student-determined learning with mobile and social media tools. *Australasian Journal of Educational Technology*, 35(3), 86–101. <https://doi.org/10.14742/ajet.3974>
- *Nazir, M., & Brouwer, N. (2019). Community of inquiry on Facebook in a formal learning setting in higher education. *Education Sciences*, 9(1). <https://doi.org/10.3390/educsci9010010>
- *Nkhoma, M., Cong, H. P., Au, B., Lam, T., Richardson, J., Smith, R., & El-Den, J. (2015). Facebook as a tool for learning purposes: Analysis of the determinants leading to improved students' learning. *Active Learning in Higher Education*, 16(2), 87–101. <https://doi.org/10.1177/1469787415574180>

- *Orawiwatnakul, W., & Wichadee, S. (2016). Achieving better learning performance through the discussion activity in facebook. *Turkish Online Journal of Educational Technology*, 15(3), 1–8.
- Ortigosa, A., Martín, J. M., & Carro, R. M. (2014). Sentiment analysis in Facebook and its application to e-learning. *Computers in Human Behavior*, 31(1), 527–541. <https://doi.org/10.1016/j.chb.2013.05.0244>
- * Owens, M., & Nussbaum, M. (2017). Twitter vs. Facebook: Using social media to promote collaborative argumentation in an online classroom. *Journal of Interactive Learning Research*, 28(3), 249-267.
- *Pai, A., McGinnis, G., Bryant, D., Cole, M., Kovacs, J., Stovall, K., & Lee, M. (2017a). Using Facebook groups to encourage science discussions in a large-enrollment biology class. *Journal of Educational Technology Systems*, 46(1), 103–136. <https://doi.org/10.1177/0047239516675898>
- *Pai, A., Cole, M., Kovacs, J., Lee, M., Stovall, K., & McGinnis, G. (2017b). As long as you are here, can i interest in you some science? Increasing student engagement by co-opting a social networking site, Facebook for science discussions. *Journal of Educational Technology Systems*, 46(2), 153–177. <https://doi.org/10.1177/0047239517729505>
- Papademetriou, C., Anastasiadou, S., Konteos, G., & Papalexandris, S. (2022). COVID-19 pandemic: the impact of the social media technology on higher education. *Education Sciences*, 12(4). <https://doi.org/10.3390/educsci12040261>
- Parks-Stamm, E. J., Zafonte, M., & Palenque, S. M. (2017). The effects of instructor participation and class size on student participation in an online class discussion forum. *British Journal of Education Technology*, 48(6), 1250-1259. <https://doi.org/10.1111/bjet.12512>
- *Peeters, W., & Pretorius, M. (2020). Facebook or fail-book: Exploring “community” in a virtual community of practice. *ReCALL*, 32(3), 291–306. <https://doi.org/10.1017/S0958344020000099>
- *Ping, N. S., & Maniam, M. (2015). The effectiveness of facebook group discussions on writing performance: A study in matriculation college. *International Journal of Evaluation and Research in Education (IJERE)*, 4(1), 30. <https://doi.org/10.11591/ijere.v4i1.4489>

- *Popescu, E., & Badea, G. (2020). Exploring a community of inquiry supported by a social media-based learning environment. *Educational Technology and Society*, 23(2), 61–76.
- *Purnamasari, A. (2019). Pre-Service EFL Teachers' Perception of Using Facebook Group for Learning. *JET (Journal of English Teaching)*, 5(2), 104. <https://doi.org/10.33541/jet.v5i2.1064>
- *Rahman, S., Ramakrishnan, T., & Ngamassi, L. (2020). Impact of social media use on student satisfaction in Higher Education. *Higher Education Quarterly*, 74(3), 304–319. <https://doi.org/10.1111/hequ.12228>
- *Riady, Y. (2014). Assisted learning through facebook: A case study of universitas terbuka's students group communities in Jakarta, Taiwan and Hong Kong. *Turkish Online Journal of Distance Education*, 15(2), 227–238. <https://doi.org/10.17718/tojde.71656>
- Rohrer, D., & Wixted, J. T. (1994). An analysis of latency and interresponse time in free recall. *Memory & Cognition*, 22(5), 511–524. <https://doi.org/10.3758/bf03198390>
- *Rubrico, J.G.U. & Hashim, F. (2014). Facebook-photovoice interface: Empowering non-native pre-service English language teachers. *Language Learning & Technology*, 18(3), 16–34.
- *Saifudin, A. M., Yacob, A., & Saad, R. (2016). The Facebook-in-action: Challenging, Harnessing and Enhancing Students Class Assignments and Projects. *Universal Journal of Educational Research*, 4(6), 1259–1265. <https://doi.org/10.13189/ujer.2016.040602>
- *Schroeder, J., & Greenbowe, T. J. (2009). The Chemistry of Facebook: Using Social Networking to Create an Online Community for the Organic Chemistry Laboratory. *Innovate: Journal of Online Education*, 5(4), 1–11.
- Sehnaz, B. G. (2015). The impact of Facebook in teaching practicum: Teacher trainees perspectives. *Educational Research and Reviews*, 10(17), 2489–2500. <https://doi.org/10.5897/err2015.2446>
- *Sheeran, N., & Cummings, D. J. (2018). An examination of the relationship between Facebook groups attached to university courses and student engagement. *Higher Education*, 76(6), 937–955. <https://doi.org/10.1007/s10734-018-0253-2>

- *Shih, R. C. (2011). Can Web 2.0 technology assist college students in learning English writing? Integrating Facebook and peer assessment with blended learning. *Australasian Journal of Educational Technology*, 27(5), 829-845. <https://doi.org/10.14742/ajet.934>
- *Sittiwong, T., & Wongnam, T. (2015). The effective of using 5 simple steps (QSCCS) learning activities on facebook to promote self-learning in the 21st century in technology printing and advertising course for undergraduate students in education technology and communications. *Universal Journal of Educational Research*, 3(11), 843–846. <https://doi.org/10.13189/ujer.2015.031110>
- *Slim, H., & Hafedh, M. (2019). Social media impact on language learning for specific purposes: A study in English for business administration. *Teaching English with Technology*, 19(1), 56–71.
- Statista (2022, March). Facebook: distribution of global audiences 2022, by age and gender. <https://www.statista.com/statistics/376128/facebook-global-user-age-distribution/>
- Statista. (2021, November). Global social networks ranked by number of users 2021 <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>
- Tawafak, R. M., AlFarsi, G., Jabbar, J., Malik, S. I., Mathew, R., AlSidiri, A., Shakir, M., & Romli, A. (2021). Impact of technologies during COVID-19 pandemic for improving behavior intention to use e-learning. *International Journal of Interactive Mobile Technologies*, 15(1), 184–198. <https://doi.org/10.3991/IJIM.V15I01.17847>
- *Teixeira, S., & Hash, K. M. (2017). Teaching Note—Tweeting macro practice: social media in the social work classroom. *Journal of Social Work Education*, 53(4), 751–758. <https://doi.org/10.1080/10437797.2017.1287025>
- Thelwall, M. (2017). Sentiment analysis. In L. Sloan and A. Quan-Haase (Eds.), *The SAGE Handbook of Social Media Research Methods* (pp. 545-556). SAGE.
- *Tran, P. (2016). Training learners to use Quizlet vocabulary activities on mobile phones in Vietnam with Facebook. *JALT CALL Journal*, 12(1), 43–56. <https://doi.org/10.29140/jaltcall.v12n1.201>

- *Tucker, V. (2015). Using social media for student collaboration. *Journal of Faculty Development*, 29(2), 45–56.
- *Tur, G., & Marín, V. I. (2014). Enhancing learning with the social media: student teachers' perceptions on Twitter in a debate activity. *Journal of New Approaches in Educational Research*, 4(1), 46–43. <https://doi.org/10.7821/naer.2015.1.102>
- Undisclosed authors. (2022). *Systematic Review of Quantitative Indices of Student Social Media Engagement in Tertiary Education [Dataset]*. Figshare. <https://doi.org/10.6084/m9.figshare.20105219>
- *VanDoorn, G., & Eklund, A. A. (2013). Face to Facebook: Social media and the learning and teaching potential of symmetrical, synchronous communication. *Journal of University Teaching & Learning Practice*, 10(1). <https://doi.org/10.53761/1.10.1.6>
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27, 3, 425. <https://doi.org/10.2307/30036540>
- *Wang, C.H., Shannon, D.M. & Ross, M.E. (2013). Students' characteristics, self-regulated learning, technology self-efficacy, and course outcomes in online learning. *Distance Education*, 34(3), 302-323. <https://doi.org/10.1080/01587919.2013.835779>
- *Whittaker, A. L., Howarth, G. S., & Lymn, K. A. (2014). Evaluation of Facebook© to create an online learning community in an undergraduate animal science class. *Educational Media International*, 51(2), 135–145. <https://doi.org/10.1080/09523987.2014.924664>
- *Wu, J. Y., Hsiao, Y. C., & Nian, M. W. (2020). Using supervised machine learning on large-scale online forums to classify course-related Facebook messages in predicting learning achievement within the personal learning environment. *Interactive Learning Environments*, 28(1), 65–80. <https://doi.org/10.1080/10494820.2018.1515085>
- *Yagci, T. (2015). Blended learning via mobile social media & Implementation of “EDMODO” in reading classes. *Advances in Language and Literary Studies*, 6(4). <https://doi.org/10.7575/aiac.all.v.6n.4p.41>

- *Yu, L.-T. (2014). A case study of using Facebook in an EFL English writing class: The perspective of a writing teacher. *The JALT CALL Journal*, 10(3), 189–202. <https://doi.org/10.29140/jaltcall.v10n3.175>
- *Zhang, Q., & Lu, Z. (2014). The writing of Chinese characters by CFL learners: Can writing on Facebook and using machine translation help? *Language Learning in Higher Education*, 4(2), 441–467. <https://doi.org/10.1515/cercles-2014-0023>
- Zhou, Jin; Ye, Jun-min (2020). *Sentiment analysis in education research: a review of journal publications*. *Interactive Learning Environments*. <https://doi.org/10.1080/10494820.2020.1826985>

Table 8.1 Characteristics of Included Studies (n = 75)

Study	n	Age	Gender (% female)	Field of study	SMP	Country	Level	Study Design	Engagement metric	Outcome variables	
										Quantitative ¹	Qualitative
Abd-El-Aal et al. (2017)	99	NR	88%	SS	F	Egypt	UG	Qualitative	SIT (ad hoc)	N/A	Satisfaction (ad hoc SIT)
Abney et al. (2019)	144	21	49.3%	B	T	US	UG	Observational	SIT (ad hoc)	N/A	Achievement (SIT)
Akcaoglu et al. (2018)	62	adults	NR	PC	F	US	NU	Intervention (WS)	SIT (ad hoc)	N/A	Satisfaction (ad hoc SIT)
Al-Azawei (2019)	143	18-20	47.6	C&T	F	Iraq	UG	Observational	SIT (ad hoc)	N/A	Satisfaction (ad hoc SIT)
Albayrak et al. (2015)	42	NR	NR	P&L	F	Turkey	UG	Qualitative	SIT (ad hoc)	N/A	Satisfaction (ad hoc SIT)
Alberth (2019)	64	NR	NR	L&C	F	Indonesia	UG	Intervention (WS)	SIT (ad hoc)	Achievement, writing skills (objective)	Other (SIT)
Alghazo et al. (2018)	322	NR	34.2%	SS	W	Saudi Arabia	UG	Intervention (BG)	Behavioral (# missed sessions)	Achievement (objective)	N/A
Altunkaya et al. (2018)	96	NR	NR	SS	F	Turkey	UG	Intervention (BG)	SIT (ad hoc)	N/A	N/A
Arabacioglu et al., (2014)	42	NR	61.9%	C&T	F	Turkey	UG	Intervention (BG)	SIT (ad hoc)	N/A	Other (SIT)
Bacile (2013)	86	NR	NR	B	F/T/I	US	UG	Intervention (WS)	Behavioral (# posts, likes)	N/A	N/A

Study	n	Age	Gender (% female)	Field of study	SMP	Country	Level	Study Design	Engagement metric	Outcome variables	
										Quantitative ¹	Qualitative
Bahati (2015)	84	NR	28.6%	SS	F	Rwanda	G	Qualitative	SIT (ad hoc)	N/A	Satisfaction (ad hoc SIT)
Bajko et al. (2016)	76	18-40	64.5%	C&T	NR	Canada	UG	Intervention (BG)	SIT (ad hoc)	N/A	Satisfaction (ad hoc SIT)
Balcikanli (2015)	113	18-23	85%	L&C	F	Turkey	UG	Qualitative	SIT (ad hoc)	N/A	N/A
Carver (2019)	83	NR	NR	L&C	I	US	UG	Intervention (WS)	SIT (ad hoc)	N/A	Satisfaction (ad hoc SIT)
Chen (2015)	134	18-40	48.6	SS	F	Taiwan	UG	Intervention (WS)	SIT (ad hoc)	Achievement (objective)	Satisfaction (ad hoc SIT)
Clements (2015)	78	NR	NR	P&L	F	Canada	UG	Intervention (WS)	Behavioral (# posts, comments, likes)	Achievement (objective)	Other (SIT)
Daniels et al. (2014)	20	19-34	NR	NR	F	US	UG	Intervention (WS)	Behavioral (# posts)	N/A	N/A
Delen, (2017)	58	NR	NR	SS	F	Turkey	UG	Intervention (WS)	Behavioral (# comments)	Achievement (objective)	N/A
Demirbilek (2015)	51	NR	80%	C&T	F	Turkey	UG	Qualitative	SIT (ad hoc)	N/A	Other (SIT)
Dizon et al. (2016)	23	NR	NR	L&C	F	Japan	UG	Intervention (WS)	SIT (ad hoc)	N/A	N/A

Study	n	Age	Gender (% female)	Field of study	SMP	Country	Level	Study Design	Engagement metric	Outcome variables	
										Quantitative ¹	Qualitative
Dougherty et al. (2014)	170	NR	NR	SS	F	US	UG	Observational	Behavioral (# posts and likes)	Achievement, completed assignments (objective)	Satisfaction (ad hoc SIT)
Duncan et al. (2016)	586	18-+25	49.1%	PC	F	US	UG	Intervention (BG)	SIT (ad hoc)	N/A	Satisfaction (ad hoc SIT)
Ercoskun et al., 2019	1450	NR	68%	SS	NR	Turkey	UG	Observational	Behavioral (# of posts)	NA	NA
Evans (2014)	252	18-24	51%	PC	T	UK	UG	Qualitative	SIT (ad hoc)	N/A	N/A
Gamble et al. (2014)	97	18	48.4 %	L&C	F	Japan	UG	Qualitative	SIT (ad hoc)	N/A	Satisfaction (ad hoc SIT)
Giannikas (2020)	14	25-60	75%	L&C	F	Republic of Cyprus	UG	Qualitative	SIT (ad hoc)	N/A	Satisfaction (ad hoc SIT)
Goktalay (2015)	41	NR	86%	SS	F	Turkey	UG	Intervention (BG)	Standardized test (UTAUT)	N/A	Achievement (SIT)
Gordon (2016)	21	NR	NR	C&T	F	US	UG	Observational	Behavioral (# posts, comments, likes)	N/A	N/A
Gregory et al. (2014)	78	NR	32.1%	PC	F	US	UG	Intervention (BG)	SIT (ad hoc)	N/A	Satisfaction and achievement (ad hoc SIT)
Gregory et al. (2016)	138	NR	NR	P&L	F	US	UG	Qualitative	SIT (ad hoc)	N/A	N/A

Study	n	Age	Gender (% female)	Field of study	SMP	Country	Level	Study Design	Engagement metric	Outcome variables	
										Quantitative ¹	Qualitative
Guo et al. (2018)	129	18-24	NR	B	F	US	UG	Intervention (BG)	SIT (ad hoc)	N/A	Satisfaction (ad hoc SIT)
Harting (2017)	9	NR	NR	L&C	F	Japan	UG	Intervention (WS)	Behavioral (# posts)	Achievement (objective)	Satisfaction (ad hoc SIT)
Hennessy et al. (2016)	150	20 a 35	53.3%	P&L	F	UK	UG	Qualitative	SIT (ad hoc)	Achievement (objective)	Satisfaction (ad hoc SIT)
Hou et al. (2015)	50	NR	NR	C&T	F	Taiwan	UG	Observational	Behavioral (# posts and comments)	N/A	N/A
Ibarra (2018)	30	NR	NR	L&C	F	Ecuador	NU	Intervention (WS)	SIT (ad hoc)	Achievement (objective)	Satisfaction (ad hoc SIT)
Lee et al. (2016)	108	NR	NR	NR	F	South Korea	NR	Intervention (BG)	Behavioral (semantic analysis metric for comments)	N/A	N/A
Luo (2018)	24	19-22	NR	SS	T	US	UG	Observational	Behavioral (# posts and characters per post)	N/A	N/A
Martinez-Cardama et al. (2019)	92	NR	NR	C&T	T	Spain	UG	Observational	Behavioral (# specific hashtags)	N/A	N/A
Miller (2013)	59	NR	NR	P&L	F	US	UG	Intervention (WS)	Behavioral (# posts and comments per student)	Achievement (objective)	N/A
Moghavvemi et al. (2018)	170	NR	NR	P&L	F	Malaysia	UG	Intervention (WS)	Standardized test (UTAUT)	N/A	N/A

Study	n	Age	Gender (% female)	Field of study	SMP	Country	Level	Study Design	Engagement metric	Outcome variables	
										Quantitative ¹	Qualitative
Mon-toneri (2015)	23	NR	NR	L&C	F	Taiwan	UG	Observational	Behavioral (# likes and views by post)	Achievement (objective)	Satisfaction (ad hoc SIT)
Mon-toneri (2017)	32	NR	NR	L&C	F	Taiwan	UG	Qualitative	SIT (ad hoc)	N/A	Other (SIT)
Naghdi-pour et al. (2016)	25	NR	52%	L&C	F	Turkey	UG	Observational	Behavioral (# comments and posts and comment-generating posts)	N/A	N/A
Narayan et al. (2019)	336	NR	NR	SS	T	New Zealand	UG	Intervention (WS)	Behavioral (# tweets, hashtags and posts)	N/A	Other (SIT)
Nazir & Brouwer (2019)	74	20-34	30%	NR	F	Netherlands	UG	Intervention (BG)	Behavioral (students and instructors # posts and comments)	N/A	Satisfaction (ad hoc SIT)
Nkhoma et al., (2015)	136	NR	NR	SS	F	Vietnam	UG	Intervention (WS)	SIT (ad hoc)	N/A	Achievement (ad hoc SIT)
Orawi-watnakul et al. (2016)	82	NR	NR	L&C	F	Turkey	UG	Intervention (WS)	Behavioral (# posts)	Achievement (objective)	Satisfaction (ad hoc SIT)
Owens et al., 2017	27	NR	93%	SS	T	US	UG	Intervention (WS)	Behavioral (# posts)	N/A	Other (SIT)
Pai et al. (2017a)	154	NR	NR	C&T	F	US	UG	Intervention (BG)	Behavioral (# posts, comments, likes)	N/A	N/A

Study	n	Age	Gender (% female)	Field of study	SMP	Country	Level	Study Design	Engagement metric	Outcome variables	
										Quantitative ¹	Qualitative
Pai et al. (2017b)	142	18	NR	C&T	F	US	UG	Intervention (BG)	Behavioral (# posts, comments, likes; textual length of posts and comments)	N/A	N/A
Peeters et al. (2020)	157	17-52	NR	L&C	F	Belgium	UG	Intervention (BG)	Behavioral (# posts across students)	N/A	N/A
Ping et al. (2015)	30	NR	NR	L&C	F	Malaysia	UG	Intervention (BG)	SIT (ad hoc)	Achievement (objective)	N/A
Popescu et al. (2020)	74	22-40	23%	SS	T	Romania	UG	Intervention (BG)	Behavioral (# tweets)	N/A	N/A
Purnamasari (2019)	56	NR	NR	L&C	F	Indonesia	UG	Observational	SIT (ad hoc)	N/A	Satisfaction (ad hoc SIT)
Rahman et al. (2020)	108	16-35	54.2%	SS	F	US	UG	Observational	SIT (ad hoc)	N/A	Satisfaction (ad hoc SIT)
Riady (2014)	1194	NR	NR	SS	F	Indonesia	UG	Qualitative	Behavioral (# posts, links, photos, events, updates, shared docs)	N/A	N/A
Rubrico et al. (2014)	59	19-27	98.3%	L&C	F	Malaysia	UG	Intervention (WS)	SIT (ad hoc)	N/A	Satisfaction (ad hoc SIT)
Saifudin et al. (2016)	58	22-26	57%	SS	F	Malaysia	UG	Intervention (WS)	SIT (ad hoc)	N/A	N/A

Study	n	Age	Gender (% female)	Field of study	SMP	Country	Level	Study Design	Engagement metric	Outcome variables	
										Quantitative ¹	Qualitative
Schroeder et al. (2009)	128	NR	NR	C&T	F	US	UG	Observational	Behavioral (# posts)	N/A	N/A
Sheeran et al. (2018)	471	17-59	76.7%	SS	F	Australia	UG	Intervention (BG)	SIT (ad hoc)	N/A	Satisfaction (ad hoc SIT)
Shih (2011)	23	NR	78.3%	L&C	F	Taiwan	UG	Intervention (WS)	Behavioral (# comments and likes)	Achievement (objective)	Satisfaction (ad hoc SIT)
Sittiwong et al. (2015)	38	NR	NR	C&T	F	Thailand	UG	Qualitative	SIT (ad hoc)	N/A	N/A
Slim et al. (2019)	102	NR	NR	L&C	F	Saudi Arabia	UG	Intervention (BG)	SIT (ad hoc)	Achievement (objective)	N/A
Teixeira et al. (2017)	45	NR	NR	SS	T	US	UG	Qualitative	SIT (ad hoc)	N/A	Satisfaction (ad hoc SIT)
Tran (2016)	21	NR	14.3%	L&C	F	Vietnam	UG	Observational	Behavioral (# posts and likes)	Achievement (objective)	Satisfaction (ad hoc SIT)
Tucker (2015)	16	NR	75%	PC	F	US	UG	Intervention (BG)	SIT (ad hoc)	N/A	Other (SIT)
Tur et al. (2015)	153	NR	NR	SS	T	Spain	UG	Observational	Behavioral (# tweets, retweets, and comments)	N/A	N/A
Van-Doorn et al. (2013)	20	NR	NR	SS	F	Australia	UG	Qualitative	SIT (ad hoc)	N/A	Achievement (SIT)
Wang et al. (2013)	415	NR	63.2%	SS	F	Taiwan	UG	Observational	SIT (ad hoc)	Personality tests (Standardized)	N/A

Study	n	Age	Gen-der (% fe-male)	Field of study	SMP	Coun-try	Lev-el	Study Design	Engage-ment met-ric	Outcome variables	
										Quanti-tative ¹	Qualita-tive
Whittak-er et al. (2014)	42	NR	NR	P&L	F	Aus-tralia	UG	Observa-tional	Behavioral (# posts)	N/A	N/A
Wu et al. (2018)	24	NR	52.2%	C&T	F	Taiwán	UG	Observa-tional	Behavioral (# posts and av-erage of words)	N/A	N/A
Yagci (2015)	177	NR	54.2%	L&C	F	Iraq	UG	Qualita-tive	SIT (ad hoc)	N/A	N/A
Yu (2014)	NR	NR	NR	L&C	F	Taiwan	UG	Qualita-tive	Behavioral (# posts)	N/A	Satisfac-tion (ad hoc SIT)
Zhang et al. (2014)	41	18-29	NR	L&C	F	Ireland	UG	Interven-tion (BG)	Behavioral (# charac-ters)	Achieve-ment (objec-tive)	N/A

Notes. B = Business; C&T = Computer and technology; F = Facebook; G = Graduate; L&C = Language & communication; N/A = Not applicable; NR = Not reported; NU = Non-university course; P&L = Physical & life sciences; P&L = Physical & life sciences; PC = Professional courses; SIT = Survey or interview; SMP = Social media platform; SS = Social sciences; T = Twitter; UG = College/Undergraduate; UTAUT = User acceptance of information technology (Venkatesh et al., 2003); W = WhatsApp. 1. On occasions when the quantitative outcome variable was an engagement metric, this is reported in the column Engagement measure.

Table 8.2 Objective Measures of Social Media Engagement in the Tertiary Education Literature

Reference	Platform	Focus	Frequency			Other
			Posts	Comments	Reactions	
Bacile (2013)	Facebook	Group	•		•	
Clements (2015)	Facebook	Group	•	•	•	
Daniels et al. (2014)	Facebook	Group	•			
Dougherty et al. (2014)	Facebook	Group	•		•	
Ercoskun et al., 2019	Facebook	Group	•			
Gordon (2016)	Facebook	Group	•	•	•	
Harting (2017)	Facebook	Group	•			
Hou et al. (2015)	Facebook	Group	•	•		
Lee et al. (2016)	Facebook	Group	•			
Lim (2010)	Facebook	Group	•	•		
Luo (2018)	Facebook	Group	•	•		•
Martinez-Cardama et al. (2019)	Twitter	Group				•
Miller (2013)	Facebook	Student	•	•		
Montoneri (2015)	Facebook	Post			•	•
Naghdipour et al. (2016)	Facebook	Group	•	•		
Narayan et al. (2019) ¹	Twitter	Group	•	•	•	
Nazir & Brouwer (2019)	Facebook	Group	•	•		
Orawiwatnakul et al. (2016)	Facebook	Group	•			
Owens et al. (2017)	Facebook	Group	•			
Pai et al. (2017a)	Facebook	Group	•	•	•	
Pai et al. (2017b)	Facebook	Group	•	•	•	
Peeters et al. (2020)	Facebook	Group	•	•		
Popescu et al. (2020) ¹	Twitter	Group	•			
Riady (2014)	Facebook	Group	•			•
Schroeder et al. (2009)	Facebook	Group	•			
Shih (2011)	Facebook	Group		•	•	
Tran (2016)	Facebook	Group	•		•	
Tur et al. (2015) ¹	Twitter	Group	•	•	•	
Whittaker et al. (2014)	Facebook	Group	•			
Wu et al. (2018)	Facebook	Group	•	•		•
Yu (2014)	Facebook	Group	•			
Zhang et al. (2014)	Facebook	Group		•		•

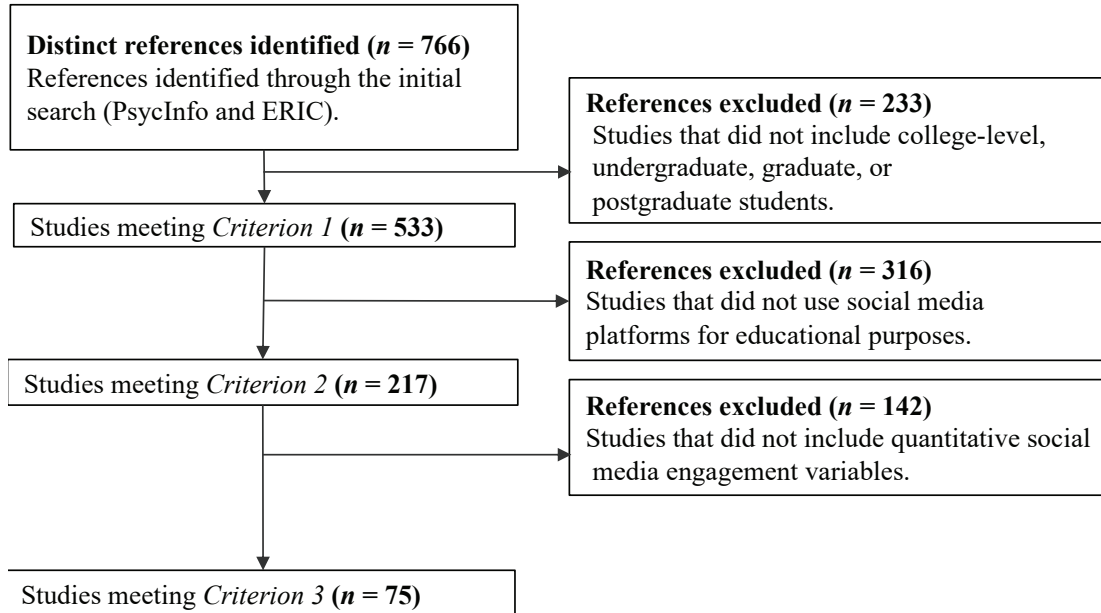
Note. 1. Posts, comments, and reactions are considered equivalent to tweets, replies, and retweets, respectively.

Table 8.3 A Proposed Taxonomy of Quantitative Metrics of Social Media Engagement

Metric	Definition	Focus	Example
Count-based			
Posting frequency	Total number of posts over a period of time (e.g., week, semester).	G, I	Pai et al. (2017a)
Commenting frequency	Total number of comments over a period of time (e.g., week, semester).	G, I, P	Naghdi-pour et al. (2016)
Reacting frequency	Total number of reactions over a period of time (e.g., week, semester).	G, I, P	Montoneri (2015)
Percentage individuals posting	Ratio of the total number of students publishing to the total number of posts.	G	N/A
Percentage individuals commenting	Ratio of the total number of students commenting to the total number of comments in the study group.	G, P	N/A
Student-instructor posting ratio	Ratio between the number of posts published by a specific group of people (e.g., students) and the total number of posts published (e.g., students and faculty members).	P	Nazir & Brouwer (2019)
Student-instructor commenting ratio	Ratio between the number of comments posted by a specific group of people (e.g., students) and the total number of comments posted (e.g., students and faculty members).	P	Nazir & Brouwer (2019)
Percentage individuals reacting	Ratio between the number of reactions posted by a specific group of people (e.g., students) and the total number of reactions posted (e.g., students and faculty).	P	N/A
Time-based			
Comment latency	Time elapsed between the appearance of a post and the appearance of a comment.	G, I, P	Lindström (2021)
Reaction latency	Time elapsed between the appearance of a post and the appearance of a reaction to that post	G, I, P	N/A
Comment inter-response time	Time elapsed between the emission of a comment made by a person and the next comment.	I, P	N/A
Reaction inter-response time	Time elapsed between the reaction issued by one person to the next reaction.	I, P	N/A
Text analysis			
Post length	Total number of characters in a post	G, I, P	Pai et al. (2017)
Comment length	Total number of characters in a comment	G, I, P	Pai et al. (2017)
Theme analysis	Occurrence and co-occurrence of words appearing in a text using an algorithmic process generated by software	G, I, P	Angus (2017)
Sentiment analysis	Total occurrence of terms analyzed by lexicon algorithms using software to detect sentiment-related patterns	G, I, P	Thelwall (2017)

Note. I = Individual; G = Group; P = Post. N/A = None available

Figure 8.1 Study Selection Flowchart



CHAPTER 9. General Methodological Approach: Conclusions

The preceding chapters lay a solid ground upon which empirical studies can be undertaken. In Chapter 5 we identified the methodological basis upon which behavioral education studies have traditionally relied upon. In Chapter 6 we provide a comprehensive outlook of the various analytical strategies that can be utilized for single-subject and group-based analyses in behavioral education. In Chapter 7 we take an in-depth transversal view into one of the methodological challenges that had been identified earlier. Specifically, we evaluated the precision of computer-based and paper-and-pencil behavioral observation systems. This assessment is important, in that it provide a firm bases for the subsequent use of these systems during the empirical-interventional section of the current dissertation. Finally, in Chapter 8 we present a systematic review and a taxonomy of quantitative and behaviorally-based metrics that can be utilized both as outcomes and as explanatory variables during our applied work. Overall, this ample collection of methodological studies allows us to draw the following conclusions that will be critical in carving out our methods in subsequent applied studies.

1. SYNERGY OF SINGLE-SUBJECT AND GROUP-BASED ANALYSES

- Even though the prevalent culture in behavior analysis is that single-subject analysis is the approach to testing novel interventions, there is a growing methodological armamentarium for combining single-subject analyses and randomized controlled trials (Virues-Ortega, 2022, Virues-Ortega et al., 2022a). This approach can be traced back to the work of Nathan Azrin (1930-2013) as recently highlighted in a monographic review (Virues-Ortega et al., 2021). There are distinct advantages of adopting this diverse approach, particularly, in the field of applied behavioral education. Specifically, a combined single-subject and group-based approach could, within the same study, (1) ascertain idiographic patterns, (2) weight the relative contribution of various intervention components (i.e., component analysis), (3) take full advantage of the analytical strategies currently available for single-subject datasets (i.e., hierarchical linear models), and (4) access wider audiences beyond the *niche* readership of applied behavior

analysis with the end goal of effecting evidence-based educational practices. While the despise towards statistical analysis in academic behavior-analytic circles is deeply rooted in the history of the field (see, for example, Michael, 1974), the philosophical and methodological considerations that led to the neglect of group-based studies are no longer relevant, at least in all contexts. The evidence reviewed is consistent with the view that it may be possible to use a combined single-subject and group-based approach for evaluating a multi-component behavioral education program via a social media platform in tertiary education.

2. VALIDITY AND PRECISION OF COMPUTER-BASED SYSTEMS FOR VIDEO-BASED BEHAVIORAL OBSERVATION

- As it has been discussed at length in Chapter 5, one of the keystones of behavioral intervention programs in education is their procedural integrity. Namely, applied procedures should be delivered as intended. Procedural integrity is integral to process-based interventions as well. Analytical approaches to multi-component interventions programs, including component analysis and multi-armed RCTs (Jaki & Wason, 2018; Ward-Horner & Sturmey, 2010) have the potential to discern the incremental validity of various intervention elements. However, characterizing these elements or processes in any detail would require a close monitoring of procedural integrity standards. The behavioral procedures that we have described as distinctively amenable to the social media channel in previous chapters include video broadcasts and asynchronous video posts. Therefore, it is critical to our procedural integrity strategy to use valid and precise video-based behavioral observation systems. The study presented in Chapter 7, which has been recently published in a top journal (Virues-Ortega et al., 2022b), presents convincing evidence that a purposely-developed mobile app can be used in precision in this context.

3. QUANTITATIVE ENGAGEMENT OUTCOMES AND A CONCEPTUAL MODEL FOR EVIDENCE-BASED INTERVENTIONS

- Lindström et al. (2021) provide a proof-of-concept study on the necessity of harvesting time-stamped user responses generated in the social media environment (i.e., re-

actions, comments, posts, views). Specifically, collecting social media user responses over time may allow to assess whether social media engagement may be a mediating factor in any indirect effects of educational interventions (as delivered via a social media platform) may have on achievement. This approach could provide some preliminary evidence on the potential of social reinforcement as a general process underlying the effects of specific educational practices that can be delivered through a social media environment. For example, a potential finding of stable commenting an increased academic achievement in social contexts where social reinforcement is highly likely (e.g., cooperative learning environment, semi-immediate instructor feedback) on the one hand, and unstable commenting and limited achievement in social media contexts where social reinforcement is less likely (e.g., lack of cooperative learning environmental design, delayed instructor feedback) on the other, suggest that social reinforcement may be a key process. However, the systematic review presented in Chapter 7 clearly indicates that quantitative engagement outcomes are rarely collected in the literature, thereby limiting our ability to ascertain key explanatory processes. The systematic review and engagement outcome taxonomy proposed in Chapter 7 provides a compelling conceptual basis that can be utilized in the applied studies that are presented in subsequent chapters.

4. PROCESS-BASED INTERVENTIONS MAY REQUIRE COMPONENT ANALYSIS AND ULTIMATELY MULTI-ARM RCTS

- Treatment evaluation and optimization is essential to evidence-based practices. The preceding chapters have emphasized the conceptual and methodological analyses needed to define a multi-component intervention and evaluate its preliminary effects on online and academic behavior (i.e., social media engagement, academic achievement). A further step in the applied scientific process is optimization, which in the particular case of multi-component interventions cannot be achieved without complex research designs including component analyses and multi-arm RCTs. The latter approaches allow the researcher to establish the relative contribution of the various elements of an intervention program, which can lead to further refinements. For example, certain components may be omitted, while others may be modified in form or

intensity. To this end, we intend to use a multi-arm RCT as the last empirical study in the current thesis project.

- In the following chapter we will attempt to bring to fruition the methodological standards that have been introduced in the preceding paragraphs.

CUARTA PARTE
ESTUDIOS EXPERIMENTALES-APLICADOS

CHAPTER 10. Effect of the Behavioral Education and Social Media (BE-Social) Program on Postgraduate Academic Performance: A Randomized Controlled Trial⁵³

ABSTRACT

Few randomized controlled trials have evaluated social media functionalities as educational aids in the context of online and blended teaching programs. We present the Behavioral Education and Social Media (BE-Social) intervention package, which integrates a collection of key evidence-informed behavioral intervention procedures delivered through a closed social media learning group. BE-Social integrates instructor-mediated cooperative learning, self-management, and video modeling components. A total of 46 students were randomly assigned to a default online program (control) or default online program plus BE-Social (intervention) group. Target outcomes included academic performance and social media engagement (reactions, comments). Results indicated that the intervention was effective in producing a 20-point increase in academic performance over 100-point scale and large increases in social media engagement. A parallel single-subject analysis revealed that intervention effects were, to a considerable extent, idiosyncratic. The theoretical, methodological, and practical implications of the study are discussed.

Keywords: Adult learning Applications, Cooperative/collaborative learning, Post-secondary education, Social Media, Teaching/learning strategies

INTRODUCTION

Online education has increased steadily over the last 30 years experiencing an exponential leap during 2020 due to the COVID-19 pandemic (UNESCO, 2020). Online and blended learning

53 El presente estudio ha sido enviado a la revista *British Journal of Educational Technologies* y se presenta en el idioma y en el formato en el que se ha enviado.

Tarifa-Rodríguez, A., Virues-Ortega, J., Calero-Elvira, A. (2022). Effect of the Behavioral Education and Social Media (BE-Social) program on postgraduate academic performance: A randomized controlled trial. *British Journal of Educational Technologies*

programs are now near-universal in tertiary education in many countries. A radical change in the *channel* of teaching will inevitably modify the social dynamics motivating academic behaviors. For example, online programs may provide a greater opportunity for automated immediate feedback and audiovisual content rehearsal, thereby enhancing content acquisition. By contrast, important elements of the traditional context of learning may be missed. For example, students in online programs may have fewer opportunities for direct social interaction during academic and non-academic activities. In this connection, some authors have suggested that peer-to-peer and instructor-mediated social media experiences may be a proxy to the social environment of traditional education (Susanto et al., 2021)

As of April 2021, there are upwards of 15 billion active social media profiles across platforms worldwide (Statista, 2021). Consistently with the widespread use of social media among college and university students, some higher education institutions have integrated these tools as part of their courses (Schworm & Gruber, 2012; Tang & Hew, 2017). Yet, the potential of social media to enhance academic participation and performance has seldom been evaluated (Aslan, 2021). Specifically, it is unclear whether the added interaction opportunities for instructors and learners brought about by social media platforms may have a quantifiable impact on academic performance. In this connection, it has not been verified whether well-established behavioral education strategies could be delivered through social media platforms *in masse* with the end goal of making tertiary online education a more effective and cost-efficient endeavor.

1.1 Social media and education

Some studies have shown that over 90% of students use social media to discuss academic content with their peers (e.g., Madge et al., 2009). Junco (2012) stated that education professionals can take advantage of these widespread educational spaces with the end goal of improving student experience and performance (see similar claims by Hicks & Graber, 2010; and Smith, 2009). For example, a recent study surveyed 409 college students about the inclusion of the social media platform Telegram as a learning channel. The authors reported a significant correlation between self-reported use of the social network and increased engagement and academic performance (Mahdiuon et. al., 2020, see also Alshuaibi et al., 2018).

Among the numerous social media platforms available, several surveys have identified Facebook as the most popular among university students in various geographical contexts (Cheung et al., 2011; Michikyan et al., 2015; Wang et al., 2011). Facebook allows for the exchange

of synchronous and asynchronous information and facilitates interaction through thematic closed groups, instant messaging, push notifications, and other key functionalities (Mazman & Usluel, 2010). Researchers have often discussed the possibility of taking advantage of Facebook capabilities in education (see for example Garcia et al., 2015; Kalelioğlu, 2017, and Selwyn, 2009). In this connection, Al-Rahmi et al. (2015) surveyed 150 university students on how they perceived the effect of a closed Facebook group on course satisfaction, cooperative learning, and academic performance. The authors encountered a positive correlation between increased self-reported participation in the social network and curricular engagement (see also Davidovitch & Belichenko, 2018). Numerous survey studies, as the ones reported, suggest that the parallel use of Facebook as an interaction channel in tertiary education may promote collaboration, communication, peer-mediated feedback, and cognitive, and social skills (Arteaga-Sánchez et al., 2019; Aydin, 2012; Doleck & Lajoie, 2018; Hurt et al., 2012; Mazman & Usluel, 2010).

In spite of the many assertions brought about by observational studies in support of the potential role of social media in education, the evidence supporting the impact of social media on academic performance is mixed. Doleck and Lajoie (2018) summarized 23 peer-reviewed studies and failed to establish consistent gains in academic performance caused by adding social media platforms to formal university courses. In their concluding remarks the authors indicated that “the disagreement in the scholarship provides immense opportunities for further theoretical and empirical research to help push forward our understanding” (p. 463). A potential caveat to the existing literature may be that its focus has often been placed on the *channel* rather than the *educational practices*. An obvious research extension in this field would involve, first, identifying evidence-based behavioral education strategies that may be compatible with social media functionalities and, second, evaluate the quantitative effects of specific evidence-based practices when delivered through social medial channels in tertiary education settings.

1.2 Evidence-based educational practices and social media

Several evidence-based behavioral education strategies may be compatible with the social media channel. These include cooperative learning strategies (including immediate feedback through instant messaging and push notifications), and self-management and video modeling of effective study habits.

1.2.1 Cooperative learning

Cooperative learning has been identified as a predictor of increased academic performance, possibly due to its incremental effect on student participation and self-reported motivation (Ataie et al., 2015; van Ryzin et al., 2020). Key aspects of cooperative learning include timely sharing of information and resources, small-group learning, and positive interdependent interactions mediated by an instructor (see a meta-analysis for school-age applications in Little et al., 2014). In addition, peer-mediated feedback has been also identified as a central component of a cooperative learning environment. For example, a recent study by Yu and Chen (2021) indicated that receiving peer-mediated feedback in a university-level science course had greater impact on academic performance than instructor-mediated feedback. These elements can be readily delivered through social media channels in tertiary education applications. Since the 2000s authors have stated that higher education institutions can benefit from the interaction opportunities and instant communication in social media platforms (see for example, Pempek et al., 2009; and Yuan & Wu, 2020). Hsu (2018) evaluated the impact of instant messaging among first-year university students by providing quick turnaround instant messages, answering questions, providing corrective feedback pertaining to the curricular material, or simply praising student engagement. By contrast, the control group received only general encouragement statements (e.g., “Keep working hard!”) unrelated to the course material and independent from engagement in the course. The results indicated that students receiving immediate and relevant feedback via instant messaging showed greater course engagement and academic performance.

1.2.2 Self-management

Self-management strategies typically involve goal setting, performance self-monitoring, self-evaluation, and self-reinforcement components (Briesch & Chafouleas, 2009; Cooper et al., 2019). Self-management skills allow students to take an active role in the acquisition of adequate study habits; a key mediator of academic performance (see for example Credé & Kuncel, 2008). While self-management strategies have been used most often in organizational and clinical settings, a variety of applications among university studies have also been reported, whether learning to play a musical instrument, studying a new language, or presenting academic material, among others (Kitsantas, 2004; Malott, 2012). Instructor-mediated interactions in the context of social media groups may provide ample opportunity for students to acquire self-management skills. In

addition, incorporating self-management procedures may be a key addition to the online teaching-learning process, as instructor support may not always be available.

1.2.3 Video modeling

Video modeling has been defined as presenting a video of a model performing a target skill to a learner (Bellini & Akullian, 2007). Video modeling is often combined with instructions, prompts, and social praise for correct performance (Shukla-Mehta et al., 2010). The procedure has been found to be effective for the acquisition of social skills (Sherer et al., 2001), activities of daily living (Aldi et al., 2016), vocational skills (Schoenherr, 2018), and academic skills (Mitchem & Richards, 2003), to mention a few examples, among students with learning disabilities. While video modeling has not been evaluated with typically developed tertiary education students, it is readily compatible with the synchronous and asynchronous video functionalities of social media, for example, for the purposes of modeling appropriate study habits.

1.3 Goals of the study

After selecting key evidence-based behavioral education strategies that would be fully compatible with the social media channel (i.e., cooperative learning, self-management, video modeling), we aim to present and evaluate them as part of a behavioral intervention package for adult students. The Behavioral Education and Social Media (BE-Social) program presented here is comprised of the following intervention components: (a) semi-immediate instructor and peer-mediated feedback (thorough instant messaging and push notifications), and (2) study habits self-management and video modeling intervention (through synchronous and asynchronous video posts).

The goal of the current study is to evaluate the overall impact of the BE-Social program on social media engagement and academic performance in a cohort of college-level adult students. An ancillary goal involves evaluating the overall contribution of semi-immediate feedback to the BE-Social intervention package. Semi-immediate feedback was chosen as an assessment goal owing to the added cost of this element of the intervention. This line of research could help to establish social media functionalities as cost-effective additions to traditional and online tertiary education. This is the first randomized controlled trial (RCT) to evaluate the impact of social media aids on academic performance.

MATERIALS AND METHODS

2.1 Participants and setting

Participants were students of a one-year online post-graduate course in applied psychology delivered in Spanish. The course started in October 2019 and ended in May 2020. The curricular content was made available through a Moodle platform in weekly lessons. In addition, all students had access to a closed Facebook group (general Facebook group), whereas participants in the intervention group had access to an additional Facebook group through which the intervention was delivered (BE-Social Facebook group). Eighty students meeting the following inclusion criteria were invited to participate via an institutional email: (a) student had been inactive in the general Facebook group during the first three months of the course, and (b) student had an average performance in weekly multiple-choice course content tests below 80%. Participants were randomly assigned to the two arms of the study (i.e., BE-Social intervention group and control group). A total of 46 students agreed to participate; 28 were assigned to the BE-Social group and 18 to the control group. The age, sex, educational level, reading comprehension skills, and pre-entry course content knowledge of participants in both groups were not significantly different (see Table 1). All students that initially accepted participation remained in the study until its completion.

The study was conducted between February and May 2020. Throughout the duration of the study, students were experiencing various forms of lockdowns in their respective residential areas due to the COVID-19 pandemic. This situation did not affect the development of academic activities, as they were online. The methods of the study were approved by the ethics committee of the Universidad Autónoma de Madrid (approval number CEI 112-2204).

2.2 Dependent variables

2.2.1 Course content tests

Participants completed tests for each of the course weekly lessons. Each test was composed of 15 multiple-choice items with three distractors and one correct answer. We computed test scores by dividing correct responses by 15 and converting that ratio into a percentage. In order to minimize content-specific effects (e.g., lower performance in weeks with particularly challenging tests or course contents), participants were allowed to complete the test of a given lesson at any time within a three-week window starting on the week when the lesson was delivered. In

addition, participants were allowed to complete missing tests during the last three weeks of the study. This meant that the timing and sequence of tests varied across participants. Tests were delivered through a Moodle platform and were marked automatically.

2.2.2 Social media engagement outcomes

Social media engagement outcomes may be important to document the teaching-learning process. For example, reacting to a course-related post (e.g., clicking the like icon) suggests a degree of exposure to the material. In addition, commenting on a course-related post may indicate a degree of elaboration of the course content referred to in the post. We computed the frequency of reactions (e.g., likes) and comments (e.g., responses to an instructor's post) per week for each participant. An independent observer tallied Facebook reactions and comments for every participant on a weekly basis. For each participant in the control group, we added the weekly number of reactions and comments in the general Facebook group. For each participant in the intervention group, we added the total number of weekly reactions and the total number of weekly comments posted in either Facebook group by each participant as separate dependent variables (i.e., general and BE-Social Facebook groups).

2.3 Experimental design

We used a two-arm simple randomization RCT. Participants were not aware of the study hypotheses and goals. We studied idiosyncratic effects with a concurrent multiple-baseline design (Bailey & Burch, 2018). The RCT and the single-subject experimental design were fully compatible and were implemented concurrently (Virues-Ortega et al., in press). As part of the single-subject analysis, participants in the intervention group were randomly assigned to four pre-determined baseline lengths staggered in successive three-week blocks (with the exception of vacation breaks, lessons were delivered weekly). Twelve participants were assigned to a six-week baseline, 7 participants were assigned to a nine-week baseline, and 9 participants were assigned to a 13-week baseline (this period included the Eastern week break). Participants in the control group underwent an extended time-series without intervention (i.e., extended baseline). Figure 1 presents a diagram with the temporal distribution of the baseline and intervention phases for all groups and baseline lengths in successive weeks and calendar dates.

2.4 Procedure

Before the start of the first baseline, a course instructor created (and administered from that point onward) the two closed Facebook groups required to implement the intervention (i.e., general and BE-Social Facebook groups). In the sections below, we describe the procedures for the control group and the intervention group (see Figure 1 for the timeline of the study phases).

2.4.1 Baseline

All participants had access to the general Facebook group for a period of at least six weeks (Figure 1). Participants in the control group remained in the general Facebook group throughout the duration of the study. The general Facebook group was intended as an opportunity for students to gain additional exposure to the course. The instructor posted daily multiple-choice study questions. Students could not create novel posts but could respond to course content-related posts. The instructor provided general feedback once a week on each post immediately before closing the discussion for that post.

2.4.2 Intervention

In addition to their participation in the general Facebook group, individuals in the intervention group were invited to participate in the BE-Social Facebook group, which incorporated the following evidence-informed intervention components. The intervention was implemented for a three-week period.

2.4.2.1 Cooperative learning

The BE-Social group allowed student direct posting and peer- and instructor-mediated feedback. The instructor posted daily open-ended questions for participants to gain additional exposure to the material and create interaction opportunities. In addition, the instructor engaged in various actions intended to create a cooperative learning environment: (a) providing personalized semi-immediate feedback (within the same day) to participant comments including corrective feedback and social praise, (b) praising students that gave feedback to peers, and (c) creating posts and comments encouraging students to share materials they had elaborated (i.e., summaries, glossaries, flashcards). Social praise was specific; praise statements referred to discrete desirable behaviors (e.g., “Thanks for lending a hand to Lizbeth on question x!”). The instructor was available over a continuous four-hour period at designated times during weekdays throughout the intervention period.

2.4.2.2 Self-management and video modeling

The instructor created daily textual and image posts presenting selected study self-management strategies. Posts were intended to promote independent study habits and organizational skills. The instructor focused on study goal setting, self-monitoring, and self-reinforcement skills (see selection of study self-management posts in the Supplementary information, Appendix 1). Self-management posts remained accessible to participants for the duration of the intervention phase. In addition, the instructor streamed three one-hour video modeling sessions (two on Week 10, and one on Week 15) intended to enhance problem-solving skills when addressing applied scenarios and study self-monitoring skills. The instructor used elements of the behavioral skills training protocol by Parsons et al. (2013). Specifically, the video modeling session started with a description of the target skill. Then, a demonstration of the target skill with additional examples (i.e., problem-solving applied scenarios, engaging in study self-recording) followed by a 10-min Q&A session (see an example of the timeline of a video modeling post in the Supplementary material, Appendix 2). Video modeling sessions closed with the instructor encouraging students to practice the target skill and thanking those who attended. Videos remained accessible through the BE-Social Facebook group for the duration of the intervention phase.

2.4.2.3 Intervention with semi-immediate feedback

This phase lasted approximately two weeks, depending on national holidays. Participants were exposed to all elements of the BE-Social multicomponent package (self-management with video modeling, cooperative learning and semi-immediate feedback). The instructor provided feedback with a maximum latency of 24 hours.

2.4.2.4 Intervention without semi-immediate feedback

Given that individualized feedback was the most cost-intensive element of the multi-component BE-Social program, we limited semi-immediate instructor feedback to the first three weeks of intervention for each participant. Specifically, elements (a) and (b) of the cooperative learning component were only available for the first three weeks of the intervention. We conducted separate analyses to determine whether participants continued to engage with the group (by reacting to posts and posting comments) following the semi-immediate feedback period. Participants continued to be exposed to all other elements of the multi-component package during this phase

(i.e., peer feedback, self-management, and video modeling posts), but they did not receive individualized feedback from the instructor.

2.5 Interobserver agreement and procedural integrity

While data generation was automated, we studied the interobserver agreement (IOA) of the data extraction process. A secondary observer extracted occurrences of Facebook reactions and comments from 25% of the BE-Social Facebook group posts. An agreement was defined as the two observers recording equal number of reactions (or comments) for a given post. We computed IOA as the number of posts with agreement divided by the total number of posts evaluated. The IOA for both reactions and comments equaled 94%.

In order to evaluate whether key aspects of the intervention were delivered as intended, we documented the percentage of student posts in the BE-Social group that received feedback within a 24-h period, and the percentage of students in the BE-Social group that visualized posts published by the instructor or by peers. We studied procedural integrity in 25% of randomly selected instructor and student posts of the BE-Social group. Two independent observers extracted the number of visualizations per posts, the timestamp of the first student input, and the timestamp of the first instructor feedback. The analysis indicated that the mean percentage of students visualizing posts was 87% (range, 65% to 96%) and that 100% of the student's inputs received feedback within a 24-hour cycle (feedback latency: $M_e = 14$ min, $M = 3.9$ hours, range, 0.0 to 21.1 hours). The IOA of the procedural integrity data extraction process was 96%, 98%, and 94% for visualizations, student input timestamp, and instructor feedback timestamp, respectively.

2.6 Social validity

We collected unsolicited feedback from all study participants as provided after the end of the intervention phase through Facebook comments, messages, and emails. We then classified comments as positive, negative, or neutral, and computed the percentage of unsolicited positive feedback for participants in the control and intervention groups. In addition, we separated feedback comments into individual sentences and conducted a theme analysis with the 3rd Eye theme analysis algorithm (Hunerberg, 2019). The algorithm groups distinct sentences that share common sets of words. The output displays these groups as topics labeled with the words shared by all the elements within them. The elements of topics are not categorized exclusively, so distinct

elements may be grouped into multiple topics (e.g., “The videos helped me solve my many questions” could be integrated into four different topics: “video,” “helping,” “solving,” and “question”).

2.7 Analysis

We conducted a two-factor mixed-effect analysis of variance (ANOVA) with Time as repeated factor (baseline, intervention, intervention without semi-immediate feedback) and Group as between-subjects factor (control group, BE-Social intervention group). The analysis was repeated for the three main outcomes: course content tests, reactions, and comments. The effect of the intervention without semi-immediate feedback was evaluated only among individuals in the intervention group with short ($n = 12$) and medium ($n = 7$) baselines. All analyses were conducted with SPSS (IBM Corp., 2019). An alpha value of 0.05 was used throughout.

In order to determine whether the attained sample size led to a sufficiently powered analysis, we conducted a series of post-hoc power analyses with the G*Power software (Faul et al., 2009, *a priori* power analyses are available upon request). The achieved power values for course content tests were 1.00 ($\eta^2 = 0.34$, effect size $f = 0.72$, $r_{\text{baseline-to-treatment}} = 0.12$, critical $F = 4.06$, $n = 46$), 0.96 ($\eta^2 = 0.21$, effect size $f = 0.52$, $r_{\text{baseline-to-treatment}} = 0.12$, critical $F = 3.01$, $n = 46$), and 0.97 ($\eta^2 = 0.30$, effect size $f = 0.65$, $r_{\text{baseline-to-treatment}} = 0.12$, critical $F = 3.56$, $n = 46$) for the Group, Time, and interaction effects, respectively.

We conducted visual and statistical analyses of the single-subject experimental design. In order to facilitate the visual analysis of the single-subject dataset, the time series of all outcomes were organized in three-week blocks with the exception of the block that included the Eastern break week, which was four weeks in duration (Week 10 through Week 13). Because content tests for a given lesson remained available through the Moodle platform for a three-week window, participants could complete varying number of tests during each three-week period. Therefore, content test performance was graphed as successive rather than weekly events. Engagement responses (reactions and comments) were aggregated in successive weekly periods. Participants were ranked by mean baseline performance and assigned to multiple-baseline triads. Each triad included two participants with different baseline lengths (e.g., six-week and nine-week baselines) and a participant with an extended baseline that did not receive the intervention. The range of baseline performance within each triad was below 10% (range, 0.1% to 7.8%).

In addition to the visual analysis, we analyzed the single-subject dataset by computing between-case standardized mean differences (BC-SMD) and 95% confidence intervals for all outcomes (tests, comments, reactions) and relevant two-term comparisons: baseline to intervention, baseline to intervention without semi-immediate feedback, intervention to intervention without semi-immediate feedback, and baseline to intervention (Hedges et al., 2012, 2013). Analyses were conducted with the SPSS macro DHPS (Marso & Shadish, 2015). BC-SMD effect sizes may be interpreted according to the usual guides for standardized mean differences (Cohen, 1988; Sawilowsky, 2009).

RESULTS

3.1 Content tests

The mixed-effect ANOVA revealed a main effect of Group, $F(1, 46) = 18.37, p < .001, \eta^2 = .34$, and Time, $F(2, 46) = 9.13, p < .001, \eta^2 = .21$, on content tests performance (Table 2). The Mauchly test was not significant ($p = .187$) and suggested that the sphericity assumption was met. The main effect of Time should be attributed to the intervention effect within the intervention group. Specifically, when the effect of Time was restricted to those in the control group, the main effect of Time was not statistically significant, $F(2, 18) = 2.78, p = .81, \eta^2 = .14$. In fact, there was a significant Time X Group interaction for content test performance, $F(2, 37) = 14.65, p < .001, \eta^2 = .30$. A pairwise comparison restricted to the intervention group using a Bonferroni adjustment revealed a statistically significant difference for the mean difference between baseline and treatment ($p < .001$) and the mean difference between baseline and intervention without semi-immediate feedback ($p = .002$). The mean difference between the two sequential forms of intervention (BE-Social with and without semi-immediate feedback) was not significant ($p = 1$), suggesting that intervention gains remained after withdrawing semi-immediate feedback.

The full multiple-baseline time series is presented in Figure 2 (participants 1 through 18), Figure 3 (participants 19 through 36), and Figure 4 (participants 37 through 46). The visual analysis reveals that the effect of the intervention was largely idiosyncratic. Specifically, 15 participants showed an immediate moderate-to-large effect upon the start of the intervention with moderate to minimal overlap (P1, P2, P4, P10, P13, P16, P22, P23, P25, P26, P31, P34, P35, P37, P40). By contrast, 12 participants showed no discernable effect through the visual analysis, with considerable overlap across the baseline and intervention time series (P5, P7, P8, P11, P14,

P17, P19, P20, P28, P29, P38, P42). Most participants showing an intervention effect maintained their performance after withdrawing semi-immediate feedback. However, for P13 and P37 the effect of the intervention quickly disappeared upon the withdrawal of semi-immediate feedback. Only one participant (P43) exposed to an extended baseline (control group) showed a consistent change in performance coinciding with intervention transition dates. In addition, three participants undergoing the extended baseline (P3, P42, and P46) showed transient increases in performance. Moreover, none of the participants receiving the intervention showed a detrimental effect in performance during the treatment and post-treatment phases. A summary of the visual analysis is available in Table 3. The single-subject analysis seemed critical to ascertain the idiosyncratic nature of the effect.

The single-subject effect size analysis confirmed the findings of the group analysis (Table 4). Specifically, large effect sizes were found in the baseline-to-intervention comparisons. As shown in the mixed-effects ANOVA, the intervention-to-intervention without semi-immediate feedback comparison revealed that the intervention gains persisted largely unaffected after the intervention had been withdrawn.

3.2 Facebook group engagement

The Mauchly test for the repeated-measures ANOVA for both Facebook group comments and reactions was significant ($p < .001$), suggesting that the sphericity assumption was not met. Therefore, we used the Greenhouse-Geisser correction for degrees of freedom and p values. The mixed-effect ANOVA of the randomized controlled trial revealed a main effect of Group on Facebook engagement responses including reactions, $F(1, 46) = 68.07$, $p = .07$, $\eta^2 = .6$, and comments, $F(1, 46) = 9.62$, $p < .05$, $\eta^2 = .18$ (Table 2). We also identified a main effect of Time for reactions, $F(1.14, 46) = 12.57$, $p < .01$, $\eta^2 = .23$, but only a statistical trend for comments, $F(1.49, 46) = 3.26$, $p = .06$, $\eta^2 = .07$. The main effect of Time on Facebook reactions may be attributed entirely to the intervention group; the control group produced no reactions or comments in the general Facebook group throughout the study. Moreover, there was a significant Time X Group interaction for reactions, $F(1.14, 46) = 12.57$, $p < .01$, $\eta^2 = .23$. A pairwise comparison using a Bonferroni adjustment revealed a statistically significant difference for mean comments between baseline and intervention ($p < .05$), but not between baseline and intervention after semi-immediate feedback was withdrawn ($p = .08$). The mean difference in comments across intervention periods (with and without semi-immediate feedback) was not significant ($p = 1$). The analysis

suggests that the increased number of comments during treatment were lost after the withdrawal of semi-immediate feedback. The pairwise comparison with Bonferroni adjustments for reactions revealed a statistically significant difference for the mean difference between baseline and intervention ($p < .001$) and between baseline and intervention after the withdrawal of semi-immediate feedback ($p < .001$). The mean reactions difference between intervention periods (with and without semi-immediate feedback) was not significant ($p = .51$), suggesting that the effect of the intervention remained unaltered after the withdrawal of semi-immediate feedback.

Figures 2, 3 and 4 present the complete multiple-baseline analysis for all participants. Visual analysis reveals that the effect of engagement, as measured by reactions and comments, was largely idiosyncratic. With the exception of P19, all participants showed a rapid increase in reactions upon the intervention starting (P19 showed a delayed increase in reactions). By contrast, only 16 participants showed a rapid increase in comments (P2, P4, P10, P13, P20, P22, P23, P25, P26, P29, P31, P32, P34, P38, P40, P41) and one additional participant (P16) demonstrated a delayed increase in comments. Overall, the increase in comments was more modest than that observed for reactions. None of the participants showed clear evidence of the increase in engagement being transient or dependent on semi-immediate feedback. Finally, a few individuals showed an apparent correlation between content tests performance and engagement (particularly reactions) during the intervention phase (P4, P7, P11, P22, P23, P25, P28, P31, P34, P37).

The single-subject effect size analysis was consistent with the findings of the group analysis both suggesting relatively larger effect sizes for reactions than comments (Table 4). In addition, baseline-to-intervention effect sizes were larger before the withdrawal of semi-immediate feedback for reactions, but not for comments. Interestingly, the single-subject effect size analysis produced identical baseline-to-intervention effect sizes for comments before and after the withdrawal of semi-immediate feedback, which is a slight departure from the results of the mixed-effects ANOVA. This difference may be attributed to the greater sensitivity to within-phase trends of the Hedges g effect size.

3.3 Social validity

Fourteen participants (50%) of the BE-Social group provided unsolicited positive feedback (P2, P10, P11, P13, P14, P19, P20, P25, P31, P32, P38, P40, P41). We did not receive any neutral or negative feedback from the BE-Social group. No unsolicited feedback was received from

participants in the control group. The top five thematic words or topics in the unsolicited feedback messages according to a sentence-by-sentence theme analysis were instructor (9 entries, 28.12% of repeated entries), support (4, 12.5%), videos (3, 9.38%), knowledge (3, 9.38%), and time (3, 9.38%). The context for these themes included gratitude towards the instructor for sharing knowledge, devoting time to students, and supporting them. Comments also mentioned the utility of the video posts.

DISCUSSION

The current study provides evidence supporting the feasibility and impact on academic performance of an evidence-informed intervention package delivered through a closed Facebook group. The BE-Social program was comprised of self-management strategies, cooperative learning, and semi-immediate feedback, and was intended as an adjunct to a one-year postgraduate course in applied psychology delivered through a Moodle platform. The overall results indicated a considerable increase in engagement indices and academic performance. Specifically, academic performance, as evaluated by frequent multiple-choice objective tests, demonstrated a 20-point increase in the intervention group over a 100-point scale. It is important to emphasize that students did not receive course credit for completing the tests, therefore there was no apparent motivation for engaging in academic dishonesty. Intervention gains persisted after semi-immediate feedback had been withdrawn. A single-subject analysis indicated that intervention gains were largely driven by a subgroup of 15 individuals responding favorably to the intervention. Interestingly, about half of these responders showed an apparent covariation between social media engagement and gain in academic performance over the course of the intervention.

4.1 Theoretical considerations

Digital social interaction through social media platforms, has already become more prevalent than face-to-face and analogical forms of social interaction. Social media platforms are designed to enhance interaction. We would expect that immediacy of communication and other response-inducing social media functionalities (e.g., push notifications, multi-device capability, labelling, engagement-dependent badges, multi-media content) were key to the effects of the intervention (Changsu et al., 2016; Méndez et al., 2014). The behavioral procedures composing the multi-component intervention were amenable to the social media channel and their effects on academic behavior might have been enhanced by social media functionalities. For example,

immediacy has been established as one of the key elements of feedback effectiveness in online education (see for example Jensen et al., 2021). Moreover, the ephemeral nature of social media feeds means that delayed feedback may be less likely to impact the audience (e.g., the recency of the relevant context may be lost).

There is ample evidence to suggest that social media engagement is, to a significant extent, a function of the social rewards that users receive for their active online behavior. For example, in a series of quantitative analyses, Lindström et al. (2021) showed that the latency of posting behavior (a dimension of behavior strength) was a function of the number of likes received in preceding posts. In addition, users tend to provide more feedback to others after they themselves have received positive feedback (Eckles et al., 2016). It has also been shown that social comparison affects social media feedback in similar ways as it affects non-social rewards (e.g., Rosenthal-von der Pütten et al., 2019).

These findings, together with the delay discounting effect, described in both basic and applied research. This phenomenon can be defined as the decrease in response rate caused by the delay of the reinforcer (Reynolds, 2006). This fact seems to explain the results of the study, where the delay of the instructor's feedback was the factor that seemed to have the greatest impact on students' academic performance. In this study, we have described the feedback delay of up to 24 hours with the term "semi-immediate" adapting to the applied characteristics of the educational program. It is likely that if the delay time were reduced, there would be higher academic performance in students. This apparent effect of feedback immediacy fits well with an online behavioral reward learning model.

Our results are consistent with the view that instructor- and peer-mediated feedback are forms of social reward that activate student engagement. Course content engagement as part of a Facebook group provides a social milieu that, in addition to replacing (totally or partially) the traditional social context of learning, may also enhance interaction opportunities (i.e., ready access to a community, regularly posted content) and timely social reinforcement (peer- and instructor-mediated feedback). Over time, these interactive interaction processes result in improved content exposure, content repetition, and content elaboration, which are known mediators of information acquisition and, ultimately, academic performance (e.g., Zheng et al., 2015). This mechanism may be captured by a *digital environment reward optimization hypothesis*. Specifically, the current evaluation of the BE-Social package suggests that increased opportunities

of reward learning brought about both by evidence-based teaching-learning strategies (e.g., video-modeling, regular content posting) and social media functionalities (e.g., semi-immediate feedback, low response effort, content sharing) may maximize social reinforcement for engagement responses, some of which are important elements of effective study behavior (e.g., content elaboration, problem-solving).

Key elements of this conceptual framework were not directly assessed here. Future studies could evaluate the impact of discrete elements of the BE-Social program on engagement frequency and latency as a function of various social reward parameters (e.g., likes received, contingent commenting and praising). The study by Lindström et al. (2021) already provides an experimental paradigm in this direction.

4.2 Methodological considerations

The current study may be considered translational in that methods borrowed from the mainstream behavioral education literature have been delivered and evaluated through a social media outlet. For example, video modelling has been used predominantly to teach social skills (Sherer et al., 2001), daily living skills (Kellems, 2016), vocational skills (Kellems et al., 2012), and academic skills (Cihak, 2009), among others, to people with developmental disability (see a systematic review in Park et al., 2019). Applications of video modeling and video self-modeling with adult population have been rare (see Moughal, 2015 for an exception), and have not focused on tertiary education. By contrast, the present study focused on college-level academic skills with typically developed individuals.

In addition, while elements of a cooperative learning environment such as instructor- and peer-mediated feedback have been evaluated at length in college-level education (e.g., Martin et al., 2002), our study presents a novel channel for delivering feedback with evidence-informed formal requirements (e.g., semi-immediate feedback, feedback specific to a target academic behavior).

Randomized controlled trials (RCTs) often mask individual or subgroup effects (Frieden, 2017). The current study featured a combination of RCT and single-subject research methodologies, a key strategy to ascertain idiosyncratic effects (Smith, 2012). While RCTs are ideal for outcome research, they may be less practical when novel multi-component interventions are being evaluated and fine-tuned. Single-subject experimental designs are increasingly used in

education to identify delayed effects, transient effects, and specific groups of non-responders (Ledford & Gast, 2018). In the current analysis only about 60% of participants in the intervention group demonstrated an apparent gain in academic performance. This suggests that the intervention may be optimized further by adding new intervention components, adjusting the intensity of the existing ones, or evaluating the impact of the various intervention elements in the BE-Social package separately (see a review of component analysis methods for behavioral interventions in Ward-Horner & Sturmey, 2010).

The current study also features an assortment of relevant methodological strategies appropriate for using a social media platform for educational purposes in naturalistic settings. Specifically, the proposed direct measurement, interobserver agreement, and procedural integrity strategies provide the basis for evaluating educational interventions in this context. Future studies would benefit from purposely-developed applications intended for real-time monitoring of key outcomes of the teaching-learning process as it unfolds in social media groups attached to online and blended learning programs. Future studies could incorporate intervention outcomes such as feedback latency, length of feedback exchanges, peer-to-instructor feedback ratio, automated text analysis, among others. Because we could only compute feedback latency retrospectively, this potentially informative outcome was restricted to monitoring procedural integrity in the current study. The computational capability of extracting these and other outcomes may allow for their use by instructors as part of the teaching-learning process.

4.3 Limitations and future directions

Various limitations to the current study should be noted. First, the amount of exposure to academic content was difficult to equate across the arms of the RCT. Specifically, the intervention procedures (e.g., additional opportunities for peer-to-peer interaction, presentation of self-management strategies) required the perfunctory addition of course-related content resulting in additional exposure to course material among those in the intervention group. More research is needed to establish minimally disruptive strategies for equating the intensity of educational interventions delivered through social media (e.g., yoking the number of social media posts across groups).

Second, we did not explore experimentally the potentially mediating role of social media engagement in academic achievement. Various engagement responses may offer different infor-

mation. Reactions provide evidence of sentiment and content reception, whereas comments may inform content elaboration and motivation, and may be subject to more complex analyses (e.g., latency of responding, text analysis). Moreover, the different level of effort involved in reacting and commenting may explain the disparity in frequencies observed in these two indices (see Billington & DiTommaso, 2003 for a discussion on the impact of response effort on choice).

Third, more robust measures of academic performance may be achieved by adding proctored testing or proctored final examinations as part of the intervention outcomes. While unsupervised tests may cause a positive bias (Steger et al., 2020). The assessment system used here was identical across groups and test items focused on applied scenarios, which cannot be solved by copying course contents or searching the Internet.

Forth, a gender gap is often observed in psychology courses (American Psychological Association, 2017) and it was certainly the case here. The unbalanced gender distribution in our sample of participants means that our findings ought to be replicated with diverse populations.

Fifth, future studies could determine whether the proposed intervention could also be implemented “autonomously” by trained students using pyramidal or peer-proctor training structures (see, for example, Erath et al., 2020; Svenningsen & Pear, 2011), which would help to optimize the cost-effectiveness of the intervention even further.

Finally, the simple randomization strategy meant that the sample sizes across the control and intervention groups was likely to be unbalanced. While simple randomization often leads to unbalanced groups, as it was the case here, it remains a randomization strategy minimally prone to bias (Schulz & Grimes, 2002).

4.4 Practical implications

The study showed how a time-efficient intervention could result in a mean 20-point gain in academic performance. A cursory cost analysis indicates that 60 instructor person-hours were required to deliver the intervention to 28 students (person-hours per student = 2.1 h). Further gains in efficiency may be possible by evaluating specific effects of the various intervention components, focusing the intervention on low-performance students, optimizing the ratio between

the feedback provided by peers and that provided by the instructor, and evaluating the scalability of the BE-Social program with larger groups.

Future studies may help to further improve the effectiveness of the BE-Social program. However, the proposed intervention has significant knowledge transfer potential as is. Policy makers, administrators, and educators can benefit from the low cost of integrating social media platforms in higher education courses. In addition, the added teaching repertoire may be easily acquired by instructors (e.g., self-management, video modeling, behavioral skills training, timely specific feedback) after adequate training. Our findings suggest that the intervention gains may be observed within days of its inception, which is consistent with the duration of semester-long or shorter courses. In addition, positive feedback from students (analyzed qualitatively in the social validity section) suggests that the intervention may also favor motivation, social climate, and positive peer-peer and peer-instructor interactions.

While more evidence is needed, it is not too soon to recommend the use of social media platforms as an educational aid in college-level courses incorporating instructors purposefully-trained to provide frequent interaction opportunities and timely and informative feedback (including praise for positive engagement).

CONCLUSIONS

With the near-universal growth of online education and social media use, social media adjuncts to online and blended learning programs are expected to increase over the foreseeable future. Unfortunately, this field has not yet emerged as an evidence-based practice. The choice of social media functionalities by educators and the design of the social dynamics delivered through this channel may have a large impact on student engagement and performance (Chen et al., 2020). The current study shows that a closed Facebook group may be an effective channel for delivering a multi-component behavioral intervention including cooperative learning, self-management, and video modeling intervention elements with adult students attending an online postgraduate course. Our findings are consistent with the view that social media platforms provide a prosthetic social milieu with the potential to enhance the social environment of traditional education by maximizing social rewards through increased interaction opportunities and timely positive feedbacks. This process is captured by the *digital environment reward optimization hypothesis*.

REFERENCES

- Aldi, C., Crigler, A., Kates-McElrath, K., Long, B., Smith, H., Rehak, K., & Wilkinson, L. (2016). Examining the effects of video modeling and prompts to teach activities of daily living skills. *Behavior Analysis in Practice*, 9(4), 384–388. <https://doi.org/10.1007/s40617-016-0127-y>
- Al-Rahmi, W. M., Othman, M. S., & Yusuf, L. M. (2015). Social media for collaborative learning and engagement: Adoption framework in higher education institutions in Malaysia. *Mediterranean Journal of Social Sciences*, 6(3 S1), 246. <https://doi.org/10.5901/mjss.2015.v6n3s1p246>
- Alshuaibi, M. S. I., Alshuaibi, A. S. I., Shamsudin, F. M., & Arshad, D. A. (2018). Use of social media, student engagement, and academic performance of business students in Malaysia. *International Journal of Educational Management*. Vol. 32 No. 4, pp. 625-640. <https://doi.org/10.1108/IJEM-08-2016-0182>
- American Psychological Association (2017). *The changing gender composition of psychology: Update and expansion of the 1995 task force report*. <http://www.apa.org/pi/women/programs/gender-composition/task-force-report.pdf>
- Arteaga-Sánchez, R., Cortijo, V., & Javed, U. (2019). Factores que impulsan la adopción de Facebook en la educación superior. *Elearning and Digital Media*, 16(6), 455-474. <https://doi.org/10.1177/2042753019863832>
- Aslan, A. (2021). Problem-based learning in live online classes: Learning achievement, problem-solving skill, communication skill, and interaction. *Computers & Education*, 171, 104237. <https://doi.org/10.1016/j.compedu.2021.104237>
- Ataie, F., Shah, A., Nasir, M., & Nazir, M. (2015). Collaborative learning, using Facebook's page and Groups. *Internal Journal of Computer System*, 02(02), 47–52.
- Aydin, S. (2012). A review of research on Facebook as an educational environment. *Educational Technology Research and Development*, 60(6), 1093–1106. <https://doi.org/10.1007/s11423-012-9260-7>
- Bailey, J. S., & Burch, M. R. (2018). *Research Methods in Applied Behavior Analysis* (2nd ed.). Routledge.
- Bellini, S., & Akullian, J. (2007). A meta-analysis of video modeling and video self-modeling interventions for children and adolescents with autism spectrum disorders. *Exceptional Children*, 73(3), 264–287. <https://doi.org/10.1177/001440290707300301>

- Billington, E., & DiTommaso, N. M. (2003). Demonstrations and applications of the matching law in education. *Journal of Behavioral Education*, 12(2), 91–104. <https://doi.org/10.1023/A:1023881502494>
- Briesch, A. M., & Chafouleas, S. M. (2009). Review and analysis of literature on self-management interventions to promote appropriate classroom behaviors (1988-2008). *School Psychology Quarterly*, 24(2), 106–118. <https://doi.org/10.1037/a0016159>
- Changsu, K., Minghui, K. & Tao, W. (2016). Influence of knowledge transfer on SNS community cohesiveness. *Online Information Review*, 40(7), 959-978. <https://doi.org/10.1108/OIR-08-2015-0258>
- Chen, T., Peng, L., Yin, X., Rong, J., Yang, J., & Cong, G. (2020). Analysis of user satisfaction with online education platforms in china during the COVID-19 pandemic. *Healthcare (Basel, Switzerland)*, 8(3), 200. <https://doi.org/10.3390/healthcare8030200>
- Cheung, C. M. K., Chiu, P. Y., & Lee, M. K. O. (2011). Online social networks: Why do students use Facebook? *Computers in Human Behavior*, 27(4), 1337–1343. <https://doi.org/10.1016/j.chb.2010.07.028>
- Cihak, D., Fahrenkrog, C., Ayres, K. M., & Smith, C. (2009). The use of video modeling via a video iPod and a system of least prompts to improve transitional behaviors for students with autism spectrum disorders in the general education classroom. *Journal of Positive Behavior Interventions*, 12(2), 103–115. <https://doi.org/10.1177/1098300709332346>
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed). Lawrence Erlbaum Associates
- Cooper, J. O., Heron, T. E., y Heward, W. L. (2019). *Applied Behavior Analysis*, 3rd ed. Pearson
- Credé, M., & Kuncel, N. (2008). Study habits, skills, and attitudes: The third pillar supporting collegiate academic performance. *Perspectives on Psychological Science*, 3(6), 425-453. <http://dx.doi.org/10.1111/j.1745-6924.2008.00089.x>
- Doleck, T., & Lajoie, S. (2018). Social networking and academic performance: A review. *Education and Information Technologies*, 23(1), 435–465. [https://doi.org/10.1007/s10639-017-9612-](https://doi.org/10.1007/s10639-017-9612-3)

- Eckles, D., Kizilcec, R. F., & Bakshy, E. (2016). Estimating peer effects in networks with peer encouragement designs. *Proceedings of the National Academy of Sciences*, 113, 7316–7322. <https://doi.org/10.1073/pnas.1511201113>
- Erath, T. G., DiGennaro Reed, F. D., Sundermeyer, H. W., Brand, D., Novak, M. D., Harbison, M. J., & Shears, R. (2020). Enhancing the training integrity of human service staff using pyramidal behavioral skills training. *Journal of Applied Behavior Analysis*, 53(1), 449–464. <https://doi.org/10.1002/jaba.608>
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2009). G*Power (Version 3.1.9.4). [Computer Software]. <https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower>
- Frieden T. R. (2017). Evidence for Health Decision Making - Beyond Randomized, Controlled Trials. *The New England Journal of Medicine*, 377(5), 465–475. <https://doi.org/10.1056/NEJMra1614394>
- Garcia, E., Elbeltagi, I. M., Dungay, K., & Hardaker, G. (2015). Student use of Facebook for informal learning and peer support. *International Journal of Information and Learning Technology*, 32(5), 286–299. <https://doi.org/10.1108/IJILT-09-2015-0024>
- Ledford, J. R., & Gast, D. L. (2018). *Single Case Research Methodology: Applications in Special Education and Behavioral Sciences* (3rd ed.). Routledge.
- Lindström, B., Bellander, M., Schultner, D. T., Chang, A., Tobler, P. N., & Amodio, D. M. (2021). A computational reward learning account of social media engagement. *Nature Communications*, 12, 1311. <https://doi.org/10.1038/s41467-020-19607-x>
- Hedges, L. V., Pustejovsky, J. E., & Shadish, W. R. (2012). A standardized mean difference effect size for single case designs. *Research Synthesis Methods*, 3, 224–239. <https://doi.org/10.1002/jrsm.1052>
- Hedges, L. V., Pustejovsky, J. E., & Shadish, W. R. (2013). A standardized mean difference effect size for multiple baseline designs across individuals. *Research Synthesis Methods*, 4, 324–341. <https://doi.org/10.1002/jrsm.1086>
- Hicks, A., & Graber, A. (2010). Shifting paradigms: Teaching, learning and Web 2.0. *Reference Services Review*, 38(4), 621–633. <https://doi.org/10.1108/00907321011090764>

- Hsu, T. C. (2018). Behavioural sequential analysis of using an instant response application to enhance peer interactions in a flipped classroom. *Interactive Learning Environments*, 26(1), 91–105. <https://doi.org/10.1080/10494820.2017.1283332>
- Hunerberg, E. (2019). 3rd Eye Theme Analysis [Online Application]. <https://3rdeyeinformation.com/home>
- Hurt, N. E., Moss, G., Bradley, C., Larson, L., Lovelace, M., Prevost, L., Riley, N., Domizi, D., & Camus, M. (2012). The 'Facebook' Effect: College Students' Perceptions of Online Discussions in the Age of Social Networking.' *International Journal for the Scholarship of Teaching and Learning*, 6(2). <https://doi.org/10.20429/ijstl.2012.060210>
- IBM Corp. (2019). *IBM SPSS Statistics for Mac (Version 26.0)* [Computer Software]. <https://www.ibm.com/products/spss-statistics>
- Jensen, L. X., Bearman, M., & Boud, D. (2021). Understanding feedback in online learning—A critical review and metaphor analysis. *Computers and Education*, 173, 104271. <https://doi.org/10.1016/j.compedu.2021.104271>
- Junco, R. (2012). The relationship between frequency of Facebook use, participation in Facebook activities, and student engagement. *Computers and Education*, 58(1), 162–171. <https://doi.org/10.1016/j.compedu.2011.08.004>
- Kalelioğlu, F. (2017). Using facebook as a learning management system: Experiences of pre-service teachers. *Informatics in Education*, 16(1), 83–101. <https://doi.org/10.15388/infe-du.2017.05>
- Kellems, R. O., & Morningstar, M. E. (2012). Using video modeling delivered through iPads to teach vocational tasks to young adults with autism spectrum disorders. *Career Development and Transition for Exceptional Individuals*, 35(3), 155-167. <https://doi.org/10.1177/0885728812443082>
- Kellems, R. O., Mourra, K., Morgan, R. L., Riesen, T., Glasgow, M., & Huddleston, R. (2016). Video Modeling and Prompting in Practice. *Career Development and Transition for Exceptional Individuals*, 39(3), 185–190. <https://doi.org/10.1177/2165143416651718>
- Kitsantas, A., Robert, A. R., & Doster, J. (2004). Developing self-regulated learners: Goal setting, self-evaluation, and organizational signals during acquisition of procedural skills. *The Journal of Experimental Education*, 72(4), 269–287. <https://doi.org/10.3200/JEXE.72.4.269-287>

- Little, S. G., Akin-Little, A., & O'Neill, K. (2014). Group contingency interventions with children—1980-2010. *Behavior Modification*, 39(2), 322–341. <https://doi.org/10.1177/0145445514554393>
- Madge, C., Meek, J., Wellens, J., & Hooley, T. (2009). Facebook, social integration and informal learning at university: “It is more for socialising and talking to friends about work than for actually doing work.” *Learning, Media and Technology*, 34(2), 141–155. <https://doi.org/10.1080/17439880902923606>
- Mahdiun, R., Salimi, G., & Raeisy, L. (2020). Effect of social media on academic engagement and performance: Perspective of graduate students. *Education & Information Technologies*, 25(4), 2427-2446. <https://doi.org/10.1007/s10639-019-10032-2>
- Malott, R. W. (2012). *I'll stop procrastinating when I get around to it*. Behaviordelia.
- Marso, D., & Shadish, W. R. (2015). *Software for Meta-analysis of Single-Case Design DHPS* (Version March 7, 2015) [Computer Software]. <https://faculty.ucmerced.edu/wshadish/software/software-meta-analysis-single-case-design/dhps-version-march-7-2015>
- Martin, T. L., Pear, J. J., & Martin, G. L. (2002). Feedback and its effectiveness in a computer-aided personalized system of instruction course. *Journal of Applied Behavior Analysis*, 35(4), 427–430. <https://doi.org/10.1901/jaba.2002.35-427>
- Mazman, S. G., & Usluel, Y. K. (2010). Modeling educational usage of Facebook. *Computers and Education*, 55(2), 444–453. <https://doi.org/10.1016/j.compedu.2010.02.008>
- Mendez, J. P., Le, K., & De la Cruz, J. (2014). Integrating Facebook in the classroom: pedagogical dilemmas. *Journal of Instructional Pedagogies*, 13.
- Michikyan, M., Subrahmanyam, K., & Dennis, J. (2015). Facebook use and academic performance among college students: A mixed-methods study with a multi-ethnic sample. *Computers in Human Behavior*, 45, 265–272. <https://doi.org/10.1016/j.chb.2014.12.033>
- Mitchem, K. J., & Richards, A. (2003). Students with learning disabilities. *Advances in Special Education*, 15(4), 99–117. <https://doi.org/10.1177/0143034393141003>
- Moughal, S. (2015). *Increasing Social Connections for Young Migrant Women in the New Zealand Community Using Video Self-Modeling*. Unpublished master's thesis. University of Auckland.

- Park, J., Bouck, E., & Duenas, A. (2019). The effect of video modeling and video prompting interventions on individuals with intellectual disability. *Journal of Special Education Technology*, 34(1), 3-16. <https://doi.org/10.1177/0162643418780464>
- Parsons, M. B., Rollyson, J. H., & Reid, D. H. (2013). Teaching practitioners to conduct behavioral skills training: A pyramidal approach for training multiple human service staff. *Behavior Analysis in Practice*, 6, 4-16. <https://doi.org/10.1007/BF03391798>
- Pempek, T. A., Yermolayeva, Y. A., & Calvert, S. L. (2009). College students' social networking experiences on Facebook. *Journal of Applied Developmental Psychology*, 30(3), 227–238. <https://doi.org/10.1016/j.appdev.2008.12.010>
- Reynolds, B. (2006). A review of delay-discounting research with humans: relations to drug use and gambling. *Behavioural Pharmacology*, 17(8), 651-667 <https://doi.org/10.1097/FBP.0b013e3280115f99>
- Rosenthal-von der Pütten, A. M., Hastall, M., Köcher, S., Meske, C., Heinrich, T., Labrenz, F., & Ocklenburg, S. (2019). “Likes” as social rewards: Their role in online social comparison and decisions to like other people’s selfies. *Computers in Human Behavior*, 92, 76–86.
- Sawilowsky, S. (2009). New effect size rules of thumb. *Journal of Modern Applied Statistical Methods*, 8, 467–474. <https://doi.org/10.22237/jmasm/1257035100>
- Schoenherr, K. (2018). Using video modeling to teach vocational skills to young adults with autism spectrum disorder. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, 79(10-B(E)). <http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc14&NEWS=N&AN=2018-34221-022>
- Schulz, K. F., & Grimes, D. A. (2002). Unequal group sizes in randomised trials: guarding against guessing. *Lancet*, 359(9310), 966–970. [https://doi.org/10.1016/S0140-6736\(02\)08029-7](https://doi.org/10.1016/S0140-6736(02)08029-7)
- Schworm, S., & Gruber, H. (2012). E-Learning in universities: Supporting help-seeking processes by instructional prompts. *British Journal of Educational Technology*, 43(2), 272–281. <https://doi.org/10.1111/j.1467-8535.2011.01176.x>
- Selwyn, N. (2009). Faceworking: Exploring students' education-related use of Facebook. *Learning, Media and Technology*, 34(2), 157–174. <https://doi.org/10.1080/17439880902923622>
- Sherer, M., Pierce, K. L., Paredes, S., Kisacky, K. L., Ingersoll, B., & Schreibman, L. (2001). Enhancing conversation skills in children with autism via video technology. Which is bet-

ter, “self” or “other” as a model? *Behavior Modification*, 25(1), 140–158. <https://doi.org/10.1177/0145445501251008>

Shukla-Mehta, S., Miller, T., & Callahan, K. J. (2010). Evaluating the effectiveness of video instruction on social and communication skills training for children with autism spectrum disorders: A review of the literature. *Focus on Autism and Other Developmental Disabilities*, 25(1), 23–36. <https://doi.org/10.1177/1088357609352901>

Smith J. D. (2012). Single-case experimental designs: a systematic review of published research and current standards. *Psychological Methods*, 17(4), 510–550. <https://doi.org/10.1037/a0029312>

Statista. (2021, April). Most popular social networks worldwide as of April 2021, ranked by number of active users. <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>

Steger, D., Schroeders, U., & Gnambs, T. (2020). A meta-analysis of test scores in proctored and unproctored ability assessments. *European Journal of Psychological Assessment*, 36(1), 174–184. <https://doi.org/10.1027/1015-5759/a000494>

Susanto, H., Yie, L. F., Mohiddin, F., Setiawan, A. A. R., Haghi, P. K., & Setiana, D. (2021). Revealing social media phenomenon in time of COVID-19 pandemic for boosting start-up businesses through digital ecosystem. *Applied System Innovation*, 4(1), 1–21. <https://doi.org/10.3390/asi4010006>

Svenningsen, L., & Pear, J. J. (2011). Effects of computer-aided personalized system of instruction in developing knowledge and critical thinking in blended learning courses. *The Behavior Analyst Today*, 12(1), 34–40. <http://dx.doi.org/10.1037/h0100709>

Tang, Y., & Hew, K. F. (2017). Is mobile instant messaging (MIM) useful in education? Examining its technological, pedagogical, and social affordances. *Educational Research Review*, 21, 85–104. <https://doi.org/10.1016/j.edurev.2017.05.001>

UNESCO (2020). *Impact of COVID-19 on education*. <https://en.unesco.org/covid19/education-response>

van Ryzin, M. J., Roseth, C. J., & Biglan, A. (2020). Mediators of effects of cooperative learning on prosocial behavior in middle school. *International Journal of Applied Positive Psychology*, 5(1–2), 37–52. <https://doi.org/10.1007/s41042-020-00026-8>

- Virues-Ortega, J., Moeyaert, M., Sivaraman, M., & Fernandez Castilla, B. Quantifying outcomes in applied behavior analysis through visual and statistical analyses: A synthesis. In J. L. Matson, *Applied Behavior Analysis: A Comprehensive Handbook*. SpringerNature. In press.
- Wang, Q., Woo, H. L., Quek, C. L., Yang, Y., & Liu, M. (2011). Using the Facebook group as a learning management system: An exploratory study. *British Journal of Educational Technology*, 43(3), 428–438. <https://doi.org/10.1111/j.1467-8535.2011.01195.x>
- Ward-Horner, J., & Sturmey, P. (2010). Component analyses using single-subject experimental designs: a review. *Journal of Applied Behavior Analysis*, 43(4), 685–704. <https://doi.org/10.1901/jaba.2010.43-685>
- Yu, F. Y., & Chen, C. Y. (2021). Student-versus teacher-generated explanations for answers to online multiple-choice questions: What are the differences? *Computers and Education*, 173, 104273. <https://doi.org/10.1016/j.compedu.2021.104273>
- Yuan, C. H., & Wu, Y. J. (2020). Mobile instant messaging or face-to-face? Group interactions in cooperative simulations. *Computers in Human Behavior*, 113, 106508. <https://doi.org/10.1016/j.chb.2020.106508>
- Zheng, L., Huang, R., Hwang, G.-J., & Yang, K. (2015). Measuring knowledge elaboration based on a computer-assisted knowledge map analytical approach to collaborative Learning. *Educational Technology & Society*, 18 (1), 321–336.

Table 10.1 Sociodemographic Characteristics of Study Participants

	BE-Social Group (n = 28)	Control Group (n = 18)	Overall (n = 46)
Age, <i>M</i> ± <i>SD</i>	41.86 (8.11)	40.67 (8.99)	41.39 (8.39)
Gender, % (<i>n</i>)			
Female	96.42 (27)	88.9(16)	93.48(43)
Male	3.70 (1)	11.11 (2)	6.52(3)
Country, % (<i>n</i>)			
United States	82.14 (23)	83.34 (15)	82.61 (38)
Other	17.86 (5)	16.67 (3)	17.39 (8)
Highest academic degree			
Bachelor's	85.71 (24)	77.78 (14)	82.61 (38)
Master's	14.28 (4)	22.22 (4)	17.39 (8)
Pre-entry reading score, <i>M</i> ± <i>SD</i>	5.39(3.10)	6.78(1.93)	5.93(2.76)
Pre-entry content score, <i>M</i> ± <i>SD</i>	4.46(2.90)	4.83(1.98)	4.61(2.56)

Notes. No significant difference identified in single-factor ANOVAs. Pre-entry reading and content scores were evaluated through ad-hoc multiple-choice tests completed before the start of the course. All pre-entry scores over a 10-point scale.

Table 10.2 Two-way Analysis of Variance during baseline (Time 1), intervention (Time 2), and Intervention without Semi-Immediate Feedback (Time 3)

	BE-Social		Control		Effect	Mixed-Effect ANOVA		η^2
	M	SD	M	SD		F ratio	df, n	
Content tests								
Time 1	38.55	10.51	39.23	10.79	G	18.366***	1, 46	.34
Time 2	61.56	16.49	38.58	15.94	T	9.130***	2, 46	.21
Time 3	58.36	18.13	33.30	13.51	T x G	14.648***	2, 37	.30
Comments								
Time 1	0.00	0.00	0.00	0.00	G	9.616*	1, 46	.18
Time 2	14.22	23.91	0.00	0.00	T	3.255	1.49, 46	.07
Time 3	21.78	39.66	0.00	0.00	T x G	3.255	1.49, 37	.07
Reactions								
Time 1	0.00	0.00	0.00	0.00	G	68.073***	1, 46	.61
Time 2	52.67	30.53	0.00	0.00	T	12.569**	1.14, 46	.23
Time 3	82.96	75.40	0.00	0.00	T x G	12.569**	1.14, 37	.23

Note. G = Group; T = Time.

* $p < .05$. ** $p < .01$. *** $p < .001$

Table 10.3 Visual Analysis Summary

Effect	Intervention group	Control group
	(n = 28)	(n = 18)
Null/negligible	12 (42.9), [P5, P7, P8, P11, P14, P17, P19, P20, P28, P29, P38, P42]	14 (77.8), [P6, P9, P12, P15, P18, P21, P21, P24, P27, P30, P33, P36, P39, P44, P45]
Moderate/large	15 (53.6), [P1, P2, P4, P10, P13, P16, P22, P23, P25, P26, P31, P34, P35, P37, P40]	1 (5.6), [P43]
Transient	2 (12.5), [P13, P37]	3 (16.7), [P3, P42, P46]

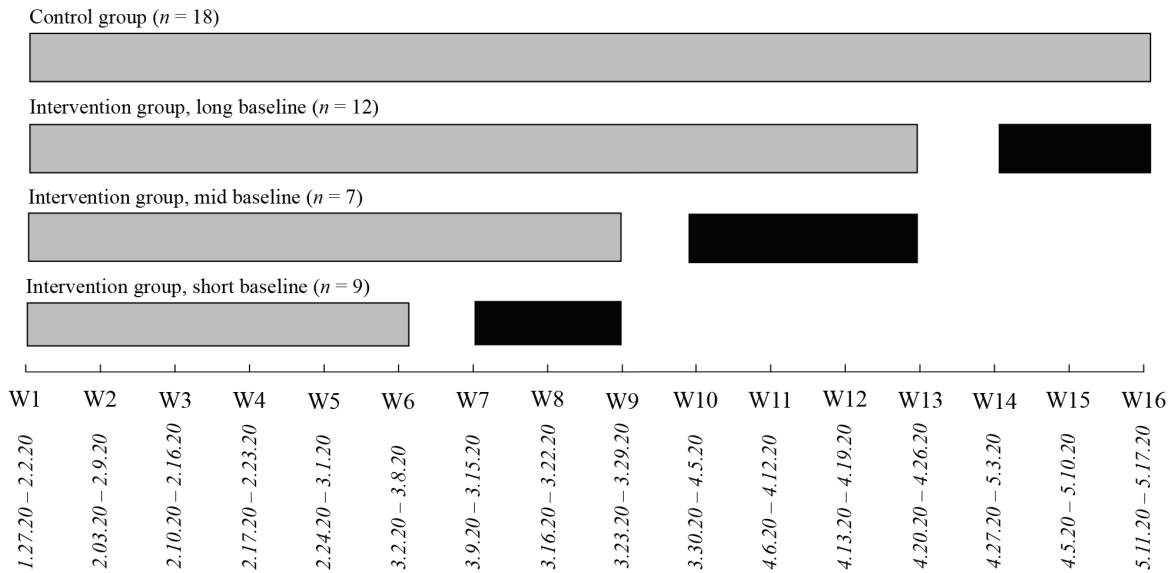
Notes. n (%), [Individual participant identifier]. Transient effects were not detectable among individuals in the intervention group with a long baseline (n = 12); percentage computed accordingly. No delayed effects identified.

Table 10.4 Multiple Baseline Design Effect Size Analysis of Participants in the Intervention Group (n = 28)

	Hedges g (95% confidence intervals)		
	Content Tests	Reactions	Comments
BL to INT W FB	1.04 (0.88, 1.20)	4.34 (3.85, 4.83)	0.81 (0.65, 0.97)
BL to INT W/O FB	0.89 (0.70, 1.08)	3.96 (3.66, 4.26)	0.81 (0.64, 0.98)
INT W FB to INT W/O FB	0.00 (-0.21, 0.21)	0.76 (0.51, 1.01)	0.11 (-0.13, 0.35)
BL to INT	0.93 (0.38, 1.48)	2.83 (2.58, 3.08)	0.66 (0.50, 0.82)

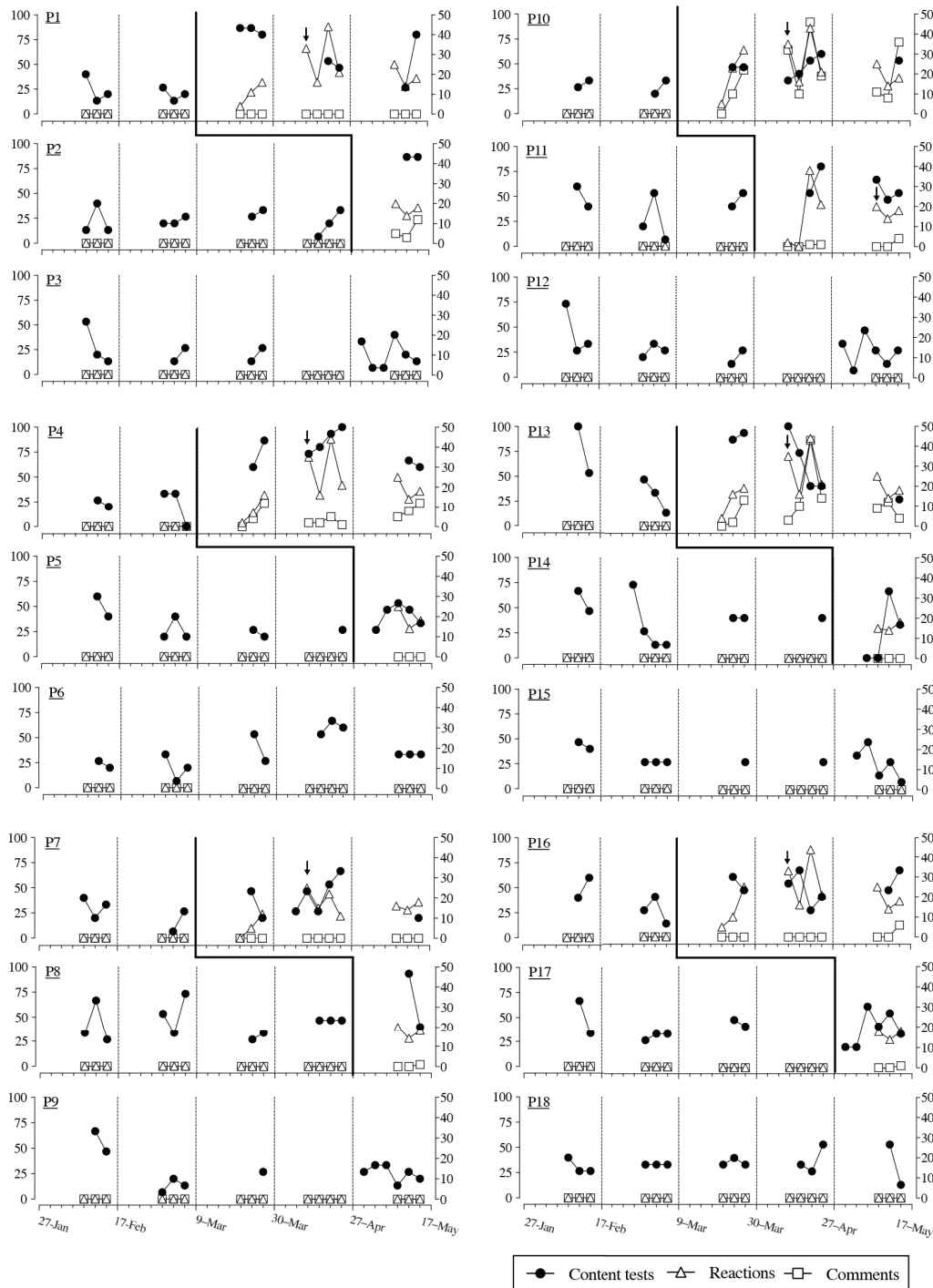
Notes. BL = Baseline; Int = Intervention; INT W FB = Intervention with semi-immediate feedback; INT /O FB = Intervention without semi-immediate feedback.

Figure 10.1 Temporal distribution of the baseline (light grey bars) and intervention (black bars) phases across the arms of the randomized controlled trial.



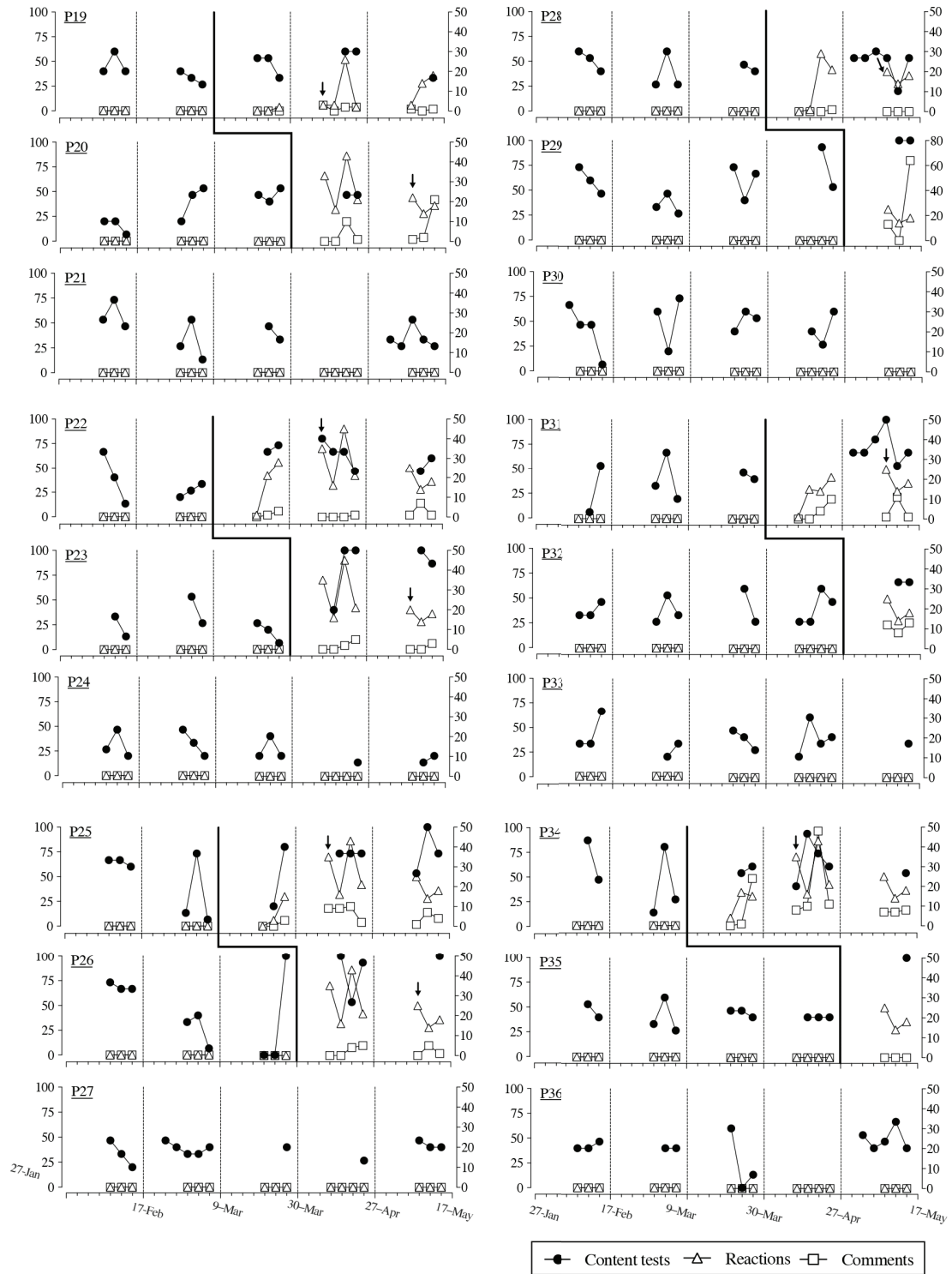
Notes. W = Weeks.

Figure 10.2 Single-Subject Analysis of Participants 1 through 18



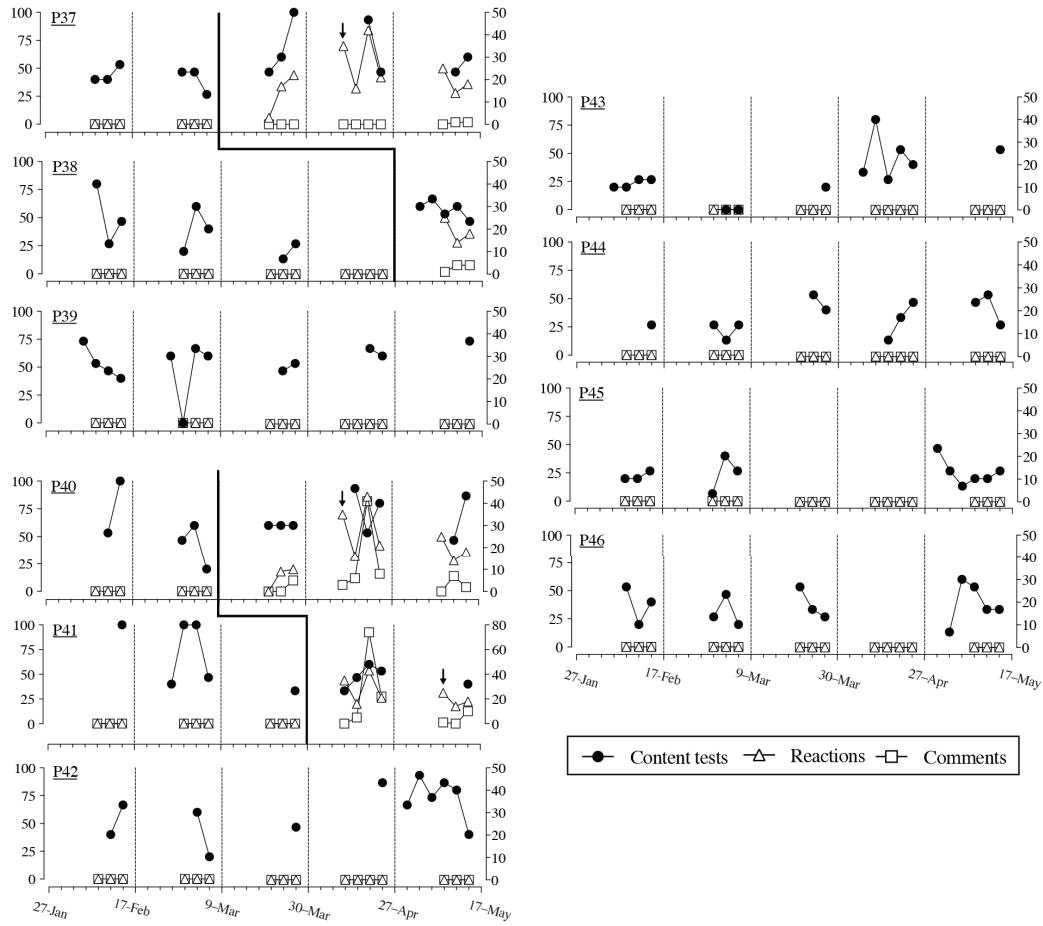
Note. Percentage of correct content tests responses and weekly frequency of engagement responses are scaled on the left and right y axes, respectively. Polygonal lines separate baseline and treatment phases. Broken vertical lines separate successive three-week periods (vacation breaks not counted). Arrows indicate the withdrawal of semi-immediate feedback.

Figure 10.3 Single-Subject Analysis of Participants 19 through 36



Note. See graphical conventions in Figure 2.

Figure 10.4 Single-Subject Analysis of Participants 37 through 46



Note. See graphical conventions in Figure 2.

CAPÍTULO 11. Enhancing Academic Performance in Tertiary Education through Social Media: A Multi-Arm RCT⁵⁴

ABSTRACT

Few randomized controlled trials (RCT) have analyzed evidence-based educational practices delivered through a social media environment. This study used a multi-arm RCT to evaluate the critical components of the Behavioral Education and Social Media (BE-Social) intervention package: self-management skills training delivered through video modeling, cooperative learning, and semi-immediate feedback. We evaluated social media engagement and academic performance among 141 students in a graduate-level applied psychology program. Students were randomly assigned to five groups: control ($n = 27$), self-management ($n = 27$), cooperative learning ($n = 33$), BE-Social without semi-immediate feedback ($n = 27$), and complete BE-Social intervention program ($n = 27$). Results indicated that participants receiving the complete BE-Social program showed the highest levels of engagement and academic performance. The analysis revealed that the semi-immediate feedback component was critical to the package effectiveness, whereas the impact of the self-management skills training intervention could not be verified. The conceptual, methodological, and practical implications of the study are discussed.

INTRODUCTION

Access to the adequate infrastructure can allow tertiary education institutions to continue their operation during challenging times (see for example Baytiyeh, 2018). For example, in the aftermath of a devastating earthquake, the University of Canterbury (New Zealand) quickly transitioned from face-to-face to online teaching preventing what would have been a major disruption (Ayebi-Arthur, 2017). Likewise, the COVID-19 pandemic has prompted students and teaching

54 El presente estudio ha sido enviado a la revista *British Journal of Educational Technologies* y se presenta en el idioma y el formato en el que se ha enviado.

Tarifa-Rodríguez, A., Virues-Ortega, J., Calero-Elvira, A. (2022). Enhancing academic performance in tertiary education through social media: A multiple-arm RCT. *British Journal of Educational Technologies*

professionals across the world to rely on e-learning solutions. The pandemic made apparent a digital breach across teaching institutions that had embraced online and blended models relative to more traditional formats (Beaunoyer et al., 2020). Numerous universities around the world now rely on online teaching as lockdowns and social distancing restrictions are imposed for extended periods of time with short notice. Adapting education to the online delivery channel have required major investments in infrastructure and human resources, particularly for the continuous education of teaching professionals unaccustomed to online education (Selvaraj et. al., 2021).

A key to successfully transitioning from traditional to online and blended education may require taking advantage of the tools that are unique to the digital channel, such as performance monitoring, semi-immediate feedback, and virtual group learning environments, to mention a few (Cheawjindakarn et. al., 2012). For example, an experimental program evaluation by Bartini (2008) showed that psychology students who received traditional instruction enhanced by a learning management system performed better academically relative to students in traditional education environments. Similar findings have been further verified in various systematic reviews and meta-analyses (McCutcheon et al., 2015; Means et al., 2013; Topping et al., 2022).

It is important to note that crucial facets of traditional university education may be undermined in an online-only environment. For example, students attending online courses enjoy fewer opportunities for social interaction, which could be detrimental to student motivation and morale (Meşe & Sevilen, 2021). In this connection, incorporating social media platforms to blended and online courses have the potential to supplement the otherwise limited social milieu of in-person education (Irwin et. al., 2012). Several studies indicate that incorporating Facebook groups to university courses may provide a strong means of communication and collaboration with the potential to impact the learning process (Arteaga et. al., 2014). For example, a study by Alshuabi et al. (2018) suggested that graduate students undertaking courses that included a social media platform showed greater cognitive engagement and academic performance. The authors suggested that cognitive engagement in the social media environment may be an important mediating factor leading to increased academic performance. Similarly, a survey study by Graham (2014) reported a positive correlation between increased participation in a parallel course social network and curricular engagement. The preliminary findings from these and other non-experimental studies ought to be verified with prospective experimental analyses.

1.1 Social Media Platforms

Facebook is one of the most popular social media channels among undergraduates and graduate students (Statista, 2022). In particular, closed Facebook groups offer a private space to interact and share information instantly. Namely, receiving private and public messages and comments through push notifications, quickly identifying online members, commenting and reacting to third-party posts, and receiving instructor moderation may help to activate student engagement and participation (e.g., Mansholt et al., 2021; Wang & Lin, 2021). Twitter has also been evaluated as a social network for academic purposes among graduate students (see for example Martinez-Cardama et al., 2019).

The facility for immediate feedback may be a key component of the potential impact of social media platforms on the teaching-learning process. Gopal et. al., (2021) and Dhawan (2020) have observed that providing feedback facilitates peer-to-peer as well as student-instructor interaction and increases student engagement with academic content. Instant messaging has become a natural means of interaction in both personal and professional life, leading to the notion that student-instructor interaction may also accommodate to this form of fast-paced communication (Giannikas, 2019). However, excessive interactions can cause stress or lead to problematic internet use, which can be deleterious to academic performance (Azizi et al., 2019) or produce resistance to using social media in education among both students and educators (Cloete & Villiers, 2009; Roblyer et al., 2010).

A balanced and safe used of social media aids in tertiary education demands the assessment of positive evidence-based educational practices that can be delivered through the social media channel. Indeed, a factor that may be obscuring the impact of social media as an educational tool is an almost exclusive focus on the channel. Studies have emphasized the content that is shared (academic curriculum) rather than the range of evidence-based educational procedures that may be delivered through this channel (Mahdiuon et al., 2019). A further limitation of this literature is the almost complete lack of experimental research, which prevents the direct evaluation of educational strategies that are compatible with the social media channel and have a measurable impact on student engagement and performance. For example, Tarifa et al. (2022) evaluated the Behavioral Education and Social Media intervention package (the BE-Social Program) incorporating a collection of evidence-based behavioral intervention strategies selected based on their adaptability to the social media channel. For example, semi-immediate instructor

feedback is readily allowed by typical social media functionalities, whereas whole-class discussions may not.

1.2 Elements of a Social Media-Based Behavioral Education Intervention

The BE-Social program is composed of self-management skills training delivered through video modeling, cooperative learning environment, and semi-immediate instructor feedback (Tarifa et al., 2022).

1.2.1 Self-Management and Video Modeling

The convenience and accessibility of online platforms allow instructors to teach key self-management skills including identifying short- and long-term objectives, prioritizing, and recording time devoted to key study behaviors (writing up study notes, working on study questions, active memorization activities, etc.). The self-management component of the BE-Social program is aimed at teaching students to (a) select and define specific target behaviors, (b) observe the target behaviors and collect data, (c) chart the data and modify the definitions of the target behaviors accordingly, and (d) deliver self-reinforcement for small academic accomplishments (e.g., grade in a weekly test of 75% or higher) (Speidel et al., 1981).

Teaching self-management strategies may empower students to assume a more active role, which may have a positive impact on personal satisfaction and academic performance (Rios et al., 2018). Self-management skills have been studied extensively as a means to enhance academic engagement and performance in people with learning disabilities (Bahri et al., 2016). However, there had been few applications with adult population in tertiary education settings. In a notable exception, Choi and Chung (2012) studied the effectiveness of self-control strategies for college-level students. A total of 84 participants were divided into three experimental groups according to the intensity of the intervention or the number of strategies applied (high, medium, and low). Results indicated that students who had received a high intensity intervention (i.e., several self-control strategies) demonstrated the greatest change in performance.

Video modeling consists of presenting a video of an expert performing a target skill with the expectation that the student will imitate the skill while or after viewing the video (Nikopoulos et al., 2016, p. 187). A meta-analysis by Bellini and Akullian (2007) strongly indicated that video modeling is an effective procedure for teaching communicative and functional skills to children and adolescents with autism spectrum disorders (ASD). In addition, video modeling has been

used to teach various motor skills. For example, Quinn et. al. (2020) reported that professional dancers receiving video modeling and video feedback showed more rapid skill acquisition relative to those in the control group.

The video functionalities of social medial platforms (e.g., live broadcast, video storage and playback, video chat) suggest that video modeling is readily compatible with the use of social media in educational settings. Specifically, social media video functionalities may be used to model study habits, demonstrate course content elaboration routines, model problem-solving strategies, and other complex repertoires. The self-management skills training component of the BE-Social package is delivered via video modeling.

1.2.2 Cooperative Learning

Cooperative learning is a method whereby instructors guide the elaboration of academic content, promote student participation, and establish group expectations (Johnson & Johnson, 1982). There is evidence to suggest that positive relationships among students have a favorable impact on academic performance (Van Ryzin et al., 2020). Moreover, cooperative learning emphasizes group collaboration, peer-to-peer communication and assistance, and information and resource sharing (Roseth et al., 2008). In particular, cooperative learning environments often include discussion scenarios presented by instructors, interdependent contingencies, and social praise; strategies are known to facilitate student participation (Popkin & Skinner, 2003). Moreover, a meta-analysis by Capar and Tarim (2015) reviewed 26 studies comparing traditional teaching vs. cooperative learning and concluded that cooperative learning environments significantly improve academic performance and student motivation.

1.2.3 Feedback

Immediate feedback has shown a near universal impact on skill acquisition and performance (Kluger & DeNisi, 1996). Immediate feedback has been shown to increase self-reported interest in academic content and lead to higher academic performance among college students (He et al., 2019). University students often use instant messaging applications as a means to efficiently access academic content (Kaysi, 2021). These findings emphasize the importance of incorporating instant peer-to-peer, and instructor-student communication to online education programs, including those using social media as educational aids. The BE-Social package includes

a semi-immediate feedback component whereby instructors respond to content-related student comments within a 24-hour cycle.

Tarifa et al. (2022) randomly assigned 46 students to a default online program (control) or default online program plus BE-Social (intervention) group. Target outcomes included academic performance and social media engagement (reactions, comments). Results indicated that the intervention was effective in producing a 20% increase in academic performance and large increases in social media engagement. A parallel single-subject analysis revealed that intervention effects were, to a considerable extent, idiosyncratic. In the present study our aim is to replicate the findings by Tarifa et al. (2022) by incorporating additional methodological standards: (1) increased sample size, (2) balanced number of participants across control and experimental groups, (3) balanced number of sessions across study phases (baseline, treatment and post-treatment), (4) comparable exposure intensity to academic content across groups, and, more importantly, and, (5) utilizing a multi-arm RCT to evaluate the incremental validity of the individual elements composing the BE-Social intervention program (self-management training via video modeling, cooperative learning environment, and semi-intermediate instructor feedback).

METHODS

2.1 Participants and setting

Participants were students enrolled in an online college-level course in applied psychology. Prospective participants received an email with a detailed description of the research project, an informed consent form, and a link to a sociodemographic information questionnaire. The curricular content was located on a Moodle platform and students had access to new materials weekly. The course began in September 2020 and ended in May 2021. We invited 170 students that had relatively low performance in weekly academic tests over the first three months of the course (40% or lower on average). Of these 141 students (mean age, 41.97; SD, 8.47) agreed to participate (see a summary of participant attrition in Figure 1). Students were primarily middle-class females with a mean 16.45 ± 5.44 years of education. Table 1 present the sociodemographic characteristic of participants (see also Table A in the Supplementary Online Information for the sociodemographic information of the intention-to-treat group). Participants experienced various lockdowns and other COVID-related restrictions throughout the study. The study was approved by the ethics committee of the Universidad Autónoma de Madrid (ethics approval number CEI 112-2204).

2.3. Experimental Design

We randomly assigned 141 participants to the five arms of the study. In order to ensure the effectiveness of the randomization process, we verified that age, gender, educational level and socioeconomic standing were not statistically different across groups. Participants were blind to their group assignment status. The study was conducted over 12 successive weeks (January 18 to April 11). The baseline phase comprised the first four weeks of the study. The intervention phase followed over the next four weeks. In order to control for a potential instructor effect, the intervention phase was divided in two two-week periods and each period within each group was semi-randomly assigned to an instructor. Participants were again exposed to baseline conditions during the last phase of the study for another four weeks. Individuals in the control group remained in baseline conditions for the complete duration of the study.

2.4. Procedure

A 33-year-old male and a 32-year-old female, both doctoral students (APB, ATR) created five closed Facebook groups with similar names (“we learn together”, “together we learn”, “study group”, “study with me,” and “learn with me”) and served as instructors and moderators within these social media groups. Each group was assigned a different set of behavioral strategies: complete BE-Social program (Group 1), BE-Social without semi-immediate feedback (Group 2), cooperative learning (Group 3), and self-management (Group 4). The fifth group was the control group.

2.4.1. Baseline

During the baseline weeks, four multiple-choice discussion questions were posted daily (Monday through Friday). All posted questions covered aspects of the course content taught during that week. The posts were identical for all groups. Instructors scheduled a weekend day to provide feedback on the twenty discussion questions that had been posted over the course of the week. Feedback was provided in the form of comments to the relevant posts. The instructor would then “turn off commenting” for each of the posts after providing weekly feedback. In addition, each of the two instructors broadcasted one-hour weekly live videos featuring the instructor engaging in office-like activities (i.e., working on a computer, writing notes, and silently reading and writing flashcards). Students had the option to watch these video broadcasts synchronously or asynchronously and imitate the instructor’s behavior. Instructors did not interact with students

during the video broadcasts. Baseline activities were intended to maintain student engagement (thereby preventing participant dropout) and equating the instructor inputs across control and intervention groups. The control group was exposed to baseline conditions for the complete duration of the study.

2.4.2. Intervention

The procedures for all intervention groups are described below.

Complete BE-Social program (Group 1). Participants were exposed to all elements of the multicomponent package (self-management via video modeling, cooperative learning, and semi-immediate feedback). Each weekday the instructors created three posts using all components of the program. One post involved a self-management strategy delivered through video modeling, a second post required cooperative learning, and a third post provided the opportunity to receive semi-immediate instructor feedback. For example, one post would involve a video broadcast with the instructor modeling self-recording skills. A second post would present a practical problem intended to prompt student discussion (cooperative learning). Finally, a third post would consist of a multiple-choice scenario for which the instructor provided feedback to student comments. Instructor feedback for this last post was individualized and semi-immediate, meaning that the delay of the instructor's response was, on average, shorter than one hour. The procedures of each of these program components are described in some detail below.

Self-management was intended to encourage students to assume an active role in the organization and performance of academic tasks by fostering self-reliance and independence. The critical elements that the instructor focused on were: goal setting, self-monitoring, self-evaluation, and self-reinforcement. All of these components were presented and modeled through video modeling sessions featuring the instructor demonstrating the target self-management skill. For the purposes of teaching goal-setting skills, instructors made 10-min video broadcasts presenting a step-by-step approach to creating a list of specific, measurable, and achievable goals using specific course topics from the program syllabus as to ensure relevance. The instructors demonstrated how to create a study planning schedule considering time available, amount of study material, reading time, and time needed to complete assignments and self-assessment tests. The students could ask questions during or after the video broadcast, and could share their

own study planning schedule to receive feedback from the instructor. The self-monitoring component encouraged students to assess their progress as they worked toward their study goals. The instructors used a video broadcast to present self-recording strategies, which required students to time the duration of academic activities that had been defined during the goal-setting phase. Then, the instructor demonstrated how to transfer daily study time records into a spreadsheet for graphing. The instructor added a weekly test performance data line to the graph and asked students to follow suit. Instructors encouraged students to check their weekly graphs and correlate study time with performance. Finally, instructors encouraged students to take extra free time (self-reinforcement), on occasions when they had improved both their study time and performance. The self-management procedures were largely implemented as an antecedent-based intervention with minimal contingent feedback beyond answering questions or providing sporadic feedback to students that presented their goal setting schedule during a video broadcast (see Appendices 1 and 2 in the Supplementary Online Information for examples of self-management training via video modeling posts).

The goal of the cooperative learning component was to establish positive interactions among students by encouraging praise and informative feedback as part of their communication style. Cooperative learning posts could contain an image, a short video, a text, a combination of text and video, or a combination of text and image. Each post involved one of two possible activities: (a) presenting an applied scenario where multiple answers could be considered correct as to generate discussion among students, and (b) presenting the outline of a student study time plan with blank spaces as to generate cooperation among students to complete the proposed plan. The instructor further facilitated interaction by (a) presenting demand-free hints and prompts when communication halted (e.g., “This scenario is so relevant to professional practice! Have you considered the potential ethical ramifications?”), (b) tagging three students to each student comment, (c) tagging students that had been inactive for the last 24 hours, and (d) providing weekly general feedback on the level of student communication and publicly praising those students who had given the most feedback to the rest of their peers. The instructor did not provide semi-immediate feedback as part of the cooperative learning procedures.

BE-Social program without semi-immediate feedback (Group 2). The intervention procedures for this group were identical to those described for Group 1 with the exception of the

semi-immediate feedback component. Specifically, instructors provided feedback on a weekly basis over the weekend (as opposed to providing feedback within a 24-cycle or sooner).

Cooperative learning (Group 3). The intervention procedures for the cooperative learning group were identical to those described for the cooperative learning component of Group 1 with the exception that all three daily posts were cooperative learning posts. Therefore, they only received the cooperative learning component of the BE-Social program.

Self-management training via video modeling (Group 4). The intervention procedures for the self-management training via video modeling group were identical to those described for this component of the BE-Social group (Group 1) with the exception that all three daily posts involved study self-management skills training delivered via video modeling. Therefore, they only received the self-management component of the BE-Social program.

2.4.3 Post-Intervention

The post-intervention phase procedures were identical to those described for the baseline phase.

2.5 Dependent variables

Participants took a total of twelve weekly tests over the course of the study. Each test was composed of 20 multiple-choice questions with three distractors and one correct answer. We calculated test scores by dividing correct responses by 20 and converting that ratio into a percentage. Ninety percent of the scenarios described practical applications of the course contents, while the remainder of the questions focused on conceptual information (e.g., definitions, characteristics, conceptual clarification, etc.). In order to minimize the effect of content difficulty on the average performance of participants, the online tests, which were published weekly, could be completed over a three-week time window.

We recorded the number of students that had viewed any given instructor post (views), and the total number of student comments and reactions (e.g., likes) generated by each instructor post. We summarized these data as *combined engagement*, defined as the sum of views, likes, and comments per post.

2.5. Interobserver agreement and procedural integrity

A secondary observer obtained data on student engagement (views, reactions, and comments) and instructor latency in resolving or responding to students. Twenty percent of instructor posts were selected from each of the five groups to assess interobserver agreement (IOA). The IOA was calculated by dividing the number of posts with agreement by the total number of posts selected for evaluation. Overall, there was 99.6% (range, 59% to 100%), 99.9% (96% to 100%), 100%, and 100% agreement on instructor feedback latency, views, comments, and likes, respectively.

In order to assess whether the study procedures were implemented as expected, an independent observer recorded key aspects of the intervention procedures throughout all study phases including: number of daily posts, number of weekly videos, latency to instructor feedback (weekly vs. semi-immediate), and video content (delivery of self-management strategies). For the purposes of video content monitoring, the independent observer conducted a 60-s partial interval recording of 20% of instructor video broadcasts. The observation protocol was intended to verify the amount of time instructors devoted to discussing self-management skills. Observations were conducted with the app Big Eye Observer, which have been validated for video-based behavioral observation applications (ABA España, 2019, Virues-Ortega et al., 2022). Instructors also recorded the following intervention-specific procedural integrity standards during the intervention phase: number of self-management posts, format of self-management posts (video vs. text), number of cooperative learning posts, and mean latency of instructor feedback. A comprehensive summary of the procedural integrity assessment is available in the online supplementary information (Table B).

2.6. Statistical analysis

2.6.1. Academic achievement

We conducted a mixed-model ANOVA with group as between-groups factor and time (Time 1: Baseline; Time 2: Intervention; Time 3: Post-Intervention) as within-subjects factor. The dependent variable was the mean performance in weekly course content tests (range, 0 to 20). Achieved power and effect size metrics are reported for the omnibus tests. In addition, we computed t-test pairwise group and time comparisons and estimated Hedges g effect sizes for any effects identified through the omnibus mixed-model ANOVA.

2.6.2. Social media engagement

The Facebook platform does not provide access to the user's name producing visualization engagement responses. This constrain meant that we should use posts as units of analysis (as opposed to subjects within a group) for the analysis of engagement responses. The number of engagement responses (i.e., views, likes, comments) across posts were not normally distributed. This, added to the longitudinal (i.e., successive posts within a group) and hierarchical structure (i.e., group-level predictors) of the engagement responses, recommended the use of hierarchical linear models (HLM). We used combined engagement, defined as the sum of views, likes, and comments per post, as our dependent variable. We used the group identifier as subject variable and post number within a group as time-based variable. A first-order autoregressive covariance structure rendered the best goodness of fit values during the model development process. We used group (BE-Social, BE-Social without semi-immediate feedback, self-management, cooperative learning, control), time (baseline, treatment, posttreatment), post type (textual, visual, video), and the interaction between group and post type as fixed-effect factors. In addition, the group-level mean academic performance was added as a group-nested random-effects factor. The Akaike's information criterion (AIC) of the unconditional model was reduced gradually by the stepwise addition of each factor (Akaike, 1974; Burhan & Anderson, 2002). The AIC values were 6991 and 6700 in the unconditional and terminal models, respectively. The addition of a random intercept did not improve model fitness. Group-nested academic performance was significantly correlated with phase and group and was the subject of a second model without obvious confounders. All model coefficients are expressed in the same metric as the dependent variable (range, 0 to 1).

All analyses were conducted with SPSS® IBM® Statistics, version 27 (IBM Corporation, 2021). We used a *p* value of 0.05 with Bonferroni adjustments for multiple comparisons.

2.7. Social Validity and Acceptability

At the end of the study, the instructors made a farewell live video for each of the groups thanking students for their participation and inviting comments about their experiences and concerns. We recorded the unsolicited comments of the live-streamed farewell videos. These comments were processed through a textual analysis tool capable of identifying common semantic themes or categories from textual samples (Hunerberg, 2019). We report the percentage of unsolicited

positive feedback and the themes resulting from the textual analysis as indirect indications of the social validity (i.e., acceptability) of the intervention.

RESULTS

3.1. Academic achievement

The Box test for the mixed-model ANOVA was not significant ($p = .179$) suggesting that the assumption of equality of covariance matrices was met. Multivariate tests following the Wilks' lambda distribution revealed a significant Time effect ($F [2, 135] = 66.11, p < .001, \eta^2 = 0.50, 1-\eta = 1.00$) and Time by Group interaction ($F [8, 270] = 4.06, p < .001, \eta^2 = 0.12, 1-\eta = .99$). The corrected within-subject effect for Time ($F [1.69, 229.11] = 42.92, p < .001, \eta^2 = 0.24, 1-\eta = 1.00$) and Time by Group interaction ($F [6.74, 229.11] = 3.25, p = .003, \eta^2 = 0.09, 1-\eta = .95$) were also significant. The main effect of group was also significant, $F (4, 136) = 4.63, p = .002, \eta^2 = 0.12, 1-\eta = .94$. The homogeneity assumption for the equality of error variances across groups was met. A main effect of Group was established as well, $F (4, 136) = 126.05, p = .002, \eta^2 = 0.12, 1-\eta = .94$. As it seems apparent from Figure 1, within-subject contrasts indicated the effect of time followed a quadratic rather than linear trend ($F [1, 136] = 7.73, p = .006, \eta^2 = 0.05$ vs. $F [1, 136] = 130.11, p < .001, \eta^2 = 0.49$) suggesting that the intervention effects were to a certain degree transient for the intervention groups, while the control group followed a deteriorating trend over time (Figure 1).

The pairwise t-test comparisons indicated that the baseline academic performance in any of the intervention groups was not statistically different from that of the control group, $p < .01$ (Table 2). The only intervention group that did not differ from the control group during the intervention phases was the self-management intervention group (Group 4). All other intervention groups showed statistically significant departures in academic performance, both during the intervention (Time 2) and post-intervention phases (Time 3) (Table 2). Specifically, the complete BE-Social group (Group 1) showed a statistically significant increase in performance during the intervention, $t(1.52) = 5.38, p < .001$, and post-intervention phases, $t(1.52) = 3.53, p < .001$ relative to the control group. The effect sizes were both within the large effect size range ($g = 1.44, 98.75\% \text{ CI } 0.68, 2.97$ and $g = 0.95, 95\% \text{ CI } 0.23, 1.65$, respectively). The slightly lower post-intervention effect suggests that the effect was partially lost after the intervention withdrawal.

The BE-Social without semi-immediate feedback group (Group 2) showed a statistically significant increase in performance during the intervention ($t[1, 52] = 3.67, p < 0.001$) and post-intervention phases ($t[1, 52] = 2.29, p = 0.03$). The pairwise effect size of Group 2 against the control group for the intervention and post-intervention phases was 0.98 (95% CI 0.27, 1.69) and 0.61 (95% CI -0.08, 1.29), respectively.

For the cooperative learning group (Group 3), a slight increase in performance was observed during the intervention and post-intervention phases ($t[1, 58] = 2.48, p = .016$ and $t[1, 58] = 2.44, p = .018$, respectively). However, using the control group as a reference, the intervention effect for Group 3 was somewhat smaller than the one observed for Groups 1 and 2. Group 3 showed nearly identical effect sizes within the moderate effect size range for both the intervention and post-intervention phases ($g = 0.64, 98.75\% \text{ CI } -0.02, 1.29$ and $g = 0.63, 95\% \text{ CI } -0.03, 1.28$, respectively). Therefore, the effect of the intervention was sustained during the post-intervention phase. However, as it was the case for all other groups, there was performance deterioration during the post-intervention phase (Figure 2).

Finally, the pairwise comparison of the self-management skills training group (Group 4) relative to the control group did not reveal statistically significant differences during the intervention and post-intervention phases ($t[1, 52] = 1.61, p = .097$ and $t[1, 58] = 1.69, p = .016$, respectively).

Overall, the multi-arm RCT analysis indicated that semi-immediate feedback was critical to the effect of the multi-component intervention, whereas the value of self-management skills training could not be verified in the unimodal intervention.

6.2 Social media engagement

The final linear mixed model analysis for combined social media engagement (Table 3) confirmed a significant effect of Group ($F[1, 800] = 16.69, p < .001$), Time ($F[2, 288] = 51.79, p < .001$), Post type ($F[2, 859] = 36.28, p < .001$), and Group by Post type interaction ($F[8, 880] = 8.48, p < .001$). Pairwise comparisons indicated that only the complete BE-Social program (Group 1) and the BE-Social program without semi-immediate feedback (Group 2) had higher levels of social media engagement relative to the control group ($i-j = 7.78 \pm 2.51, p = .008$ and $= 7.66 \pm 2.54, p = .01$, respectively). Engagement levels across groups did not change during the intervention phase relative to baseline, whereas engagement decreased across groups during the post-in-

tervention phase relative to baseline ($i-j = 8.46 \pm 0.91$, $p < .001$). Finally, textual posts induced greater interaction across groups than visual (e.g., image with text) or video posts, but only mean difference between textual and video posts was statistically significant in the pairwise analysis ($i-j = 9.59 \pm 1.14$, $p < .001$). While a mediation analysis was not feasible (some of the engagement responses could not be affiliated with individual participants), there seem to be an apparent correlation between increments in social media engagement and increments in academic performance as expressed in the mean differences of both outcomes during the intervention phase (Figure 3).

3.3 Social Validity and Acceptability

A total of 46 participants provided unsolicited positive comments during the farewell video broadcasts. Therefore, 32.6% of participants produced unsolicited positive comments. The BE-Social without semi-immediate feedback group provided the most positive comments (51.8%), followed by the complete BE-Social program group (44.5%). Only 18% of participants in the self-management and cooperative learning groups volunteered positive comments. Finally, 33% of participants in the control group shared positive comments. No neutral or negative comments were posted in any of the groups (see Supplementary Online Material, Table C). The sentence-by-sentence thematic analysis indicated that the five most prevalent themes in the feedback messages were about knowledge acquisition (35 occurrences, 7.5% of the text), high-quality teaching (27 occurrences, 5.8% of the text), commendation for the video broadcasts (14 occurrences, 3.0% of the text), appreciation for the problem-solving strategies presented (10 occurrences, 2.2% of the text), and perceived value of the experience (9 occurrences, 1.9% of the text).

DISCUSSION

The present study expands the evidence base of social media platforms as an aid to tertiary education. We implemented the multi-component BE-Social program as an adjunct to an online applied psychology course to assess the efficacy of the key elements of this program individually and in combination. The results suggest that participants who were exposed to the complete intervention package (including semi-immediate feedback) demonstrated the highest academic performance. While the presence of semi-immediate feedback did not affect the total amount of social media engagement, the current multi-arm RCT seems to suggest that feedback was a critical element of the program. Our results also indicate that part of the intervention gains may

be quickly lost after the intervention is withdrawn, thereby suggesting that the program ought to be present for the duration of the course for optimal results. The isolated implementation of discrete components of the program (particularly self-management training) had relatively small effects on performance.

Our findings are in line with the empirical literature suggesting that fast turn-around feedback is conducive to student learning. For example, a systematic review by Liu (2021) found that the immediacy of teacher responses had a positive impact on student engagement with the academic material and, ultimately, on performance. In addition, immediacy has been identified as a key dimension to effective feedback (Er et al., 2021; Henderson & Wen, 1976). Also, there is evidence to suggest that students prefer receiving feedback from educators through instant messaging as opposed to email (Gopal et. al., 2021). While our study does not demonstrate reward learning effects at the individual level, these findings are in line with a delay-discounting effect, which has been described in detail in the animal and human operant literatures (Reynolds, 2006). Specifically, a relatively immediate reward (e.g., immediate instructor feedback) may have a greater impact on engagement and other relevant academic behaviors than more delayed ones.

Interestingly, students in the BE-Social group without semi-immediate feedback (Group 2) showed high levels of social media engagement, while their level of academic achievement was relatively lower than the one observed in the complete BE-Social group (Group 1). This could indicate that social media engagement may not be an accurate predictor of course content elaboration. An alternative interpretation may be that, although frequent student posting can still occur under delayed instructor feedback (for example, as a result of being exposed to a cooperative learning environment), it may be less likely to induce successful content acquisition and elaboration.

Social media engagement has been shown to accommodate to simple social reward optimization processes (Lindström et al., 2021). Such realization has obvious practical implications that were indirectly evaluated here. Specifically, our analysis suggests that the cooperative learning and semi-immediate feedback strategies may be particularly effective in mobilizing student online behavior in study groups. While the mediating role of social media responses in academic achievement cannot be asserted, the study provides observational evidence that those groups that showed more active online behavior showed also relatively higher achievement. The study also provides a demonstration that the facility of social media environments to produce imme-

mediate interactions (e.g., push notifications, tagging, reacting, commenting, etc.) can be easily integrated in evidence-based educational practices including the design of a cooperative learning environment, the delivery of self-management skills training, and the effective use of instructor-mediated feedback.

The analytical approach of the multi-arm RCT suggests that the elements of the proposed multi-component intervention are likely to be additive with the cooperative learning and semi-immediate feedback manipulations being the most likely contributors to the compounded effect observed in the full implementation of the intervention package. While the self-management training intervention had an apparent effect on performance (Figure 1) the effect was not statistically significant. Given that the multi-arm RCT was intended to evaluate discrete components of the intervention, whether isolated or in combination, additional analyses would be needed to substantiate the potential additivity of the critical elements of the BE-Social package (i.e., self-management training, cooperative learning environment, instructor-mediated semi-immediate feedback). While an *additive effect hypothesis* can hardly be verified with between-groups datasets (see for example Van Iddekinge et al., 2017), we can tentatively examine the mismatch between empirical effect sizes and the theoretical additive effects. Specifically, a plausible additive effect would mean that the sum of the unimodal interventions without semi-immediate feedback would approach the effect size of the BE-Social group without semi-immediate feedback,

$$g_3 + g_4 \approx g_2 \quad (1)$$

$$0.64 + 0.45 \approx 0.98$$

Where g_2 , g_3 , and g_4 are the pairwise comparison Hedges g effect size for Groups 2, 3, and 4, respectively. Similarly, the additive effect hypothesis would be strengthened if the incremental effect of semi-immediate feedback and the separate effects of the unimodal interventions (Group 3 and Group 4) would approach the effect size of the full multi-component intervention (Group 1),

$$g_3 + g_4 + (g_1 - g_2) \approx g_1 \quad (2)$$

$$0.64 + 0.45 + (1.44 - 0.98) \approx 1.44$$

Where g_1 , g_2 , g_3 , g_4 , and g_5 are the pairwise comparison Hedges g effect sizes for Groups 1 through 5, respectively. The observed empirical effect sizes match the theoretical additive relation within 0.1 effect size units in both vases (0.98 vs. 1.09 and 1.44 vs. 1.55, respectively). While

this demonstration is anecdotal, it does suggest that the multi-arm approach may be an analytic tool to weight the relative contribution of multi-component educational programs.

In sum, the results show that the combination of all the selected educational strategies may have a synergic effect in that each of the intervention components, when applied individually, seem to detract from the effect of the complete program. However, the current analytical approach was not comprehensive. Specifically, we could not assess semi-immediate feedback in the absence of other intervention procedures. Moreover, the baseline condition to which the control group was exposed for the duration of the study involved minimal interactions that were deemed essential to maintain student morale and engagement before the implementation of the intervention phase. Critical comparisons involved the addition of elements to a background intervention rather than comparing the absence of intervention to a discrete intervention component. Therefore, complex interactions between the instructor baseline behavior and subsequent performance cannot be discarded.

Interestingly, instructor textual posts generated more engagement than video broadcasts (i.e., self-management training via video modeling posts) or mixed posts with both text and images. While studies that quantitatively evaluate engagement are rare, some surveys suggest that text-based posting tends to generate more interactive engagement, whereas video-based and mixed posts seem to impose a one-sided communication dynamic (see for example Swartzwelder et al., 2019). The nature of the content may also play a role. For example, text-based posts presented multiple-choice scenarios that frequently prompted student guessing and follow up questions.

Several limitations and future extensions to the current study should be noted. First, we have examined the compatibility of an incidental collection of behavioral education strategies using one of the most popular social media platforms for tertiary education students. Future studies could replicate our findings with alternative combinations of evidence-based strategies and social media platforms. Also, the specific parameters of the strategies utilized here could also be modified (e.g., elements of a cooperative learning environment, aspects of the self-management protocol). For example, the proposed self-management protocol may be enhanced by adding performance-based feedback as opposed to relying primarily on video modeling. In sum, further replications varying procedural aspects of the intervention may lead to a greater optimization of the proposed protocol.

Second, our data collection strategy did not allow us to reconcile discrete online events (e.g., reactions, views, posts, comments) with individual students, which meant that the impact of the intervention could only be assessed at the group level and that mediation analyses with social media engagement as a potential mediating factor could not be conducted. This also meant that idiosyncratic patterns of treatment effects could not be studied in detail (for an analysis of idiosyncratic effects of the BE-Social program see Tarifa et al., 2022). Future studies in this area could minimize these concerns by way of purposely-developed application programming interfaces (APIs) that would automate aspects of the data collection process in the social media environment. Additional analytical enhancements include the integration of text and theme analysis systems that could provide hints on the social dynamics and emotional factors that involved in social media exchanges in educational contexts (see for example Drus & Khalid, 2019).

Third, our outcome measures were limited to *ad hoc* course content tests and standard social media responses (i.e., views, likes, comments, posts). These outcomes were important in that they allowed frequent measurements across study phases and were ecologically valid in that they were naturalistically integrated with the course. However, future analyses would benefit from the addition of standardized measurements of student satisfaction and end-of-course assessments. The latter were not practical in the current study owing to the multi-phase structure of the design, which had to be fitted within a one-year course. While frequent testing allowed the opportunity to monitor the intervention effects over time, it prevented a more comprehensive evaluation of the overall effect of the program on student academic achievement.

Fourth, the apparent deteriorating academic achievement trend observed in the control group, which was continuously exposed to baseline conditions, may be attributed to selecting low-performance students (see inclusion criteria), or to the hierarchical nature of course contents (i.e., latter lessons relied in the foundational knowledge of earlier ones). It is interesting to note that the elements of the BE-Social program were sufficient to offset this deteriorating trend, even though intervention gains were not fully sustained during the post-intervention phase.

CONCLUSIONS

The following conclusions can be drawn from the current analysis: (1) the unique combination of self-management training via video modeling, cooperative learning online environment, and semi-immediate instructor feedback included in the BE-Social program had a large positive im-

pact on social media engagement and academic achievement, (2) the effect of the elements of the intervention seem to be additive, although we could only verify the incremental validity of semi-immediate feedback, (3) increased social media engagement may be a mediating factor to increased academic achievement, and (4) the multi-component intervention may have optimal effects on achievement when delivered for the complete duration of the course.

With the exponential growth of the casual use of social media in tertiary education, it would be important for educators to evaluate the potential of the social media channel as a medium for evidence-based educational practices. Integrating a social media study group in tertiary education courses may be a cost-effective approach to enhancing peer-to-peer and student-instructor interactions with the end goal of improving academic achievement. The proposed program demonstrated that key evidence-based educational practices such as the design of a cooperative learning environment and the delivery of semi-immediate instructor feedback are readily amenable to a social media study group and can quickly enhance student participation and performance.

REFERENCES

- ABA España (2019). *Big Eye Observer* (Version 2.0). [Mobile App]. <https://apps.apple.com/us/app/big-eye-observer/id1299214188>
- Ali, W. (2020). Online and remote learning in higher education institutes: a necessity in light of COVID-19 pandemic. *Higher Education Studies*, 10(3), 16. <https://doi.org/10.5539/hes.v10n3p16>
- Alkatout, I., Dhanawat, J., Ackermann, J., Freytag, D., Peters, G., Maass, N., Mettler, L., & Pape, J. M. (2021). Video feedback and video modeling in teaching laparoscopic surgery: a visionary concept from Kiel. *Journal of Clinical Medicine*, 10(1), 163. <https://doi.org/10.3390/jcm10010163>
- Alshuaibi, M. S. I., Alshuaibi, A. S. I., Shamsudin, F. M., & Arshad, D. A. (2018). Use of social media, student engagement, and academic performance of business students in Malaysia. *International Journal of Educational Management*, 32(4), 625–640. <https://doi.org/10.1108/IJEM-08-2016-0182>
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723. <https://doi.org/10.1109/TAC.1974.1100705>

- Arteaga Sánchez, R., Cortijo, V., & Javed, U. (2014). Students' perceptions of Facebook for academic purposes. *Computers & Education*, 70, 138–149. <https://doi.org/10.1016/J.COMPE-DU.2013.08.012>
- Ayebi-Arthur, K. (2017). E-learning, resilience and change in higher education: helping a university cope after a natural disaster. *E-Learning and Digital Media*, 14(5), 259–274. <https://doi.org/10.1177/2042753017751712>
- Azizi, S. M., Soroush, A., & Khatony, A. (2019). The relationship between social networking addiction and academic performance in Iranian students of medical sciences: a cross-sectional study. *BMC Psychology*, 7(1), 28. <https://doi.org/10.1186/s40359-019-0305-0>
- Bahri, L., Mirnasab, M., Noorazar, G., Fathi Azar, E., & Asadi, S. (2016). The effect of self-management strategies on improving task performance in students with learning disabilities. *Journal of Analytical Research in Clinical Medicine*, 4(1), 53–65. <https://doi.org/10.15171/jarcm.2016.009>
- Bartini, M. (2008). An empirical comparison of traditional and web-enhanced classrooms. *Journal of Instructional Psychology*, 35(1), 3–11. <https://doi.org/10.1145/1562741.1562743>
- Baytiyeh, H. (2018). Online learning during post-earthquake school closures. *Disaster Prevention and Management: An International Journal*, 27(2), 215–227. <https://doi.org/10.1108/DPM-07-2017-0173>
- Beaunoyer, E., Dupéré, S., & Guitton, M. J. (2020). COVID-19 and digital inequalities: Reciprocal impacts and mitigation strategies. *Computers in Human Behavior*, 111, 106424. <https://doi.org/10.1016/j.chb.2020.106424>
- Bellini, S., & Akullian, J. (2007). A meta-analysis of video modeling and video self-modeling interventions for children and adolescents with autism spectrum disorders. *Exceptional Children*, 73(3), 264–287. <https://doi.org/10.1177/001440290707300301>
- Bertucci, A., Conte, S., Johnson, D. W., & Johnson, R. T. (2010). The impact of size of cooperative group on achievement, social support, and self-esteem. *Journal of General Psychology*, 137(3), 256–272. <https://doi.org/10.1080/00221309.2010.484448>
- Burnham, K. P., & Anderson, D. R. (2002). *Model Selection and Multimodel Inference* (2nd ed.). Springer.

- Capar, G., & Tarim, K. (2015). Efficacy of the cooperative learning method on mathematics achievement and attitude: A meta-analysis research. *Educational Sciences: Theory & Practice*, 15(6), 553–559. <https://doi.org/10.12738/estp.2015.2.2098>
- Cheawjindakarn, B., Suwannatthachote, P., & Theeraroungchaisri, A. (2012). Critical success factors for online distance learning in higher education: a review of the literature. *Creative Education*, 3(8), 61–66. <https://doi.org/10.4236/ce.2012.38b014>
- Cloete, S., & Villiers, C. de. (2009). Facebook as an academic tool for ICT lecturers. *Proceedings of the 2009 Annual Conference of the Southern African Computer Lecturers' Association*, 16-22. <https://doi.org/10.1145/1562741.1562743>
- Choi, J. H., & Chung, K. M. (2012). Effectiveness of a college-level self-management course on successful behavior change. *Behavior Modification*, 36(1), 18–36. <https://doi.org/10.1177/0145445511418102>
- Cooper, J. O., Heron, T. E., & Heward, W. L. (2019). *Applied Behavior Analysis* (3rd ed.). Pearson.
- Dhawan, S. (2020). Online learning: a panacea in the time of COVID-19 crisis. *Journal of Educational Technology Systems*, 49(1), 5–22. <https://doi.org/10.1177/0047239520934018>
- Drus, Z., & Khalid, H. (2019). Sentiment analysis in social media and its application: Systematic literature review. *Procedia Computer Science*, 161, 707–714. <https://doi.org/10.1016/j.procs.2019.11.174>
- Er, E., Dimitriadis, Y., & Gašević, D. (2021). A collaborative learning approach to dialogic peer feedback: a theoretical framework. *Assessment and Evaluation in Higher Education*, 46(4), 586–600. <https://doi.org/10.1080/02602938.2020.1786497>
- Giannikas, C. (2019). Facebook in tertiary education: the impact of social media in e-learning. *Journal of University Teaching and Learning Practice*, 17(1). <https://doi.org/10.53761/1.17.1.3>
- Gopal, R., Singh, V., & Aggarwal, A. (2021). Impact of online classes on the satisfaction and performance of students during the pandemic period of COVID 19. *Education and Information Technologies*, 26, 6923–6947. <https://doi.org/10.1007/s10639-021-10523-1>
- Graham, M. (2014). Social media as a tool for increased student participation and engagement outside the classroom in higher education. *Journal of Perspectives in Applied Academic Practice*, 2(3), 16–24. <https://doi.org/10.14297/jpaap.v2i3.113>

- He, J., Hao, W., & Kim, J.-W. (2019). The effects of instant feedback system on course interest and academic achievement in gamification learning. *Educational Innovations and Applications*, 467–470. <https://doi.org/10.35745/ecei2019v2.119>
- Henderson, W. T., & Wen, S. S. (1976). Effects of immediate positive reinforcement on undergraduates' course achievement. *Psychological Reports*, 39(2), 568-570. <https://doi.org/10.2466/pr0.1976.39.2.568>
- Hunerberg, E. (2019). *3rd Eye Theme Analysis* [Online Application]. <https://3rdeyeinformation.com/home>
- Irwin, C., Ball, L., Desbrow, B., & Leveritt, M. (2012). Students' perceptions of using Facebook as an interactive learning resource at university. *Australasian Journal of Educational Technology*, 28(7), 1221–1232. <https://doi.org/10.14742/ajet.798>
- Johnson, D. W., & Johnson, R. T. (1982). The effects of cooperative and individualistic instruction on handicapped and nonhandicapped students. *The Journal of Social Psychology*, 118(2), 257–268. <https://doi.org/10.1080/00224545.1982.9922805>
- Kaysi, F. (2021). Mobile instant messaging application habits among university students. *Interactive Learning Environments*, 1–19. <https://doi.org/10.1080/10494820.2021.1922460>
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254–284. <https://doi.org/10.1037/0033-2909.119.2.254>
- Kusumawati, A. J. (2020). Redesigning face-to-face into online learning for speaking competence during covid-19: ESP for higher education in Indonesia. *International Journal of Language Education*, 4(2), 276–288. <https://doi.org/10.26858/ijole.v4i2.14745>
- Li, S., Zhu, Y., & Ellis, R. (2016). The effects of the timing of corrective feedback on the acquisition of a new linguistic structure. *Modern Language Journal*, 100(1), 276–295. <https://doi.org/10.1111/modl.12315>
- Liu, W. (2021). Does teacher immediacy affect students? a systematic review of the association between teacher verbal and non-verbal immediacy and student motivation. *Frontiers in Psychology*, 12, 713978. <https://doi.org/10.3389/fpsyg.2021.713978>

- Mah, E., Yu, J., Deck, M., Lyster, K., Kawchuk, J., Turnquist, A., & Thoma, B. (2021). Immersive video modeling versus traditional video modeling for teaching central venous catheter insertion to medical residents. *Cureus*, 13(3) <https://doi.org/10.7759/cureus.13661>
- Mahdiuon, R., Salimi, G., & Raeisy, L. (2019). Effect of social media on academic engagement and performance: Perspective of graduate students. *Education and Information Technologies*, 2427–2446. <https://doi.org/10.1007/s10639-019-10032-2>
- Mansholt, H., Götz, N. A., & Babitsch, B. (2021). Chances and barriers of online problem-based learning (ePBL) for advanced training in the healthcare sector. *Studies in Health Technology and Informatics*, 281, 822–823. <https://doi.org/10.3233/SHTI210292>
- Martínez-Cardama, S., & Caridad-Sebastián, M. (2019). Social media and new visual literacies: Proposal based on an innovative teaching project. *Education for Information*, 35(3), 337–352. <https://doi.org/10.3233/EFI-180214>
- McCutcheon, K., Lohan, M., Traynor, M., & Martin, D. (2015). A systematic review evaluating the impact of online or blended learning vs. face-to-face learning of clinical skills in undergraduate nurse education. *Journal of Advanced Nursing*, 71(2), 255–270. <https://doi.org/10.1111/jan.12509>
- Means, B., Toyama, Y., Murphy, R., & Baki, M. (2013). The effectiveness of online and blended learning: A meta-analysis of the empirical literature. *Teachers College Record*, 115(3). <https://doi.org/10.1177/016146811311500307>
- Nikopoulos, C., Luiselli, J. K., & Fischer, A. J. (2016). Video Modeling. In J. K. Luiselli, & Aaron J. Fischer (Eds.), *Computer-Assisted and Web-Based Innovations in Psychology, Special Education, and Health* (pp. 187–210). Academic Press. <https://doi.org/10.1016/B978-0-12-802075-3.00008-5>
- Meşe, E. & Sevilen, Ç. (2021). Factors influencing EFL students' motivation in online learning: A qualitative case study. *Journal of Educational Technology and Online Learning*, 4(1), 11-22 . Retrieved from <https://dergipark.org.tr/en/pub/jetol/issue/60134/817680>
- Popkin, J., & Skinner, C. H. (2003). Enhancing academic performance in a classroom serving students with serious emotional disturbance: Interdependent group contingencies with randomly selected components. *School Psychology Review*, 32(2), 282–295. <https://doi.org/10.1080/02796015.2003.12086199>

- Quinn, M., Narozanick, T., Miltenberger, R., Greenberg, L., & Schenk, M. (2020). Evaluating video modeling and video modeling with video feedback to enhance the performance of competitive dancers. *Behavioral Interventions*, 35(1), 76–83. <https://doi.org/10.1002/bin.1691>
- Rega, A., Somma, F., & Iovino, L. (2020). Development of emotional skills through video modeling: A case study with a non-verbal participant. *Technology and Disability*, 32(3), 211–218. <https://doi.org/10.3233/TAD-180222>
- Reynolds B. (2006). A review of delay-discounting research with humans: relations to drug use and gambling. *Behavioural Pharmacology*, 17(8), 651–667. <https://doi.org/10.1097/FBP.0b013e3280115f99>
- Rios, D., Kazemi, E., & Peterson, S. M. (2018). Best practices and considerations for effective service provision via remote technology. *Behavior Analysis: Research and Practice*, 18(3), 277–287. <https://doi.org/10.1037/bar0000072>
- Roblyer, M. D., McDaniel, M., Webb, M., Herman, J., & Witty, J. V. (2010). Findings on Facebook in higher education: A comparison of college faculty and student uses and perceptions of social networking sites. *Internet and Higher Education*, 13(3), 134–140. <https://doi.org/10.1016/j.iheduc.2010.03.002>
- Roseth, C. J., Johnson, D. W., & Johnson, R. T. (2008). Promoting early adolescents' achievement and peer relationships: The effects of cooperative, competitive, and individualistic goal structures. *Psychological Bulletin*, 134(2), 223–246. <https://doi.org/10.1037/0033-2909.134.2.223>
- Selvaraj, A., Radhin, V., KA, N., Benson, N., & Mathew, A. J. (2021). Effect of pandemic based online education on teaching and learning system. *International Journal of Educational Development*, 85, 102444. <https://doi.org/10.1016/j.ijedudev.2021.102444>
- Speidel, G. E., Tharp, R. G., & Speidel, G. E. (1981). What does self-reinforcement reinforce? An empirical analysis of the contingencies in self-determined reinforcement. *Child Behavior Therapy*, 2(4), 1–22. https://doi.org/10.1300/J473V02N04_01
- Statista (2022, March.). Facebook: Distribution of Global Audiences 2022, by Age and Gender. <https://www.statista.com/statistics/376128/facebook-global-user-age-distribution/>

- Swartzwelder, K., Murphy, J., & Murphy, G. (2019). The impact of text-based and video discussions on student engagement and interactivity in an online course. *Journal of Educators Online*, 16(1). <https://doi.org/10.9743/jeo.2019.16.1.13>
- Topping, K. J., Douglas, W., Robertson, D., & Ferguson, N. (2022). Effectiveness of online and blended learning from schools: A systematic review. *Review of Education*, 10(2), e3353 <https://doi.org/10.1002/rev3.3353>
- Tran, P. (2016). Training learners to use quizlet vocabulary activities on mobile phones in Vietnam with Facebook. *JALT CALL Journal*, 12(1), 43–56. <https://doi.org/10.29140/jaltcall.v12n1.201>
- Van Iddekinge, C. H., Aguinis, H., Mackey, J. D., & DeOrtentiis, P. S. (2017). A meta-analysis of the interactive, additive, and relative effects of cognitive ability and motivation on performance. *Journal of Management*, 20(10), 1-31. <https://doi.org/10.1177/0149206317702220>
- Van Ryzin, M. J., Roseth, C. J., & Biglan, A. (2020). Mediators of effects of cooperative learning on prosocial behavior in middle school. *International Journal of Applied Positive Psychology*, 5(1–2), 37–52. <https://doi.org/10.1007/s41042-020-00026-8>
- Virues-Ortega, J., Delgado Casas, C., Martin, N., Tarifa-Rodriguez, A., Reina Hidalgo, A. J., & Cox, A. D., & Navarro Guzmán, J. I. (2022) Interobserver agreement in software-aided and paper-and-pencil behavioral observation. *Behavior Research Methods*. <https://doi.org/10.3758/s13428-022-01861-0>
- Wang, W. T., & Lin, Y. L. (2021). Evaluating factors influencing knowledge-sharing behavior of students in online problem-based learning. *Frontiers in Psychology*, 12, 691755. <https://doi.org/10.3389/fpsyg.2021.691755>

Table 11.1 Sociodemographic Characteristics of Participants (n = 141)

	Total (n = 141)	G1 (n = 27)	G2 (n = 33)	G3 (n = 27)	G4 (n = 27)	G5 (n = 27)
Gender, % (n)						
Female	90.07 (127)	88.89 (24)	90.91 (30)	96.30 (26)	96.30 (26)	77.78 (21)
Male	9.92 (14)	11.11 (3)	6.06 (3)	3.70 (1)	3.70 (1)	22.22 (6)
Mean age, years (SD)	41.97 (8.47)	40.11 (8.66)	41.48 (8.34)	42.78 (7.12)	44.56 (9.47)	41.04 (8.52)
Years of education (SD)	16.45 (5.44)	17.70 (5.09)	16.67 (5.61)	15.96 (5.27)	17.44 (5.25)	14.41 (5.65)
Mean BL performance (SD)	12.33 (2.18)	12.35 (1.72)	12.46 (2.05)	12.58 (2.15)	12.57 (2.52)	11.67 (2.45)
Country, % (n)						
U. S	88.65 (125)	88.89 (24)	87.88 (29)	92.59 (25)	96.30 (26)	81.48 (22)
Others	11.35 (16)	11.11(3)	12.12 (4)	7.41 (2)	3.70 (1)	18.52 (5)
Socioeconomic Status, % (n)						
Lower-middle	61.70 (87)	30.56 (11)	64.71 (23)	58.06 (18)	41.18 (14)	60 (21)
Middle	2.84 (4)	2.78 (1)	2.94 (1)	3.23 (1)	3.23 (1)	0 (0)
Upper-middle	31.91 (45)	36.11 (11)	23.53 (9)	16.13 (5)	32.35 (11)	17.14 (6)
Upper class	2.13 (3)	5.56 (2)	0 (0)	3.23 (1)	0 (0)	0 (0)

Notes. G1 = BE-Social; G2 = BE-Social without semi-immediate feedback; G3 = Cooperative learning; G4 = Self-management; G5 = Control group.

Table 11.2 Pairwise t Test Comparisons for Academic Achievement across Groups

	Group		Pairwise t tests			
	Mean	SD	t	df	p	Hedges g
	[98.75% CI]					
G1: BE-Social						
Time 1	13.22	3.48	1.67	1, 52	.102	0.45 [-0.23, 1.12]
Time 2	16.04	2.67	5.38	1, 52	<.001	1.44 [0.68, 2.97]
Time 3	13.48	4.17	3.53	1, 52	<.001	0.95 [0.23, 1.65]
G2: BE-Social*						
Time 1	12.07	3.09	0.49	1, 52	.63	0.13 [-0.54, 0.80]
Time 2	14.86	3.23	3.67	1, 52	<.001	0.98 [0.27, 1.69]
Time 3	11.99	4.08	2.29	1, 52	.03	0.61 [-0.08, 1.29]
G3: Cooperative learning						
Time 1	12.76	2.87	1.38	1, 58	.174	0.35 [-0.29, 0.99]
Time 2	13.55	2.88	2.48	1, 58	.016	0.64 [-0.02, 1.29]
Time 3	11.90	3.71	2.44	1, 58	.018	0.63 [-0.03, 1.28]
G4: Self-Management						
Time 1	12.21	3.09	0.65	1, 52	.52	0.17 [-0.49, 0.84]
Time 2	13.02	3.15	1.61	1, 52	.103	0.45 [-0.24, 1.12]
Time 3	11.15	3.42	1.69	1, 52	.097	0.45 [-0.23, 1.13]
G5: Control						
Time 1	11.63	3.52				
Time 2	11.53	3.43				
Time 3	9.32	4.49				

Note. All pairwise comparisons against control group. Bonferroni adjusted confidence intervals. CI = confidence interval. * BE-Social group without semi-immediate feedback.

Time 1 = Baseline; Time 2 = Intervention; Time 3 = Post-intervention.

Table 11.3 Linear Mixed Effects Model for Combined Engagement

Fixed effects	F	df	p
(Intercept)	2127.37	1, 799.66	<.001
Group	19.69	4, 628.11	<.001
Time	51.79	2, 287.99	<.001
Post type	36.28	2, 858.97	<.001
Group by Post type	8.48	8, 880.13	<.001
Pairwise comparisons, <i>i</i> - <i>j</i>		df	p
BE-Social - Control	7.78±2.51	1, 909.92	.008
BE-Social* - Control	7.66±2.54	1, 906.58	.010
Cooperative learning - Control	-0.16±2.66	1, 911.03	1.000
Self-management - Control	-4.09±2.46	1, 889.42	.389
Treatment - Baseline	1.11±0.91	1, 308.16	.449
Posttreatment - Baseline	-8.46±0.91	1, 233.41	<.001
Text - Visual	2.78±1.49	1, 870.87	.124
Text - Video	9.59±1.14	1, 854.24	<.001

Notes. = mean differences ± standard errors. * BE-Social group without semi-immediate feedback.

Figure 11.1 Participant Attrition Flowchart

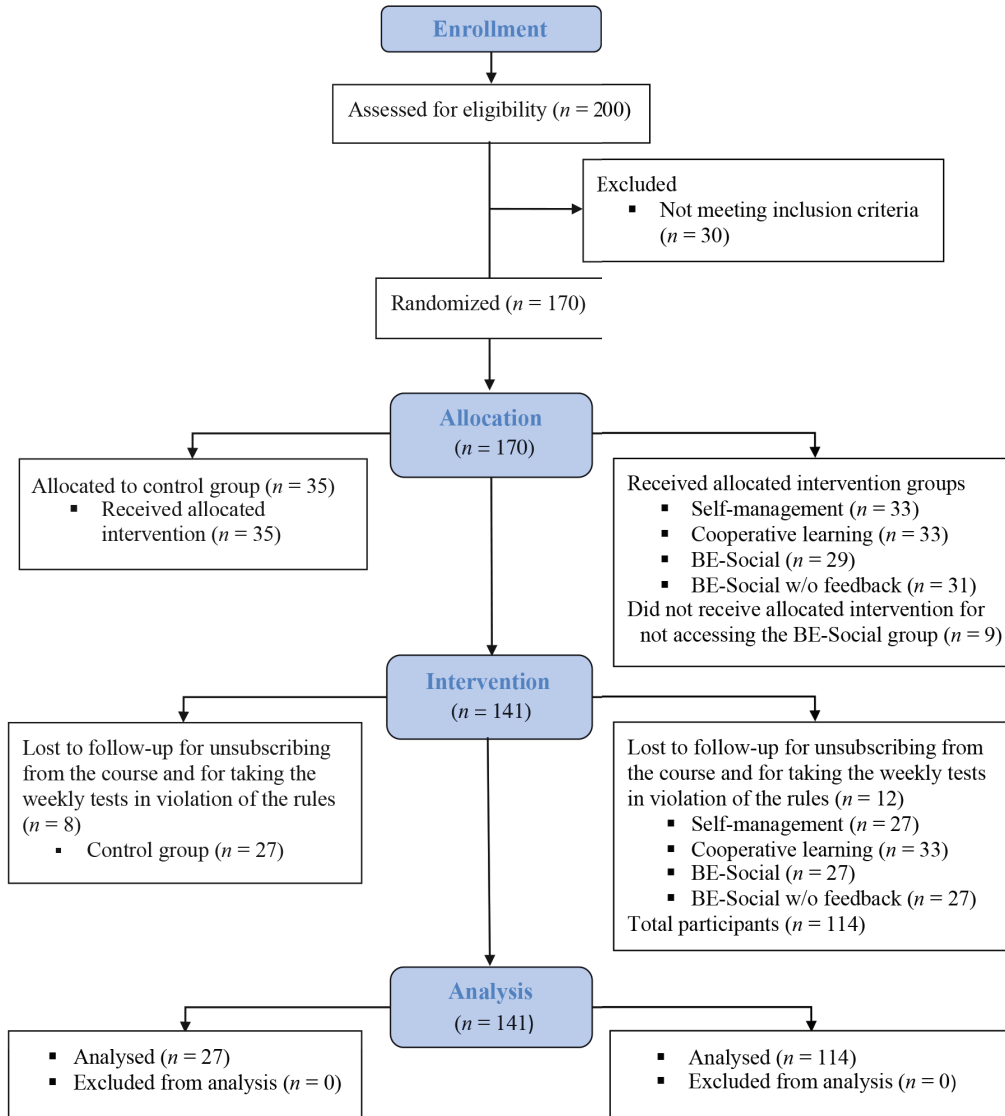


Figure 11.2 Changes in Academic Achievement during Pre-Intervention, Intervention, and Post-Intervention

Note. Time 1 = Baseline; Time 2 = Intervention; Time 3 = Post-intervention.

* BE-Social group without semi-immediate feedback.

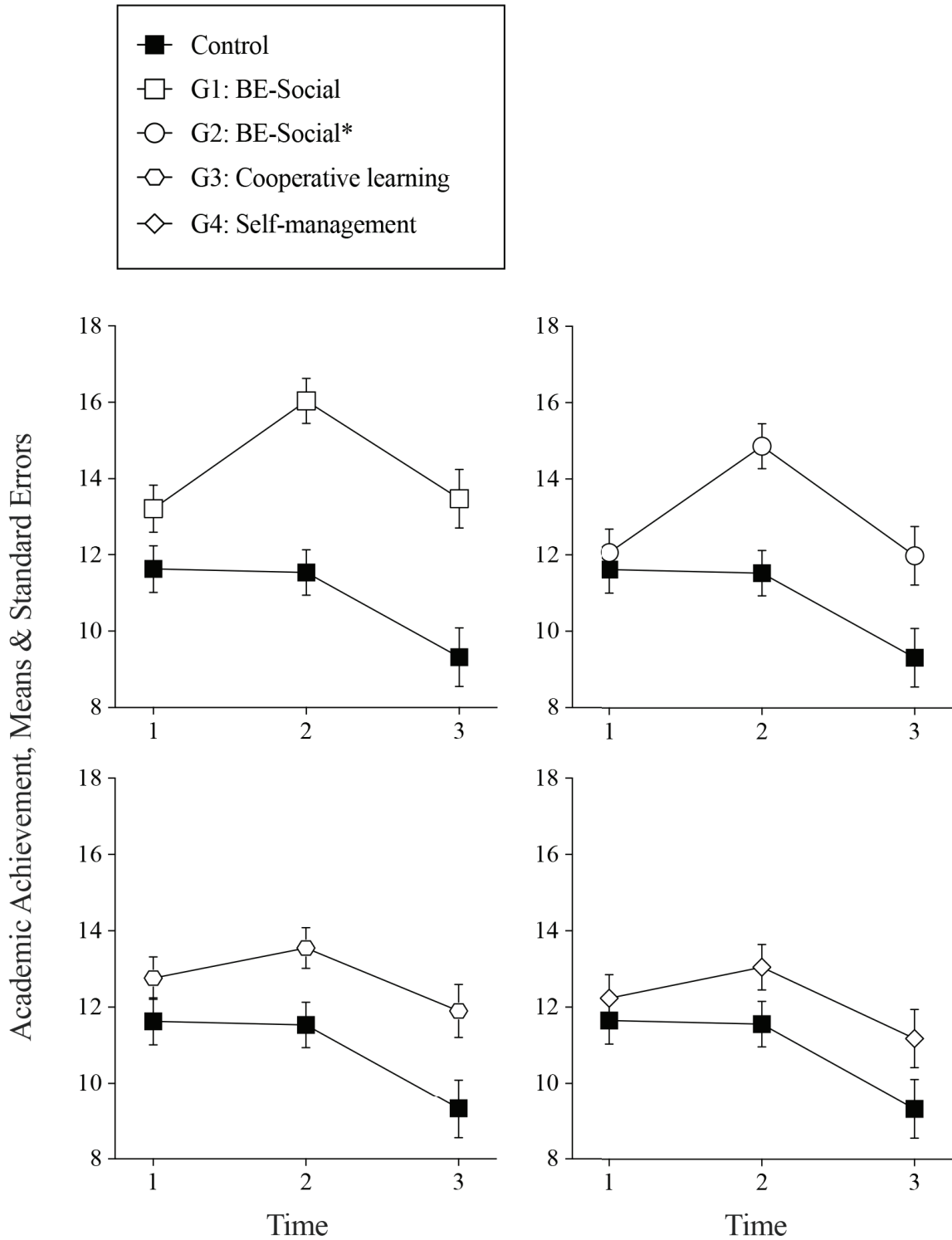
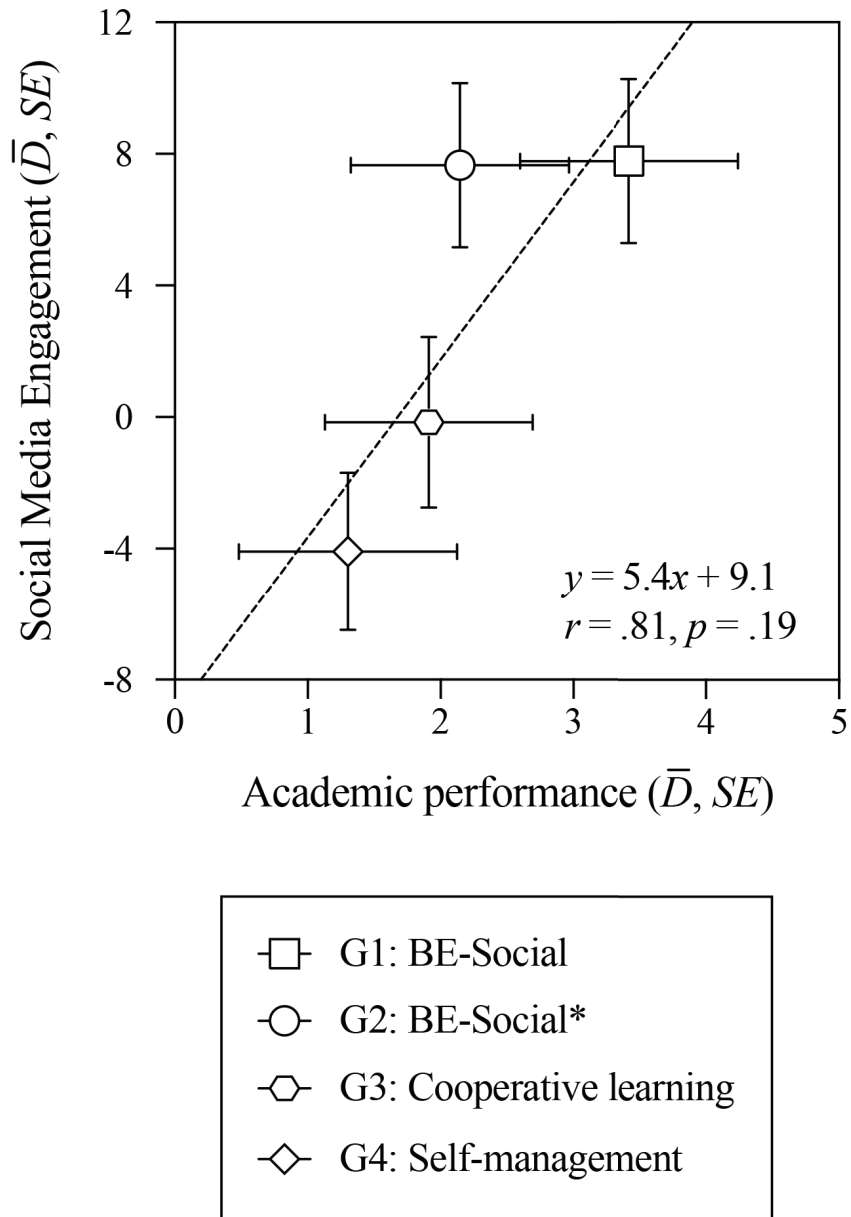


Figure 11.3 Academic Performance and Social Media Engagement across Intervention Groups

Notes. All mean differences and standard errors during the intervention phase. Broken line denotes linear regression. * BE-Social group without semi-immediate feedback.



QUINTA PARTE
DISCUSIÓN

CAPÍTULO 12. Discusión, limitaciones y líneas futuras

DISCUSIÓN

Este capítulo trata de destacar las principales conclusiones extraídas de los estudios presentados a lo largo de la presente tesis doctoral, examinando de forma crítica las consideraciones conceptuales y metodológicas. También, ofrece varias recomendaciones aplicadas, tanto para profesionales como para investigadores del área, a raíz de los resultados extraídos de los estudios metodológicos y aplicados. Finalmente, el capítulo recoge las conclusiones generales de este proyecto de investigación.

1. Fundamentos metodológicos

Nathan Azrin (1930-2013) fue un precursor del uso combinado de diseños de caso único y ensayos controlados aleatorizados. La primera investigación presentada en la tesis, el capítulo 6, presenta y argumenta claras ventajas de la combinación de ambas tradiciones metodológicas dentro del ámbito del análisis aplicado de conducta. Su uso permite identificar patrones idiosincráticos de los sujetos a lo largo del tiempo, tal y como hemos visto en el estudio presentado en el capítulo 10. También fomenta la comprensión de los efectos individuales de los procedimientos que forman una intervención multicomponente y facilita la aplicación de modelos de análisis jerárquico debido a la estructura multinivel de los diseños analítico-conductuales típicos, tal y como se ha comentado en detalle en el capítulo 11. La revisión realizada en el capítulo 6 confirma que la combinación de diseños de caso único y de grupo constituye una formulación práctica y válida para evaluar programas conductuales multicomponentes de intervenciones educativas mediadas por plataformas de redes sociales.

Utilizar sistemas de observación conductual rigurosos y válidos permite garantizar que los procedimientos aplicados se han realizado tal y como se describieron en el diseño inicial de intervención, con el fin de promover la integridad de los procedimientos de intervención. Al aplicar una intervención conductual de carácter multicomponente cabe el riesgo de que la intervención no sea aplicada como se ha descrito debido a procesos de deriva, olvido, o ausencia de conse-

cuencias correctivas por errores de procedimiento. En el caso del modelo propuesto, la presentación de varios elementos del procedimiento a través de publicaciones de vídeo recomendaba la validación del sistema de observación diseñado específicamente para este tipo de observación conductual. El capítulo 7 expone un estudio que valida una aplicación móvil como herramienta para registrar observaciones conductuales en contextos aplicados y así, garantizar la integridad de procedimiento y que fue utilizada extensamente a tal efecto en el estudio aplicado descrito en el capítulo 11.

El capítulo 8 consistió en realizar una revisión sistemática de 75 estudios recientes, con un total de 11.605 participantes de educación terciaria que utilizaban las redes sociales como canal de enseñanza-aprendizaje. La mayoría de investigaciones empleaban encuestas cualitativas y cuestionarios *ad hoc* y una minoría se centraba en utilizar métricas cuantitativas, siendo la frecuencia grupal, la medida común. Solo seis estudios empíricos evaluaron simultáneamente la interacción y la participación de los participantes y el rendimiento académico. Ambos sucesos, limitan conocer y demostrar si las respuestas discretas que generan las redes sociales sobre la interacción y la participación de los estudiantes en forma de publicaciones, comentarios y reacciones pueden funcionar como variables mediadoras para conocer los efectos en el rendimiento académico de estudiantes de posgrado. Este enfoque permitiría conocer si la cantidad y la estabilidad de publicaciones y comentarios relacionados con el temario correlacionaría con el incremento en pruebas objetivas de rendimiento. Este suceso sería probable debido a la retroalimentación semi-inmediata de los iguales o de los educadores, en dinámicas de aprendizaje cooperativo llevados a cabo en plataformas de redes sociales que permitan crear grupos privados. En cambio, observar un patrón de publicaciones inestables y limitados, podría sugerir una falta de retroalimentación semi-inmediata o ausencia de estrategias que promuevan el aprendizaje cooperativo. Por este motivo, el capítulo 8, además de presentar una revisión sistemática, propone una taxonomía de métricas cuantitativas relevantes basadas en el análisis de conducta. Algunas de las métricas no se han utilizado en investigaciones, pero consideramos que su aplicación podría facilitar futuros análisis de respuestas operantes generadas en las redes sociales y, así intentar correlacionar o demostrar respuestas relacionadas con el rendimiento académico.

Programa BE-Social: Efectos combinados y patrones idiosincráticos

Los capítulos 6, 7 y 8 tratan de ofrecer una perspectiva conceptual y metodológica para diseñar una intervención multicomponente y evaluar sus efectos en las respuestas de interacción y

participación ofrecida a través de plataformas de redes sociales. En el capítulo 10 se presenta un paquete de intervención analítico-conductual aplicado mediante grupos privados de Facebook a una muestra de 46 estudiantes en un curso de especialización profesional de posgrado. El paquete de intervención está compuesto por estrategias de autogestión con videomodelado, aprendizaje cooperativo y retroalimentación semi-inmediata, denominado programa BE-Social (Behavioral Education and Social Media). El objetivo del estudio fue evaluar el impacto de la intervención sobre las respuestas de interacción y participación de los estudiantes; y observar efectos inmediatos y demorados en rendimiento académico. Los procedimientos se aplicaron como una intervención única, es decir, el grupo experimental recibía una combinación de todas las estrategias que conformaban el paquete BE-Social y el grupo control carecía de acceso al programa. Los resultados generales indicaron un aumento considerable en los índices de respuestas de interacción y participación en la plataforma de la red social y en las pruebas objetivas semanales que evaluaban el rendimiento académico. Las ganancias de la intervención sobre la variable dependiente persistieron después de finalizar la fase experimental durante el periodo de seguimiento. Sin embargo, mediante un análisis de caso único, en concreto, un diseño de línea base múltiple, se observó que solo quince estudiantes respondieron favorablemente a la intervención BE-Social. Aproximadamente la mitad de los participantes mostraron un aumento de las respuestas de interacción y participación en el grupo privado además de obtener ganancias en el rendimiento académico durante el transcurso de la intervención. Así pues, el estudio presentado en el capítulo 10 de la presente tesis doctoral muestra que (a) el grupo cerrado de Facebook puede ser un canal eficaz para proporcionar una intervención conductual multicomponente, (b) los componentes que forman el programa BE-Social son compatibles con las funcionalidades que facilita Facebook, en concreto, los grupos privados, (c) la intervención BE-Social, modifica los patrones de respuestas de interacción y participación en un grupo privado de Facebook, (d) el uso de un diseño de caso único permite detectar patrones idiosincráticos entre los participantes y (e) se observa una correlación entre las respuestas de interacción y participación con los índices de rendimiento académico, aunque no puede aun aseverarse una relación mediadora de la interacción en red social.

Programa BE-Social: Evaluación específica de los componentes de la intervención

La evaluación y optimización del tratamiento es esencial para las prácticas basadas en la evidencia. La presente tesis doctoral ha enfatizado los análisis conceptuales y metodológicos necesarios para definir una intervención multicomponente y evaluar sus efectos preliminares sobre

el rendimiento académico y la conducta en redes sociales. Un paso más en el proceso científico aplicado es la optimización, que en el caso particular de las intervenciones multicomponentes no puede lograrse sin diseños de investigación complejos que incluyan análisis de componentes o estudios con múltiples grupos. Estos últimos enfoques permiten al investigador establecer la contribución relativa de los distintos elementos de un programa de intervención, lo que puede conducir a nuevos ajustes del programa de intervención. Por ejemplo, pueden omitirse ciertos componentes, mientras que otros pueden modificarse en forma o intensidad.

El capítulo 11 trata de replicar el estudio presentado en el capítulo 10 incorporando normas metodológicas adicionales: (1) una mayor muestra de participantes (141 estudiantes), (2) un número equilibrado entre el grupo control y los grupos experimentales, (3) un número equiparable entre las sesiones proporcionadas durante las fases del estudio (líneabase, intervención, post-tratamiento), (4) una intensidad comparable de presentación de contenido académico entre los grupos, control y experimental y (5) la evaluación individualizada de los componentes del programa BE-Social. Todas estas variaciones pretendían evaluar la validez incremental de los componentes de la intervención BE-Social y conocer qué procedimiento o conjunto de ellos, promueve las respuestas de interacción y participación en las redes sociales; además de verificar si existe correlación entre dichas respuestas y el rendimiento académico. En resumen, este segundo estudio de ensayo controlado aleatorizado (ECA) permitiría optimizar la intervención detectando posibles debilidades mediante la evaluación relativa de sus varios componentes al efecto combinado del programa que ya había sido evaluado en el estudio anterior. Para lograr una evaluación precisa, se empleó un ECA multi-grupo con cinco brazos.

Los resultados indicaron que los estudiantes expuestos al paquete de intervención completo, BE-Social con retroalimentación semi-inmediata mostraban mayor rendimiento académico. En cambio, en comparación con el estudio anterior (capítulo 10), los efectos de la intervención tienden a desvanecerse después de su retirada. Este hecho, sugiere que el factor crítico de la intervención es la retroalimentación semi-inmediata y debería de permanecer durante todo el curso para mantener resultados óptimos. Por otro lado, el grupo BE-Social sin feedback semi-inmediato obtuvo un índice elevado de respuestas de interacción y participación en el grupo privado de Facebook, pero el índice de rendimiento académico fue menor que la intervención con el programa completo (BE-Social con retroalimentación semi-inmediata). Este hallazgo sugiere que las respuestas de interacción y participación puede que no cumplan la función de variable media-

dora sobre el rendimiento académico que habíamos planteado como hipótesis. Un análisis mucho más molecular se hace necesario para determinar bajo qué circunstancias la interacción en red social es un buen predictor de los efectos del programa sobre rendimiento académico, como sí parece serlo cuando los componentes de automanejo y aprendizaje cooperativo fueron evaluados de forma aislada. Un hallazgo importante de este estudio es el aparente valor aditivo de los componentes que integran el programa BE-Social. En particular, observamos que las estrategias que más contribuyen a la efectividad del programa completo son el aprendizaje cooperativo y la retroalimentación. En cambio, el procedimiento de autogestión con videomodelado no obtuvo efectos significativos durante la fase de intervención. Curiosamente, durante el análisis de los resultados, se observaron mayores respuestas de interacción y participación en las publicaciones textuales, en lugar de las publicaciones mixtas (texto con imágenes) o las transmisiones en video.

La amplia evaluación del programa presentada, y que resume la aplicación del mismo con la participación de casi dos centenares de participantes, sugiere que la eficacia del mismo es significativa, su viabilidad a través del canal de las redes sociales (concretamente un grupo cerrado de Facebook) se adaptada adecuadamente a las funcionalidades de comunicación e interacción disponibles en el mismo, y aporta un análisis muy aproximado a la eficacia relativa y sinergias entre los componentes del programa.

Implicaciones conceptuales

Uno de los objetivos principales del estudio del Capítulo 11 era el de analizar los diferentes componentes del programa BE-Social por separado con el fin de establecer la contribución de cada uno de los componentes del programa y la posibilidad de que existieran efectos sinérgicos entre ellos. Pese a que las características de nuestro estudio (comparaciones entre grupos en lugar de análisis individuales) no permiten verificar directamente la hipótesis de que los diferentes componentes del programa se ajustan a un modelo de efectos aditivos (a tal efecto un análisis intra-sujeto o de caso único hubiera sido más concluyente), la suma del efecto encontrado para cada uno de los elementos del programa BE-Social por separado, es decir, estrategias de autogestión con videomodelado, aprendizaje cooperativo y retroalimentación semi-inmediata, se aproxima al efecto del programa completo. Esto sugiere que el programa BE-Social está compuesto por diferentes componentes que operan bajo procesos operantes relativamente independientes y que contribuyen de forma aditiva al efecto global del programa.

En concreto, el componente de autogestión con videomodelado sería una intervención basada en antecedentes. El objetivo de este componente es que los estudiantes realizaran las conductas modeladas por los experimentadores en los videos para así poder contactar con las consecuencias naturales de la autogestión (p.ej., aumento del tiempo libre, aumento del tiempo efectivo de estudio, reducción del tiempo dedicado a actividades distractoras que interfieren con el tiempo de estudio), las cuales en muchos casos eran especificadas en los videos. Sin embargo, este componente por sí solo no tuvo un efecto estadísticamente significativo sobre el rendimiento académico de los estudiantes.

El componente de aprendizaje colaborativo se trataría de una intervención basada en antecedentes orientada hacia la creación de un contexto en el que la tasa de reforzamiento social mediada por terceros, es decir, los otros estudiantes del grupo, fuese elevada. El propósito de este componente era el de incrementar el nivel de interacción y participación de los estudiantes, sobre la base de que el aumento de la participación mejoraría el rendimiento académico de los mismos (véase, por ejemplo, Alshuabi et al., 2018, Arteaga et. al., 2014, Graham, 2014). Recientemente, Lindström et al., (2021) demostraron que la participación en redes sociales puede ser explicada mediante un modelo operante. Concretamente, demostraron que el intervalo entre publicaciones, cuando dicha conducta está bajo una operante libre, es función del reforzamiento social, operativizado en el estudio como el número de comentarios y reacciones recibidos por la conducta de publicar en la red social. Por tanto, favorecer un contexto con una alta tasa de reforzamiento social mediado por terceros debería de aumentar la participación de los estudiantes y, con ello, su propio rendimiento académico, sin requerir que el propio instructor se encargue de administrar los reforzadores sociales, algo que ya ha sido utilizado en otros modelos de educación conductual como CAPSI (Martin et al., 2002b).

Por último, el componente de retroalimentación semi-inmediata sería una intervención basada en el proceso de reforzamiento social, mediado por el propio instructor. Este componente podría estar contribuyendo al efecto de la intervención sobre el rendimiento académico de varias formas. En primer lugar, es posible que la retroalimentación del instructor actúe como reforzador por la mera participación en el grupo de Facebook de los propios estudiantes. Sin embargo, en nuestro estudio, jamás se administró retroalimentación semi-inmediata en ausencia del componente de aprendizaje cooperativo, por lo que el efecto del feedback semi-inmediato del instructor sobre la participación de los estudiantes podría haber quedado ensombrecido por

los efectos del componente de aprendizaje cooperativo con retroalimentación demorada, explicando así el hecho de que la participación de los estudiantes en el grupo apenas incrementase. Otra posibilidad que parece ajustarse más a los resultados encontrados en nuestro estudio es que la retroalimentación del instructor esté reforzando diferencialmente los comentarios de los estudiantes que resultan más acertados y acordes a los contenidos del curso. De esta forma, más que un gran cambio en la participación total de los estudiantes, cabría esperar una mejoría en el nivel de acierto y calidad de los comentarios, redundando así en un aumento del rendimiento académico. Desgraciadamente, en nuestro estudio no se realizó un análisis pormenorizado del contenido de las aportaciones de los estudiantes que permita corroborar la hipótesis de que la retroalimentación del instructor opera reforzando diferencialmente las aportaciones de las publicaciones, pero el hecho de que aumentase el rendimiento académico en ausencia de un aumento en la participación en el grupo con retroalimentación semi-inmediata parece sugerir un proceso de estas características.

En cualquier caso, atendiendo tanto a los resultados del estudio del Capítulo 10, como a los del estudio del Capítulo 11, el componente que parece tener un mayor impacto sobre el rendimiento académico es la retroalimentación semi-inmediata. Nótese que el grupo 2 del estudio del Capítulo 11 estaba sometido a un programa idéntico al programa BE-Social, con la única excepción de que la retroalimentación se administraba semanalmente en lugar de cada 24 horas, por lo que la diferencia en el efecto sobre el rendimiento entre ambos grupos debe ser atribuida completamente a la inmediatez de la retroalimentación, lo que nos da una idea de la importancia de este parámetro.

En este sentido, resulta particularmente relevante la literatura básica sobre el fenómeno que ha sido denominado descuento por demora⁵⁵. Este fenómeno puede definirse como la disminución de la tasa de respuesta ocasionada por la demora del reforzador (véase, por ejemplo, Lattal, 2010, para una revisión de la literatura básica). En animales de laboratorio es fácil apreciar como una demora de apenas unos pocos segundos ocasiona reducciones muy marcadas o incluso completas en la tasa de respuesta (p.ej., Perin, 1943). Si bien el ser humano no es extremadamente sensible, como sí ocurre con animales no humanos, a la demora del reforzador, lo cierto es que el fenómeno de descuento por demora también se aprecia muy claramente en seres humanos (véase Reynolds, 2006 para una revisión). Esto parece explicar los resultados de

55 *Delay discounting*, en la literatura original.

nuestro estudio, en los que la demora de la retroalimentación del instructor fue el factor que pareció tener un mayor impacto sobre el rendimiento académico de los estudiantes. Hay que tener en cuenta que pese a que hemos utilizado el término “semi-inmediato” para describir la demora del feedback en nuestro estudio, en realidad existía la posibilidad de que la retroalimentación se demorase hasta 24 horas en el primer ECA y de una hora de demora como media en el segundo ECA. Es probable que, en línea con la investigación básica, el efecto sobre el rendimiento académico del programa BE-Social hubiera sido sensiblemente superior utilizando una retroalimentación más inmediata, pese a que ello quizá no hubiera sido práctico de cara a establecer la validez ecológica del programa.

Limitaciones y líneas futuras

El estudio del capítulo 7, Accuracy of paper-and-pencil systematic observation versus computer-aided systems valida el programa Big Eye Observer como sistema preciso y fiable para registrar conductas mediante observación directa. En futuros análisis, sería interesante replicar el estudio con conductas de diferentes topografías, extensiones y operacionalmente complejas para así, evaluar experimentalmente su potencial. Uno de los hallazgos que informa este estudio, es que observadores sin experiencia previa en el uso de registro de conductas observables pueden adquirir las habilidades con recursos de observación en contextos naturales y generar registros fiables. El estudio se desarrolló con observadores recién entrenados y los resultados sugirieron que la observación con papel y lápiz obtuvo mayor nivel de precisión que los sistemas computarizados. Este resultado podría deberse a la escasa experiencia de los observadores al registrar eventos naturales. En futuros estudios, podría replicarse con participantes hábiles en el registro de eventos naturales tanto con recursos tecnológicos como con papel y lápiz. Otra hipótesis, referente a los resultados obtenidos, podría ser que el acceso al papel y al lápiz es más común y sencillo, en comparación de la disposición de sistemas tecnológicos complejos. Por lo general, en las aplicaciones Observer TX y Big Eye Observer, para acceder a la pantalla de registro, primero el observador ha de configurar pantallas previas con información relevante de la conducta, por ejemplo, la duración de la sesión, el tipo de registro que se desea aplicar, la creación de etiquetas para las conductas a observar, etc. Todos estos pasos, requieren de un esfuerzo por parte del observador. En futuras investigaciones podría emplearse un entrenamiento en habilidades conductuales a los observadores con el fin de reducir el tiempo y el esfuerzo del observador. En relación al concepto de esfuerzo, utilizar sistemas de registros asistidos por ordenador, obliga

al observador a disponer de un equipo de grabación, grabar las sesiones en video, transferir los videos a un ordenador, eliminar los clips de video repetidos, pausar los videos y registrar. Quizá, para futuros estudios, la cantidad de tiempo dedicado a estas respuestas podría adjudicarse a otras personas del equipo, como técnicos, ayudantes de investigación o estudiantes universitarios que todavía cursan sus estudios, y así tratar de equilibrar la carga laboral y centrar a los observadores en su tarea.

La mayoría de la bibliografía empleada en la revisión sistemática del capítulo 8 *Quantitative indices of student social media engagement in tertiary education: a systematic review and a taxonomy* utilizaban cuestionarios y encuestas *ad hoc* para medir las respuestas de interacción y participación de los estudiantes de forma transversal. Por otro lado, los estudios que utilizaban medidas cuantitativas estaban basados en cuestionarios *ad hoc* de tipo Likert y solo pocos estudios utilizaron medidas derivadas de la observación conductual, por ejemplo: la frecuencia de publicación, la latencia de respuesta de los estudiantes al comentario del educador o el tiempo entre respuesta entre los comentarios de un estudiante específico. También comentar que un tercio de las investigaciones revisadas se realizaron en contextos de aprendizaje de un segundo idioma. Este hecho puede limitar la diversidad y los resultados obtenidos en esta revisión; ya que es diferente la dinámica y la motivación de los estudiantes en cursos para aprender segundos idiomas que un curso académico reglado. Cabe resaltar que tan solo seis estudios investigaron simultáneamente en el estudio de las respuestas de interacción y participación con el rendimiento académico. Por último, ninguno de los estudios revisados utilizó un ECA o un grupo control que equiparara la intensidad de interacción o entrega del educador, en relación al grupo experimental. Normalmente, los grupos control consistían en la ausencia de acceso o participación al uso de las redes sociales con fines educativos. En futuras investigaciones, sería interesante replicar la metodología de búsqueda hasta la fecha actual, ya que el estudio se inició en octubre del 2020, en las primeras fases de la pandemia mundial COVID-19. Para evitar la interrupción de la educación durante largos meses, la mayoría de instituciones educativas de primaria, secundaria y universidades se adaptaron y comenzaron a aplicar softwares educativos y plataformas de redes sociales, compatibles con la entrega en línea de estrategias de enseñanza. Una posible búsqueda hasta la fecha, modificaría el número total de estudios publicados y probablemente la cantidad de estudios que emplean metodologías cuantitativas.

El estudio del capítulo 10, *Effect of the behavioral education and Social Media (BE-Social) program on postgraduate academic performance: A randomized controlled trial* presenta varias limitaciones. En primer lugar, la muestra estaba sesgada en cuanto al género femenino. En futuros estudios es importante alcanzar un equilibrio en la distribución sociodemográfica de los participantes. En segundo lugar, la cantidad de exposición del contenido no fue equiparable en los dos grupos (control y experimental). Para futuras investigaciones es necesario establecer un diseño de investigación que equipare la intensidad de estímulos (oportunidades de interacción con los estudiantes, número de estrategias educativas) que se van a presentar los grupos experimentales y de control. En tercer lugar, la eficacia del programa BE-Social queda enmascarada, al tratarse de una intervención formada por múltiples componentes. En futuras réplicas sería necesario realizar un análisis de componentes y discernir qué procedimiento o combinación es eficaz y suficiente para provocar cambios en el rendimiento académico. En cuarto lugar, no se realizó un análisis experimental de las respuestas de interacción y participación en las redes sociales (publicaciones, comentarios y reacciones) sobre la posible función que podrían desempeñar como variables mediadoras en relación al rendimiento académico. Los comentarios podrían informar del nivel de compromiso y motivación con el contenido curricular, respuestas que podrían operacionalizarse y medirse a través de métricas cuantitativas, tal y como se propone en la taxonomía del estudio *Quantitative indices of student social media engagement in tertiary education: a systematic review and a taxonomy*. En futuros estudios, se podrían replicar los resultados incluyendo nuevas métricas cuantitativas extraídas de la taxonomía propuesta en el estudio de la revisión sistemática. Por otro lado, las reacciones podrían comunicar sentimientos, cómo perciben los estudiantes los conocimientos que se imparten. No obstante, ello requeriría de análisis más complejos, también propuesto en la taxonomía de métricas operantes.

El capítulo 11 presenta el estudio *Enhancing academic performance in tertiary education through Social Media: A multiple-armed RCT*. A pesar de elaborarse con fines de resolver una gran cantidad de limitaciones del estudio anterior, cabe resaltar las limitaciones que presenta. Los datos obtenidos sobre las respuestas de interacción y participación en los grupos privados de Facebook (publicaciones, comentarios, reacciones) son grupales. La estrategia de recopilación de datos no permitía identificar los miembros individualmente. Esto implica que tan solo conocemos el impacto de la intervención a nivel grupal y no individual. En futuras investigaciones habría que diseñar desde el inicio un plan de integridad de procedimiento que incluya tomar datos individuales de cada respuesta de interacción y participación. Incluso, se podría entrenar a los propios par-

participantes a autorregistrar sus propias respuestas. Por otro lado, en relación al deterioro aparente del grupo control en la fase de post-tratamiento, puede atribuirse a la dificultad del contenido académico o bien a la selección de los alumnos, cabe recordar que el estudio se llevó a cabo con estudiantes que habían presentado bajo rendimiento en las pruebas semanales objetivas durante los primeros cuatro meses de curso. Es interesante observar que, en todos los grupos experimentales, indiferentemente de la cantidad de ganancias, no se observó dicho declive, invitando a hipotetizar que el programa BE-Social tiende a provocar mejoras en el rendimiento académico. Respecto a la naturaleza de las pruebas objetivas semanales, utilizadas como medidas objetivas para registrar el rendimiento académico, hay que resaltar que se tratan de pruebas *ad hoc*, esto implica que carecen de supervisión o de una validación estándar. No obstante, su uso facilitó la obtención de medidas frecuentes a lo largo de todas las fases. Es importante recordar que, los resultados obtenidos en dichas pruebas no se verían reflejados en el expediente académico, esto elimina cualquier factor de deshonestidad académica.

Este último estudio experimental-aplicado combina estrategias basadas en la evidencia y plataformas de redes sociales. Estudios futuros pueden replicar los resultados de esta investigación utilizando combinaciones alternativas. Por ejemplo, la intervención basada en estrategias de autogestión podría mejorarse añadiendo retroalimentación basada en el desempeño. Es decir, tras la observación (mediante una grabación o un video en directo) del desempeño del estudiante, el educador puede proporcionar retroalimentación correctiva y positiva. También podría replicarse este estudio, añadiendo retroalimentación inmediata del educador al componente de aprendizaje cooperativo. El procedimiento a seguir sería exactamente el mismo al descrito en el estudio, pero en este caso, el educador ofrecería retroalimentación inmediata a cada pregunta de los estudios. Podría incrementar las ganancias, pero habría que considerar si la estrategia de aprendizaje cooperativo con este cambio, seguiría ofreciendo un espacio de reflexión y pensamiento crítico de los estudiantes. Otro aspecto que sería conveniente considerar para futuros estudios, es la posibilidad de ofrecer retroalimentación inmediata, y no semi-inmediata. La razón por la cual se determinó el uso de fijar un periodo de latencia entre la pregunta de cualquier estudiante y la respuesta del educador fue debido a la incompatibilidad de estar presente en un lugar físico y en línea en una red social. Es más probable que el educador esté en línea durante un tiempo determinado, previamente establecido con los estudiantes, que no durante todo un día.

Recomendaciones prácticas

El programa de intervención conductual BE-Social ha demostrado ser una intervención eficiente en términos de coste (tiempo)-beneficio. La aplicación del programa en el primer estudio experimental-aplicado dio lugar a un incremento de 20 puntos de media en el rendimiento académico. Un análisis somero de costes indica que cada estudiante necesita semanalmente, dos horas aproximadamente de tiempo del educador para recibir la intervención. A raíz de los resultados del segundo estudio experimental-aplicado, es posible reducir el tiempo del educador y mantener las ganancias en el rendimiento académico aplicando el programa BE-Social con retroalimentación semi-inmediata empleando grupos privados de Facebook, como red social. Los resultados obtenidos en ambos estudios permiten afirmar que se trata de una intervención con un importante potencial de transferencia de conocimientos. Las instituciones educativas, los decisores sociales, los propios educadores pueden beneficiarse del bajo coste que comportaría integrar algunas plataformas de redes sociales en cursos de formación terciaria. Además, las estrategias conductuales que forman la intervención BE-Social (estrategias de autogestión con videomodelado, aprendizaje cooperativo y retroalimentación semi-inmediata) son fáciles de adquirir y aplicar en contextos naturales y en línea, siempre y cuando los educadores reciban una formación específica.

Por otro lado, los resultados del primer estudio experimental-aplicado, expresados tanto a través de un diseño de caso único como de grupo, mostraron un rápido incremento en el rendimiento académico una vez iniciada la intervención. Posteriormente, el segundo estudio experimental-aplicado muestra un incremento en el rendimiento académico justo después de introducir la intervención BE-Social con reforzamiento semi-inmediato. Alcanzar un incremento óptimo en tan poco tiempo es compatible con la duración de cualquier curso de posgrado distribuido en trimestres o semestres.

En concreto, en ambos estudios experimentales, el programa BE-Social se diseñó con el fin de incrementar las respuestas de interacción y participación generadas en la plataforma de red social. De esta forma, mediante sistemas de registro de observación directa era sencillo cuantificar las respuestas objetivo y obtener datos que pudieran utilizarse más tarde, para compararlos con los datos del rendimiento académico. Aparentemente, los resultados del primer estudio sugirieron que podía existir una correlación entre ambos índices (respuestas de interacción y participación en las redes sociales y rendimiento académico). Sin embargo, en su replicación presen-

tada en el capítulo 11, no se aprecia correlación entre ambas respuestas. Ya que, los estudiantes que recibieron la intervención BE-Social sin retroalimentación semi-inmediata mostraron índices superiores de respuestas de participación e interacción en la red social en comparación a los estudiantes que recibieron el mismo programa, pero con retroalimentación inmediata y los índices de rendimiento académico fueron más elevados en el programa BE-Social con retroalimentación inmediata. Por lo tanto, no podemos afirmar que ambas respuestas correlacionan. Así pues, para futuras investigaciones, sería interesante diseñar programas con estrategias conductuales focalizadas en el incremento de rendimiento académico, sin contemplar variables mediadoras como las respuestas de interacción y participación.

En relación a los resultados previamente comentados, sería más eficiente aplicar la intervención BE-Social a través de grupos privados de Facebook con estudiantes de posgrado, con el objetivo de alcanzar índices superiores de rendimiento académico. Esto implica que el educador entregue estrategias de autogestión y videomodelado, aprendizaje cooperativo y, como elemento esencial, que proporcione retroalimentación semi-inmediata. Para evitar altos costes de tiempo al educador, sería conveniente determinar un periodo de tiempo dónde el educador y los alumnos se conecten en línea, de esta forma, facilitaría la entrega de retroalimentación con pocos segundos de demora desde que un estudiante solicita su ayuda. También, referente al componente de aprendizaje cooperativo, con el fin de agilizar la dinámica participativa y comunicativa entre el alumnado y que dicha tarea no recaiga en su totalidad en el educador, podría adaptarse al sistema CAPSI (Martin et al., 2002b). Es decir, los estudiantes que dominen un área o que hayan finalizado un módulo de contenido académico, pueden ejercer el rol de supervisor a un par de sus compañeros con el propósito de resolver sus dudas. Esta dinámica ha demostrado ser beneficiosa para ambos roles (estudiante y estudiante-supervisor).

También es importante mencionar que la intervención BE-Social se ha diseñado operatizando las estrategias de autogestión con videomodelado y aprendizaje cooperativo, de forma concreta. Esto no excluye otras alternativas para aplicar ambos procedimientos. Por ejemplo, tal y como han sugerido los resultados del último estudio experimental, las estrategias de autogestión con videomodelado no provocaron efectos aparentes durante la fase de intervención, en comparación con el resto de fases de línea base. Una posible aplicación alternativa de este componente sería ofrecer retroalimentación basado en el desempeño del estudiante y no la realización de la tarea en sí. Asimismo, ocurre con el procedimiento de aprendizaje cooperativo, el educador po-

dría seleccionar a estudiantes más avanzados y ofrecerles la tarea de proporcionar retroalimentación a un par de compañeros durante unas semanas, siempre y cuando soliciten ayuda. De este modo, el educador y los alumnos seleccionados, durante esas semanas, distribuirían la función de entregar retroalimentación al resto de estudiantes.

Cabe destacar que los resultados obtenidos sobre la validez social de la intervención propuesta en ambos estudios son positivos. Ya que, de forma espontánea, la mayoría de estudiantes enviaron comentarios positivos el último día del estudio. Esto sugiere que proporcionar la intervención BE-Social mediante grupos privados de Facebook, fomenta un clima de participación y motivación entre iguales y crea vínculos positivos entre los educadores y el alumnado. No obstante, no se realizó una evaluación de validez social formal constituyendo ello un elemento a considerar en futuros estudios en este ámbito (ver diferentes enfoques de evaluación de escalas de validez social en Anderson et al., 2021).

Después de obtener resultados positivos tras realizar dos estudios experimentales-aplicados sobre la intervención conductual BE-Social, podemos confirmar que la intervención con todos los componentes (estrategias de autogestión con videomodelado, el aprendizaje cooperativo y la retroalimentación semi-inmediata) es exitosa con estudiantes que presentan un perfil bajo de rendimiento. Este fenómeno adquiere importancia cuando, en algunos casos, los estudiantes que presentan dificultades de aprendizaje tienden a desarrollar problemas de conducta (p.ej., agresividad, adicciones) (Crosnoe, 2006) o bien, deciden abandonar los estudios por falta de apoyo de las instituciones educativas (Au-Yong-Oliveira et al., 2017).

Se ha demostrado que la intervención BE-Social es efectiva para estudiantes de posgrado con bajo rendimiento durante la línea base. Los resultados obtenidos permiten realizar una recomendación preliminar del programa en este ámbito, quizá dirigida a este perfil. Los resultados obtenidos hacen posible recomendar el uso de plataformas de redes sociales, en concreto, grupos privados de Facebook, como canal para proporcionar una intervención conductual que requiera de interacciones frecuentes (síncronas o asíncronas) y de retroalimentación inmediata o semi-inmediata en educación terciaria. La presente colección de estudios son los primeros en la literatura en utilizar metodología experimental, tanto de caso único como de grupo, para evaluar una intervención educativa mediante red social.

A continuación, se detallan las recomendaciones prácticas argumentadas anteriormente:

1. El programa BE-Social se adapta a las funcionalidades de los grupos cerrados de Facebook, como canal de interacción y participación.
2. Instituciones educativas y educadores sin experiencia previa en análisis de conducta, pueden adquirir con relativa facilidad los procedimientos analítico-conductuales de la intervención BE-Social mediante formación específica.
3. El paquete de intervención propuesto alcanza incrementos óptimos en el rendimiento académico de los estudiantes de forma inmediata o transcurridos pocos días de iniciarse la aplicación.
4. Nuestras observaciones sugieren que en un subgrupo considerable de estudiantes la intervención pierde su efecto una vez es retirada, lo que hace recomendable su aplicación continuada durante toda la duración del curso o actividad académica de interés.
5. Nuestros estudios sugieren que, aunque el efecto de la intervención es considerable, existe un patrón idiosincrático que hace recomendable la evaluación de las necesidades individualizadas de estudiantes que no responden a la intervención.
6. El paquete de intervención propuesto es compatible con el calendario académico, distribuido en trimestres o semestres, debido a su rápido efecto en el rendimiento académico.
7. Entidades educativas y educadores pueden aplicar el programa BE-Social con estudiantes de bajo rendimiento; los estudios empíricos presentados validan su aplicación entre estudiantes con dicho perfil.
8. El componente esencial de la intervención BE-Social, la retroalimentación semi-inmediata, puede adaptarse a fin de su aplicación práctica, por ejemplo, determinando horarios específicos o señales que indiquen que el instructor está conectado o disponible.
9. Los programas educativos de posgrado pueden utilizar el programa BE-Social para fomentar vínculos positivos entre docentes y estudiantes, creando así un clima positivo de motivación y participación.
10. Instituciones educativas de educación terciaria pueden incorporar grupos cerrados de Facebook como canal para proporcionar una intervención conductual que requiera de respuestas de interacción y participación entre educadores y estudiantes.

CONCLUSIONES GENERALES

La presente tesis doctoral y la colección de estudios que incorpora permite establecer un conjunto de conclusiones preliminares sobre la utilización de intervenciones conductuales educativas mediante red social en educación terciaria que se presentan de forma resumida a continuación:

1. Las funcionalidades de plataformas de red social en su uso en educación terciaria son compatibles con el uso de estrategias de educación conductual basadas en la evidencia y en particular el entrenamiento de habilidades de autogestión de la conducta de estudio, el diseño de un ambiente de aprendizaje cooperativo y la presentación de feedback semi-inmediato por parte del educador.
2. Los elementos de intervención conductual educativa mencionados se integraron en una intervención multicomponente (programa BE-Social) evaluada mediante dos estudios ECA observándose incrementos en rendimiento académico estadística y socialmente significativos.
3. Los componentes del programa BE-Social tienen validez incremental y posiblemente efectos aditivos sobre rendimiento académico. Ello sugiere que la intervención puede estar operando a través de procesos operantes independientes. Los análisis de carácter correlacional presentados sugieren que indicadores cuantitativos de interacción (p.ej., reacciones, comentarios, visualizaciones) pueden, bajo ciertas circunstancias, ser buenos indicadores del eventual efecto sobre el rendimiento académico de la intervención.
4. El feedback semi-inmediato por parte del educador parece ser el elemento de la intervención con un impacto más claramente discernible sobre el rendimiento académico, lo cual es congruente con la amplia literatura existente sobre los efectos del feedback en el aprendizaje, así como con los estudios del efecto del descuento de demora en el aprendizaje operante.
5. Dado que se han detectado efectos parcialmente transitorios, especialmente en el segundo estudio ECA, la optimización del efecto del programa requiere de una aplicación continua durante la duración del curso o actividad educativa de interés.

6. Es posible que en torno a un tercio de estudiantes no responda al modelo de intervención propuesto por lo que actividades de tutorización individualizadas, así como la evaluación de nuevos componentes (o cambios en la intensidad de los elementos de intervención propuestos) se haga necesaria a fin de alcanzar efectos óptimos.

GENERAL CONCLUSIONS

The current doctoral dissertation allows us to establish a set of preliminary conclusions on behavioral educational interventions through a social media in tertiary education:

1. The functionalities of social media platforms are compatible with the use of evidence-based behavioral educational strategies and, in particular, with training of study self-management skills, designing a cooperative learning environment, and the delivering instructor semi-immediate feedback.
2. The above behavioral educational intervention elements were integrated into a multicomponent intervention (BE-Social program) evaluated through two RCT studies showing statistically and socially significant increases in academic achievement.
3. The components of the BE-Social program have incremental validity and possibly additive effects on academic achievement. This suggests that the intervention may be operating through independent operant processes. The correlational analyses presented herein suggest that quantitative indicators of social media engagement (e.g., reactions, comments, visualizations) may, under certain circumstances, be good indicators of the eventual effect of the intervention on academic achievement.
4. Instructor semi-immediate feedback appears to be the element of the intervention with a distinct impact on academic performance, which is in line with the vast body of literature on the effect of feedback on learning and also with the human operant literature on delay discounting.
5. Given that somewhat transient effects observed, especially in the second RCT study, optimizing the effect of the program requires continuous application for the duration of the course or educational activity of interest.
6. Approximately one third of students may not respond to the proposed program. Therefore, individualized tutoring strategies, as well as the evaluation of new components (or changes in the intensity of the proposed intervention elements) may be necessary in order to optimize the effects of the BE-Social program.

REFERENCES

- Acar, C., Tekin-Iftar, E., & Yikmis, A. (2017). Effects of mother-delivered social stories and video modeling in teaching social skills to children with autism spectrum disorders. *Journal of Special Education, 50*(4), 215-226.
- Adams, G. L. & Engelmann, S. (1996). *Research on Direct Instruction: 25 Years beyond Distar*. Seattle: Educational Achievement Systems.
- Agnew, J. L. (1998). The establishing operation in organizational behavior management. *Journal of Organizational Behavior Management, 18*(1), 7–19. https://doi.org/10.1300/J075v18n01_02
- Aichner, T., Grünfelder, M., Maurer, O., & Jegeni, D. (2021). Twenty-five years of social media: a review of social media applications and definitions from 1994 to 2019. *Cyberpsychology, Behavior, and Social Networking, 24*(4), 215–222. <https://doi.org/10.1089/cyber.2020.0134>
- Akemoglu, Y., Muharib, R., & Meadan, H. (2020). A systematic and quality review of parent-implemented language and communication interventions conducted via telepractice. *Journal of Behavioral Education, 29*(2), 282-316. <https://doi.org/10.1007/s10864-019-09356-3>
- Al-Rahmi, W. M., & Zeki, A. M. (2017). A model of using social media for collaborative learning to enhance learners' performance on learning. *Journal of King Saud University - Computer and Information Sciences, 29*(4), 526–535. <https://doi.org/10.1016/j.jksuci.2016.09.002>
- Albright, L., Reeve, K. F., Reeve, S. A., & Kisamore, A. N. (2015). Teaching statistical variability with equivalence-based instruction. *Journal of Applied Behavior Analysis, 48*(4), 883–894. <https://doi.org/10.1002/jaba.249>
- Albright, L., Schnell, L., Reeve, K. F., & Sidener, T. M. (2016). Using stimulus equivalence-based instruction to teach graduate students in applied behavior analysis to interpret oper-

ant functions of behavior. *Journal of Behavioral Education*, 25(3), 290–309. <https://doi.org/10.1007/s10864-016-9249-0>

Alkathlan, A., & Al-Daraiseh, A. (2017). An analytical study of the use of social networks for collaborative learning in higher education. *International Journal of Modern Education and Computer Science*, 9(2), 1–13. <https://doi.org/10.5815/ijmecs.2017.02.01>

Alvero, A. M., Bucklin, B. R., & Austin, J. (2001). An objective review of the effectiveness and essential characteristics of performance feedback in organizational settings (1985-1998). *Journal of Organizational Behavior Management*, 21(1), 3–29. https://doi.org/10.1300/J075v21n01_02

American Telemedicine Association. (2016). *What is telemedicine?* Retrieved from www.americantelemed.org/about-telemedicine/what-is-telemedicine#.Vsky1ce-0Gsl.

Anderson, D. M., & Keel, M. C. (2002). Using "reasoning and writing" to teach writing skills to students with learning disabilities and behavioral disorders. *Journal of Direct Instruction*, 2(1), 49-55.

Anderson, R., Taylor, S., Taylor, T., & Virues-Ortega, J. (2021). Thematic and textual analysis methods for developing social validity questionnaires in applied behavior analysis. *Behavioral Interventions* 37(3). 732-753. <https://doi.org/10.1002/bin.1832>

Aravamudhan, S., & Awasthi, S. (2021). The use of prompts and precision teaching to address speech sound disorders in a 17-year-old girl with autism. *Behavior Analysis in Practice*, 14(3), 644–659. <https://doi.org/10.1007/s40617-020-00470-7>

Armbruster, B. B., & Osborn, J. (2002). *Reading Instruction and Assessment: Understanding the IRA Standards*. Allyn & Bacon.

Armendariz, F., & Umbreit, J. (1999). Using active responding to reduce disruptive behavior in a general education classroom. *Journal of Positive Behavior Interventions*, 1(3), 152–158. <https://doi.org/10.1177/109830079900100303>

Ausenhuis, J. A., & Higgins, W. J. (2019). An evaluation of real-time feedback delivered via telehealth: training staff to conduct preference assessments. *Behavior Analysis in Practice*, 12(3), 643–648. <https://doi.org/10.1007/s40617-018-00326-1>

- Au-Yong-Oliveira, M., Vitória, A., Silva, C., Carlos, V., Moutinho, V., Moreira, G., & Paiva Dias, G. (2017). Higher education and the problem of abandonment – how can we keep students from leaving? *Proceedings of the International Technology, Education and Development Conference*. <https://doi.org/10.21125/inted.2017.1690>
- Axelrod, S., & Greer, R. D. (1994). Cooperative learning revisited. *Journal of Behavioral Education*, 4(1), 41–48. <https://doi.org/10.1007/BF01560508>
- Ayres, K. M., & Langone, J. (2005). Intervention and instruction with video for students with autism: A review of the literature. *Education and Training in Developmental Disabilities*, 40(2), 183–196.
- Baer, D. M., & Wolf, M. M. (1987). Some still-current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*, 20(4), 313–327. <https://doi.org/10.1901/jaba.1987.20-313>
- Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*, 1(1), 91–97. <https://doi.org/10.1901/jaba.1968.1-91>
- Bahadourian, A. J., Tam, K. Y. (Brian), Greer, R. D., & Rousseau, M. K. (2006). The effects of learn units on student performance in two college courses. *International Journal of Behavioral Consultation and Therapy*, 2(2), 246–264. <https://doi.org/10.1037/h0100780>
- Bailey, J. S., & Burch, M. R. (2017). *Research Methods in Applied Behavior Analysis* (2nd ed.). Routledge.
- Bailey, J. S., & Burch, M. R. (2019). *Ética para analistas de conducta* (3ra ed.). ABA España. <https://doi.org/10.26741/abaspain/2019/Bailey> (Original publicado en 2016)
- Balcazar, F. E., Hopkins, B. L., & Suarez, Y. (1985). A critical, objective review of performance feedback. *Journal of Organizational Behavior Management*, 7(3-4), 65–89. https://doi.org/10.1300/J075v07n03_05
- Barbetta, P. M., Heron, T. E., & Heward, W. L. (1993). Effects of Active Student Response During Error Correction on the Acquisition, Maintenance, and Generalization of Sight Words By

Students With Developmental Disabilities. *Journal of Applied Behavior Analysis*, 26(1), 111–119. <https://doi.org/10.1901/jaba.1993.26-111>

Barczak, M. A., & Cannella-Malone, H. I. (2021). Self-management of vocational skills for people with significant intellectual disabilities: A systematic review. *Journal of Intellectual Disabilities*. <https://doi.org/10.1177/1744629520987768>

Barretto, A., Wacker, D. P., Harding, J., Lee, J., & Berg, W. K. (2006). Using telemedicine to conduct behavioral assessments. *Journal of Applied Behavior Analysis*, 39(3), 333–340. <https://doi.org/10.1901/jaba.2006.173-04>

Baruah, T. D. (2012). Effectiveness of social media as a tool of communication and its potential for technology enabled connections: A micro-level study. *International Journal of Scientific and Research Publications*. *International Journal of Scientific and Research Publications*, 2(5), 1–10. www.ijsrp.org

Bauer, J., and Prenzel, M. (2012). Science education. European teacher training reforms. *Science* 336, 1642–1643. doi: 10.1126/science.1218387

Bećirović, S., Dubravac, V., & Brdarević-Čeljo, A. (2022). Cooperative learning as a pathway to strengthening motivation and improving achievement in an EFL Classroom. *SAGE Open*, 12(1), 215824402210780. <https://doi.org/10.1177/21582440221078016>

Beck, R., & Clement, R. (1991). The great falls precision teaching project: An historical examination. *Journal of Precision Teaching*, 8(2), 8–12.

Becker, A., McLaughlin, T., Weber, K. P., & Gower, J. (2017). The effects of copy, cover and compare with and without additional error drill on multiplication fact fluency and accuracy. *Electronic Journal of Research in Education Psychology*, 7(18), 747–760. <https://doi.org/10.25115/ejrep.v7i18.1368>

Begeny, J. C., Laugle, K. M., Krouse, H. E., Lynn, A. E., Tayrose, M. P., & Stage, S. A. (2010). A control-group comparison of two reading fluency. *School Psychology Review*, 39(1), 137–155.

- Bellini, S., & Akullian, J. (2007). A meta-analysis of video modeling and video self-modeling interventions for children and adolescents with autism spectrum disorders. *Exceptional Children, 73*(3), 264–287. <https://doi.org/10.1177/001440290707300301>
- Biemiller, A., & Boote, C. (2006). An effective method for building meaning vocabulary in primary grades. *Journal of Educational Psychology, 98*(1), 44–62. <https://doi.org/10.1037/0022-0663.98.1.44>
- Binder, C. (1996). Behavioral fluency: Evolution of a new paradigm. *Behavior Analyst, 19*(2), 163–197. <https://doi.org/10.1007/BF03393163>
- Binder, C., & Watkins, C. L. (2013). Article update: precision teaching and direct instruction - measurably superior instructional technologies in schools. *Performance Improvement Quarterly, 26*(2), 96–115. <https://doi.org/10.1002/piq.21145>
- Bitzer, D. L., & Johnson, R. L. (1971). Plato: A computer-based system used in the engineering of education. *Proceedings of the IEEE, 59*(6), 960–968. <https://doi.org/10.1109/PROC.1971.8294>
- Bolt, T. D., Hansen, B. D., Caldarella, P., Richard Young, K., Williams, L., & Wills, H. P. (2019). Varying opportunities to respond to improve behavior of elementary students with developmental disabilities. *International Electronic Journal of Elementary Education, 11*(4), 327–334. <https://doi.org/10.26822/iejee.2019450791>
- Boutain, A. R., Sheldon, J. B., & Sherman, J. A. (2020). Evaluation of a telehealth parent training program in teaching self-care skills to children with autism. *Journal of Applied Behavior Analysis, 53*(3), 1259–1275. <https://doi.org/10.1002/jaba.743>
- Bowman-Perrott, L., Davis, H., Vannest, K., Williams, L., Parker, R., & Greenwood, C. (2013). Academic benefits of peer tutoring: A meta-analytic review of single-case research. *School Psychology Review, 42*(1), 39–55. <https://doi.org/10.1080/02796015.2013.12087490>
- Boyce, T. E., & Hinline, P. N. (2002). Interteaching: A strategy for enhancing the user-friendliness of behavioral arrangements in the college classroom. *Behavior Analyst, 25*(2), 215–226. <https://doi.org/10.1007/BF03392059>

- Boyd, D. M. (2007). Why youth (heart) Social network sites: the role of networked publics in teenage social life. *MacArthur Foundation Series on Digital Learning – Youth, Identity, and Digital Media*, 7641(41), 1–26. <https://doi.org/10.1162/dmal.9780262524834.119>
- Boyd, D. M., & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1), 210–230. <https://doi.org/10.1111/j.1083-6101.2007.00393.x>
- Brandi Simonsen, Sarah Fairbanks, Amy Briesch, Diane Myers, & George Sugai. (2008). Evidence-based Practices in Classroom Management: Considerations for Research to Practice. *Education and Treatment of Children*, 31(1), 351–380. <https://doi.org/10.1353/etc.0.0007>
- Brasch, T. L., Williams, R. L., & McLaughlin, T. F. (2008). The effects of a direct instruction flashcard system on multiplication fact mastery by two high school students with ADHD and ODD. *Child and Family Behavior Therapy*, 30(1), 51–59. https://doi.org/10.1300/J019v30n01_04
- Brett, A., Rothlein, L., & Hurley, M. (1996). Vocabulary acquisition from listening to stories and explanations of target words. *Elementary School Journal*, 96(4), 420–422. <https://doi.org/10.1086/461836>
- Briesch, A. M., & Daniels, B. (2013). Using self-management interventions to address general education behavioral needs: Assessment of effectiveness and feasibility. *Psychology in the Schools*, 50(4), 366–381. <https://doi.org/10.1002/pits.21679>
- Brodsky, J., & Fienup, D. M. (2018). Sidman Goes to College: A meta-analysis of equivalence-based instruction in higher education. *Perspectives on Behavior Science*, 41(1), 95–119. <https://doi.org/10.1007/s40614-018-0150-0>
- Brooke, R. R., & Ruthven, A. J. (1984). The effects of contingency contracting on student performance in a PSI class. *Teaching of Psychology*, 11, 87–89 https://doi.org/10.1207/s15328023top1102_7
- Brosnan, J., Moeyaert, M., Brooks Newsome, K., Healy, O., Heyvaert, M., Onghena, P., & van den Noortgate, W. (2018). multilevel analysis of multiple-baseline data evaluating precision teaching as an intervention for improving fluency in foundational reading skills for at risk

- readers. *Exceptionality*, 26(3), 137–161. <https://doi.org/10.1080/09362835.2016.1238378>
- Brown, C., Schildkamp, K., & Hubers, M. D. (2017). Combining the best of two worlds: a conceptual proposal for evidence-informed school improvement. *Educational Research*, 59(2), 154–172. <https://doi.org/10.1080/00131881.2017.1304327>
- Bruhn, A. L., Vogelgesang, K., Fernando, J., & Lugo, W. (2016). Using data to individualize a multicomponent, technology-based self-monitoring intervention. *Journal of Special Education Technology*, 31(2), 64–76. <https://doi.org/10.1177/0162643416650024>
- Bruno, L. P., Lewis, A. M., Kaldenberg, E. R., Bahr, P. A., & Immerfall, J. (2020). Direct instruction of text-to-speech software for students with intellectual disability. *Education and Training in Autism and Developmental Disabilities*, 55(4), 424–437.
- Buchs, C., & Butera, F. (2015). Cooperative learning and social skills development. *Collaborative Learning: Developments in Research and Practice*, January, 201–217.
- Budney, A. J., Higgins, S. T., Delaney, D. D., Kent, L., & Bickel, W. K. (1991). Contingent reinforcement of abstinence with individuals abusing cocaine and marijuana. *Journal of Applied Behavior Analysis*, 24(4), 657–665. <https://doi.org/10.1901/jaba.1991.24-657>
- Bufford, R. K. (1976). Evaluation of a reinforcement procedure for accelerating work rate in a self-paced course¹. *Journal of Applied Behavior Analysis*, 9(2), 208–208. <https://doi.org/10.1901/jaba.1976.9-208>
- Buga, R., Căpeneată, I., Chirasnel, C., & Popa, A. (2014). Facebook in Foreign Language Teaching – A Tool to Improve Communication Competences. *Procedia - Social and Behavioral Sciences*, 128, 93–98. <https://doi.org/10.1016/j.sbspro.2014.03.124>
- Calkin, A. B. (2005). Precision teaching: The Standard Celeration Charts. *The Behavior Analyst Today*, 6(4), 207–215. <https://doi.org/10.1037/h0100073>
- Carl Hughes, J., Beverley, M., & Whitehead, J. (2007). Using precision teaching to increase the fluency of word reading with problem readers. *European Journal of Behavior Analysis*, 8(2), 221–238. <https://doi.org/10.1080/15021149.2007.11434284>

- Carmody, E., & Stauch, T. (2020). Building play skills using video modeling and matrix training. *Journal of Behavioral Education*, 1 20. <https://doi.org/10.1007/s10864-020-09414-1>
- Carpenter, S. K., & Vul, E. (2011). Delaying feedback by three seconds benefits retention of face-name pairs: the role of active anticipatory processing. *Memory & cognition*, 39(7), 1211–1221. <https://doi.org/10.3758/s13421-011-0092-1>
- Carr, M. E., Moore, D. W., & Anderson, A. (2014). Self-management interventions on students with autism: A meta-analysis of single-subject research. *Exceptional Children*, 81, 28–44. <https://doi.org/10.1177/0014402914532235>.
- Charlop-Christy, M. H., Le, L., & Freeman, K. A. (2000). A comparison of video modeling with in vivo modeling for teaching children with autism. *Journal of Autism and Developmental Disorders*, 30, 537–552. doi:10.1023/A:1005635326276
- Chassiakos, Y. R., Radesky, J., Christakis, D., Moreno, M. A., Cross, C., Hill, D., Ameenuddin, N., Hutchinson, J., Boyd, R., Mendelson, R., Smith, J., & Swanson, W. S. (2016). Children and adolescents and digital media. *Pediatrics*, 138(5). <https://doi.org/10.1542/peds.2016-2593>
- Chen, A., Lu, Y., Chau, P. Y. K., & Gupta, S. (2014). Classifying, Measuring, and Predicting Users' Overall Active Behavior on Social Networking Sites. *Journal of Management Information Systems*, 31(3), 213–253. <https://doi.org/10.1080/07421222.2014.995557>
- Cheng, I. N. Y., Chan, J. K. Y., Kong, S. S. Y., & Leung, K. M. Y. (2016). Effectiveness and obstacle of using Facebook as a tool to facilitate student-centred learning in higher education. *Asia-Pacific Forum on Science Learning and Teaching*, 17(2), 1–15.
- Chernikova, O., Heitzmann, N., Fink, M. C., Timothy, V., Seidel, T., Fischer, F. (2019). Facilitating diagnostic competences in higher education: A meta-analysis in medical and teacher education. *Educational Psychology Review*, 32(1), 157–196. <https://doi.org/10.1007/s10648-019-09492-2>
- Chiesa, M., & Robertson, A. (2000). Precision teaching and fluency training: Making maths easier for pupils and teachers. *Educational Psychology in Practice*, 16(3), 297–310. <https://doi.org/10.1080/713666088>

- Choudhury, I. (2002). Use of reciprocal peer tutoring technique in an environmental control systems course at an undergraduate level. *Journal of Construction Education*, 7(3), 137–142.
- Christie, C. A., & Schuster, J. W. (2003). The effects of using response cards on student participation, academic achievement, and on-task behavior during whole-class, math instruction. *Journal of Behavioral Education*, 12, 147–165.
- Cihon, T. M., White, R., Zimmerman, V. L., Gesick, J., Stordahl, S., & Eshleman, J. (2017). The effects of precision teaching with textual or tact relations on intraverbal relations. *Behavioral Development Bulletin*, 22(1), 129–146. <https://doi.org/10.1037/bdb0000056>
- Clayton, M., & Headley, A. (2019). The use of behavioral skills training to improve staff performance of discrete trial training. *Behavioral Interventions*, 34(1), 136–143. <https://doi.org/10.1002/bin.1656>
- Clearly, T., & Zimmerman, B. J. (2001). Self-regulation differences during athletic practice by experts, no-experts, and novices. *Journal of Applied Sport Psychology*, 13, 61–82. <https://doi.org/10.1080/104132001753149883>
- Conroy, M. A., Sutherland, K. S., & Snyder, A. L., Marsh, S. (2008). Classwide Interventions. Effective instruction makes a difference. *Teaching Exceptional Children* 40(6), 24–30.
- Cooke, N.L., Gibbs, S.L., Campbell, M.L., & Shalvis, S.L. (2004). A Comparison of “Reading Mastery Fast Cycle” and “Horizons Fast Track A-B” on the Reading Achievement of Students with Mild Disabilities. *Journal of Direct Instruction* 4(2), 139-151
- Cooper, J. O., Heron, T. E., & Heward, W. L. (2020). *Applied behavior analysis* (3rd ed.). Hoboken, NJ: Pearson.
- Copeland, S. R., & Hughes, C. (2002). Effects of goal setting on task performance of persons with mental retardation. *Education and Training in Mental Retardation and Developmental Disabilities*, 37(1), 40–54.
- Corbett, B. A., & Abdullah, M. (2005). Video modeling: Why does it work for children with autism? *Journal of Early and Intensive Behavior Intervention*, 2(1), 2-8. <http://dx.doi.org/10.1037/h0100294>

- Crawford, D.B., & Snider, V.E. (2000). Effective mathematics instruction the importance of curriculum. *Education and Treatment of Children*, 23, 122.
- Crone-Todd, D. E., Eyre, H. L., Hutchens, S. A., Jones, J. R., & Pear, J. J. (2007). The effect of changing grading criteria on student writing. *The Behavior Analyst Today*, 8(1), 35-42. <http://dx.doi.org/10.1037/h0100105>
- Crosbie, J. & Kelly, G. (1993). A computer-based Personalized System of Instruction course in applied behavior analysis. *Behavior Research Methods, Instruments, and Computers*, 25(3), 366–370. <https://doi.org/10.3758/BF03204527>
- Crosnoe R. (2006). The connection between academic failure and adolescent drinking in secondary school. *Sociology of Education*, 79(1), 44–60. <https://doi.org/10.1177/003804070607900103>
- Cymbal, D., Wilder, D. A., Cruz, N., Ingraham, G., Llinas, M., Clark, R., & Kamlowky, M. (2021). Procedural integrity reporting in the Journal of Organizational Behavior Management (2000–2020). *Journal of Organizational Behavior Management* <https://doi.org/10.1080/01608061.2021.2014380>
- Dabner, N. (2012). “Breaking Ground” in the use of social media: A case study of a university earthquake response to inform educational design with Facebook. *Internet and Higher Education*, 15(1), 69–78. <https://doi.org/10.1016/j.iheduc.2011.06.001>
- Datchuk, S. M., & Rodgers, D. B. (2019). Text writing within simple sentences: a writing fluency intervention for students with high-incidence disabilities. *Learning Disabilities Research and Practice*, 34(1), 23–34. <https://doi.org/10.1111/ldrp.12185>
- de Melo Wider, L. B., da Silva Barros, R., & Varella, A. (2020). Equivalence class formation in individuals with autism: predictions from ABLA-R levels. *The Analysis of verbal behavior*, 36(2), 215–232. <https://doi.org/10.1007/s40616-020-00134-1>
- Demant, M. S., & Yates, G. C. R. (2003). Primary teachers’ attitudes toward the direct instruction construct. *Educational Psychology*, 23(5), 483–489. <https://doi.org/10.1080/0144341032000123741>

- Demir, M. (2018). Developing a scale for using Facebook as a learning tool. *Educational Technology Research and Development*, 66(6), 1457–1477. <http://www.jstor.org/stable/45018685>
- Desrochers, M. N., Clemmons, T., Grady, M., & Justice, B. (2000). An evaluation of Simulations in Developmental Disabilities (SIDD) : Instructional software that provides practice in behavioral assessment and treatment decisions. *Journal of Technology in Human Services*, 17(4), 15–27. https://doi.org/10.1300/J017v17n04_02
- Dickerson, F. B., Tenhula, W. N., & Green-Paden, L. D. (2005). The token economy for schizophrenia: Review of the literature and recommendations for future research. *Schizophrenia Research*, 75(2–3), 405–416. <https://doi.org/10.1016/j.schres.2004.08.026>
- Didion, L., & Toste, J. R. (2021). Data mountain: self-monitoring, goal setting, and positive attributions to enhance the oral reading fluency of elementary students with or at-risk for reading disabilities. *Journal of Learning Disabilities*. <https://doi.org/10.1177/002221942111043482>
- Dipiro J. T. (2009). Why do we still lecture? *American Journal of Pharmaceutical Education*, 73(8), 137. <https://doi.org/10.5688/aj7308137>
- Dixon, M. R., Belisle, J., Stanley, C. R., Daar, J. H., & Williams, L. A. (2016). Derived equivalence relations of geometry skills in students with autism: an application of the PEAK-E curriculum. *The Analysis of Verbal Behavior*, 32(1), 38–45. <https://doi.org/10.1007/s40616-016-0051-9>
- Donlan, L. (2014). Exploring the views of students on the use of Facebook in university teaching and learning. *Journal of Further and Higher Education*, 38(4), 572–588. <https://doi.org/10.1080/0309877X.2012.726973>
- Duncan, P. K., & Bruwelheide, L. R. (1985-1986). Feedback: Use and possible behavioral functions. *Journal of Organizational Behavior Management*, 7(3-4), 91–114. https://doi.org/10.1300/J075v07n03_06
- Dunlap, G., dePerczel, M., Clarke, S., Wilson, D., Wright, S., White, R., & Gomez, A. (1994). Choice making to promote adaptive behavior for students with emotional and behavioral challenges. *Journal of Applied Behavior Analysis*, 27(3), 505–518. <https://doi.org/10.1901/jaba.1994.27-505>

- Ehrlich, R. J., Nosik, M. R., Carr, J. E., & Wine, B. (2020). Teaching Employees How to Receive Feedback: A Preliminary Investigation. *Journal of Organizational Behavior Management*, 40(1–2), 19–29. <https://doi.org/10.1080/01608061.2020.1746470>
- Ellerani, P., & Gentile, M. (2013). The role of teachers as facilitators to develop empowering leadership and school communities supported by the method of cooperative learning. *Procedia - Social and Behavioral Sciences*, 93, 12–17. <https://doi.org/10.1016/j.sbspro.2013.09.144>
- Ellison, K. S., Guidry, J., Picou, P., Adenuga, P., & Davis, T. E., 3rd (2021). Telehealth and autism prior to and in the age of COVID-19: A systematic and critical review of the last decade. *Clinical child and family psychology review*, 24(3), 599–630. <https://doi.org/10.1007/s10567-021-00358-0>
- Emurian, H. H. (2005). Web-based programmed instruction: Evidence of rule-governed learning. *Computers in Human Behavior*, 21(6), 893–915. <https://doi.org/10.1016/j.chb.2004.03.002>
- Engelmann, S., & Carnine, D. (1982). *Theory of instruction: Principles and applications*. New York: Irvington Publishers.
- Eppler, M. A., & Ironsmith, M. (2004). PSI and distance learning in a developmental psychology course. *Teaching of Psychology*, 31(2), 131–134
- Erbey, R., McLaughlin, T. F., Derby, K. M., & Everson, M. (2011). The effects of using flashcards with reading racetrack to teach letter sounds, sight words, and math facts to elementary students with learning disabilities. *International Electronic Journal of Elementary Education*, 3(3), 213–226.
- Erjavec, K. (2013). Informal Learning through Facebook among Slovenian Pupils. *Comunicar*, 21(41), 117–126. <https://doi.org/10.3916/C41-2013-11>
- Estrapala, S., & Reed, D. K. (2019). Goal-setting instruction: A step-by-step guide for high school students. *Intervention in School and Clinic*, 55, 286–293. <https://doi.org/10.1177/1053451219881717>

- Ethington, A. T., Spriggs, A. D., Shepley, S. B., & Bausch, M. E. (2021). Behavior skills training for teaching and generalizing self-instruction skills for students with intellectual disability. *Journal of Intellectual Disabilities*. <https://doi.org/10.1177/1744629521995349>
- Fajfar, P., Campitelli, G., & Labollita, M. (2012). Effects of immediacy of feedback on estimations and performance. *Australian Journal of Psychology*, 64(3), 169–177. <https://doi.org/10.1111/j.1742-9536.2011.00048.x>
- Farros, J. N., Shawler, L. A., Gatzunis, K. S., & Weiss, M. J. (2020). The effect of synchronous discussion sessions in an asynchronous course. *Journal of Behavioral Education*, 1-13. <https://doi.org/10.1007/s10864-020-09421-2>
- Ferguson, J., Craig, E. A., & Dounavi, K. (2019). Telehealth as a model for providing behaviour analytic interventions to individuals with autism spectrum disorder: A systematic review. *Journal of autism and developmental disorders*, 49(2), 582–616. <https://doi.org/10.1007/s10803-018-3724-5>
- Fienup, D. M., Covey, D. P., & Critchfield, T. S. (2010). Teaching brain-behavior relations economically with stimulus equivalence technology. *Journal of Applied Behavior Analysis*, 43(1), 19–33. <https://doi.org/10.1901/jaba.2010.43-19>
- Fienup, D. M., Critchfield, T. S., & Ninness, C. (2010). Efficiently establishing concepts of inferential statistics and hypothesis decision making through contextually controlled equivalence classes. *Journal of Applied Behavior Analysis*, 43(3), 437–462. <https://doi.org/10.1901/jaba.2010.43-437>
- Fiske K. E. (2008). Treatment integrity of school-based behavior analytic interventions: a review of the research. *Behavior Analysis in Practice*, 1(2), 19–25. <https://doi.org/10.1007/BF03391724>
- Freishtat, R. L., & Sandlin, J. A. (2010). shaping youth discourse about technology: Technological colonization, manifest destiny, and the frontier myth in Facebook's public pedagogy. *Educational Studies*, 46(5), 503–523. <https://doi.org/10.1080/00131946.2010.510408>
- Fuentes, B., de la Fuente-Gómez, L., Sempere-Iborra, C., Delgado-Fernández, C., Tarifa-Rodríguez, A., Alonso de Leciñana, M., de Celis-Ruiz, E., Gutiérrez-Zúñiga, R., López-Tàpper, J., Martín Alonso, M., Pastor-Yborra, S., Rigual, R., Ruiz-Ares, G., Rodríguez-Pardo, J.,

- Virués-Ortega, J., Borobia, A. M., Blanco, P., & Bueno-Guerra, N. (2022). DUBbing Language-therapy CINema-based in Aphasia post-Stroke (DULCINEA): study protocol for a randomized crossover pilot trial. *Trials*, 23(1), 21. <https://doi.org/10.1186/s13063-021-05956-5>
- Funk, S. C., & Dickson, K. L. (2011). Multiple-choice and short-answer exam performance in a college classroom. *Teaching of Psychology*, 38(4), 273–277. <https://doi.org/10.1177/0098628311421329>
- Gamage, S., Ayres, J. R., & Behrend, M. B. (2022). A systematic review on trends in using Moodle for teaching and learning. *International journal of STEM education*, 9(1), 9. <https://doi.org/10.1186/s40594-021-00323-x>
- Geiger, K. B., LeBlanc, L. A., Hubik, K., Jenkins, S. R., & Carr, J. E. (2018). Live training versus e-learning to teach implementation of listener response programs. *Journal of Applied Behavior Analysis*, 51(2), 220–235. <https://doi.org/10.1002/jaba.444>
- Goh, C. F., Rasli, A., Tan, O. K., & Sang, L. C. (2019). Determinants and academic achievement effect of Facebook use in educational communication among university students. *Aslib Journal of Information Management*, 71(1), 105-123. <https://doi.org/10.1108/AJIM-05-2018-0116>
- Gorgan, E. M., & Kodak, T. (2019). Comparison of interventions to treat prompt dependence for children with developmental disabilities. *Journal of Applied Behavior Analysis*, 52(4), 1049–1063. <https://doi.org/10.1002/jaba.638>
- Grant, L.K., & Spencer, R.E. (2003). The personalized system of instruction: Review and applications to distance education. *The International Review of Research in Open and Distributed Learning*, 4(2), 1–17. <https://doi.org/10.19173/irrodl.v4i2.152>
- Greenberb, S. (2010). *The effects of procedural integrity on skill acquisition* [Master's thesis]. Northeastern University, Boston, Massachusetts. <https://repository.library.northeastern.edu/files/neu:506/fulltext.pdf>
- Greer, R. D. (1994). A systems analysis of the behaviors of schooling. *Journal of Behavioral Education*, 4(3), 255–264. <https://doi.org/10.1007/BF01531981>

- Greer, R. D. (1997). The comprehensive application of behavior analysis to schooling (CABAS®). *Behavior and Social Issues*, 7(1), 59–63. <https://doi.org/10.5210/bsi.v7i1.300>
- Greer, R. D., & McDonough, S. H. (1999). Is the learn unit a fundamental measure of pedagogy? *Behavior Analyst*, 22(1), 5–16. <https://doi.org/10.1007/BF03391973>
- Grisante, P. C., de Rose, J. C., & McIlvane, W. J. (2014). Controlling relations in stimulus equivalence classes of preschool children and individuals with down syndrome. *The Psychological record*, 64(2), 195–208. <https://doi.org/10.1007/s40732-014-0021-3>
- Grosbeck, G., Bran, R., & Tiru, L. (2011). Dear teacher, what should I write on my wall? A case study on academic uses of Facebook. *Procedia - Social and Behavioral Sciences*, 15, 1425–1430. <https://doi.org/10.1016/j.sbspro.2011.03.306>
- Grossen, B. (2004). Success of a direct instruction model at a secondary level school with high-risk students. *Reading and Writing Quarterly*, 20(2), 161–178. <https://doi.org/10.1080/10573560490262091>
- Gureasko-Moore, S., DuPaul, G. J., & White, G. P. (2006). The effects of self-management in general education classrooms on the organizational skills of adolescents with ADHD. *Behavior Modification*, 30(2), 159–183. <https://doi.org/10.1177/0145445503259387>
- Haas, J. R., & Hayes, S. C. (2006). When knowing you are doing well hinders performance: exploring the interaction between rules and feedback. *Journal of Organizational Behavior Management*, 26(1-2), 91–111. https://doi.org/10.1300/J075v26n01_04
- Hagopian L. P. (2020). The consecutive controlled case series: Design, data-analytics, and reporting methods supporting the study of generality. *Journal of Applied Behavior Analysis*, 53(2), 596–619. <https://doi.org/10.1002/jaba.691>
- Haladyna, T. M. (2004). *Developing and Validating Multiple-Choice Test Items*, 3rd Ed. Erlbaum.
- Hall, S. S., Debernardis, G. M., & Reiss, A. L. (2006). The acquisition of stimulus equivalence in individuals with fragile X syndrome. *Journal of Intellectual Disability Research: JIDR*, 50(Pt 9), 643–651. <https://doi.org/10.1111/j.1365-2788.2006.00814.x>

- Hallahan, D.P., Lloyd, J.W., Kosiewicz, M.M., Kauffman, J.M., & Graves, A.W. (1979). Self-monitoring of attention as a treatment for a learning-disabled boy's off-task behavior. *Learning Disability Quarterly*, 2, 24 - 32. <https://doi.org/10.2307/1511021>
- Hanley, G. P. (2017). Editor's Note. *Journal of Applied Behavior Analysis*, 50(1), 3–7. doi:10.1002/jaba.366
- Hansen, B. D., & Wills, H. P. (2014). The effects of goal setting, contingent reward, and instruction on writing skills. *Journal of Applied Behavior Analysis*, 47(1), 171–175. <https://doi.org/10.1002/jaba.92>
- Harris, K. R., Barbara, D. F., Saddler, B., Frizzelle, R., & Graham, S. (2005). Self-monitoring of attention versus self-monitoring of academic performance: Effects among students with ADHD in the general education classroom. *The Journal of Special Education*, 39(3), 145-156. <http://dx.doi.org/10.1177/0022466905039003020>
- Hattie, J. (2008). Visible learning: A synthesis of over 800 meta-analyses relating to achievement. *In Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement*. <https://doi.org/10.4324/9780203887332>
- Haydon, T., Macsuga-Gage, A. S., Simonsen, B., & Hawkins, R. (2012). Opportunities to Respond: A Key Component of Effective Instruction. *Beyond Behavior*, 22(1), 23–31. <https://doi.org/10.1177/107429561202200105>
- Haydon, T., Mancil, G. R., & Van Loan, C. (2009). Using opportunities to respond in a general education classroom: A case study. *Education & Treatment of Children*, 32, 267–278. <http://dx.doi.org/10.1353/etc.0.0052>
- Haydon, T., Mancil, G. R., Kroeger, S. D., McLeskey, J., & Lin, W.-Y. J. (2011). A review of the effectiveness of guided notes for students who struggle learning academic content. *Preventing School Failure: Alternative Education for Children and Youth*, 55(4), 226–231. <https://doi.org/10.1080/1045988x.2010.548415>
- Haydon, T., Marsicano, R., & Scott, T. M. (2013). A comparison of choral and individual responding: A review of the literature. *Preventing School Failure: Alternative Education for Children and Youth*, 57(4), 181–188. <https://doi.org/10.1080/1045988x.2012.682184>

- Hayes, S. C., Rosenfarb, I., Wulfert, E., Munt, E. D., Korn, Z., & Zettle, R. D. (1985). Self-reinforcement effects: An artifact of social standard setting? *Journal of Applied Behavior Analysis*, 18(3), 201–214. <https://doi.org/10.1901/jaba.1985.18-201>
- Hayirođlu, M. İ. (2019). Telemedicine: Current concepts and future perceptions. *Anatolian Journal of Cardiology*, 22, 21–22. <https://doi.org/10.14744/AnatolJCardiol.2019.12525>
- Helbig, K. A., Radley, K. C., Schrieber, S. R., & Derieux, J. R. (2021). Vocational social skills training for individuals with intellectual and developmental disabilities: A Pilot Study. *Journal of Behavioral Education*, 0123456789. <https://doi.org/10.1007/s10864-021-09445-2>
- Hempenstall, K. (2014). What works? Evidence-based practice in education is complex. *Australian Journal of Learning Difficulties*, 19(2), 113–127. <https://doi.org/10.1080/19404158.2014.921631>
- Hillman, W. (2003). Learning how to learn: problem-based learning. *Australian Journal of Teacher Education*, 28(2). <http://dx.doi.org/10.14221/ajte.2003v28n2.1>
- Hirsch, S. E., Ennis, R. P., & McDaniel, S. C. (2013). Student self-graphing as a strategy to increase teacher effectiveness and student motivation. *Beyond Behavior*, 22(3), 31–39. <https://doi.org/10.1177/107429561302200305>
- Hirst, J. M., DiGennaro Reed, F. D., & Reed, D. D. (2013). Effects of Varying Feedback Accuracy on Task Acquisition: A Computerized Translational Study. *Journal of Behavioral Education*, 22(1), 1–15. <https://doi.org/10.1007/s10864-012-9162-0>
- Houmanfar, R. (2013). Performance feedback: From component analysis to application. *Journal of Organizational Behavior Management*, 33(2), 85–88. <https://doi.org/10.1080/01608061.2013.787002>
- Huang, R., Du, J., Chang, Tw., Spector, M., Zhang, Y., Zhang, A. (2017). A Conceptual Framework for a Smart Learning Engine. *Innovations in Smart Learning*. Lecture Notes in Educational Technology. Springer, Singapore. https://doi.org/10.1007/978-981-10-2419-1_11
- Hughes, E. M., & Yakubova, G. (2019). Addressing the mathematics gap for students with ASD: an evidence-based systematic review of video-based mathematics interventions. *Review*

Journal of Autism and Developmental Disorders, 6(2), 147–158. <https://doi.org/10.1007/s40489-019-00160-3>

Hurst, E. J. (2016). Evolutions in telemedicine: from smoke signals to mobile health solutions. *Journal of Hospital Librarianship*, 16(2), 174–185. <https://doi.org/10.1080/15323269.2016.1150750>

Hurt, N., Moss, G., Bradley, C., Larson, L., Lovelace, M., Prevost, L., Riley, N., Domizi, D. and Camus, M., (2012). The 'Facebook' Effect: College students' perceptions of online discussions in the age of social networking. *International Journal for the Scholarship of Teaching and Learning*, 6(2). <https://doi.org/10.20429/ijstl.2012.060210>

Hurtado-Parrado, C., Pfaller-Sadovsky, N., Medina, L., Gayman, C. M., Rost, K. A., & Schofill, D. (2022). A systematic review and quantitative analysis of interteaching. *Journal of Behavioral Education*, 31(1), 157–185 . <https://doi.org/10.1007/s10864-021-09452-3>

Ingham, P., & Greer, R. D. (1992). Changes in student and teacher responses in observed and generalized settings as a function of supervisor observations. *Journal of Applied Behavior Analysis*, 25(1), 153–164. <https://doi.org/10.1901/jaba.1992.25-153>

Ivy, J. W., Meindl, J. N., Overley, E., & Robson, K. M. (2017). Token economy: A systematic review of procedural descriptions. *Behavior Modification*, 41(5), 708–737. <https://doi.org/10.1177/0145445517699559>

Jaki, T., & Wason, J. (2018). Multi-arm multi-stage trials can improve the efficiency of finding effective treatments for stroke: a case study. *BMC Cardiovascular Disorders*, 18(1), 215. <https://doi.org/10.1186/s12872-018-0956-4>

Jimenez, S. T., & Gayman, C. M. (2021). Interteaching: How much does each component increase its efficacy?. *Journal of Applied Behavior Analysis*, 54(4), 1503–1513. <https://doi.org/10.1002/jaba.848>

Johnson, D. W., & Johnson, R. T. (1982). The effects of cooperative and individualistic instruction on handicapped and nonhandicapped students. *The Journal of Social Psychology*, 118(2), 257–268. <https://doi.org/10.1080/00224545.1982.9922805>

- Johnson, K., & Street, E. M. (2012). From the laboratory to the field and back again: Morningside Academy's 32 years of improving students' academic performance. *The Behavior Analyst Today*, 13(1), 20–40. <https://doi.org/10.1037/h0100715>
- Jones, R. T., Kazdin, A. E., & Haney, J. I. (1981). Social validation and training of emergency fire safety skills for potential injury prevention and lifesaving. *Journal of Applied Behavior Analysis*, 14(3), 249–260. <https://doi.org/10.1901/jaba.1981.14-249>
- Kabilan, M. K., Ahmad, N., & Abidin, M. J. Z. (2010). Facebook: An online environment for learning of English in institutions of higher education? *The Internet and Higher Education*, 13(4), 179–187. <https://doi.org/10.1016/J.IHEDUC.2010.07.003>
- Kame'enui, E. J., & Carnine, D. W. (1986). Pre-teaching versus concurrent teaching of the component skills of a subtraction algorithm: A component analyses of direct instruction. *Exceptional Child*, 33(2), 103–115. <https://doi.org/10.1080/0156655860330203>
- Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of social media. *Business Horizons*, 53(1), 59–68. <https://doi.org/10.1016/j.bushor.2009.09.003>
- Karle, A. (2013). Kaveri Subrahmanyam and David Šmahel: Digital youth: The role of media in development. *Journal of Youth and Adolescence*, 42(2), 308–310. <https://doi.org/10.1007/s10964-012-9852-x>
- Keller, F. S. (1968). "Good-bye teacher". *Journal of Applied Behavior Analysis*, 1, 79–89
- Kelly, M. B. (1976). A review of academic permanent-product data collection and reliability procedures in applied behavior analysis research. *Journal of Applied Behavior Analysis*, 9(2), 211. <https://doi.org/10.1901/jaba.1976.9-211>
- Keohane, D.-D., & Greer, R. D. (2005). Teachers' use of a verbally governed algorithm and student learning. *International Journal of Behavioral Consultation and Therapy*, 1(3), 252–271. <https://doi.org/10.1037/h0100749>
- Keohane, D.-D., Luke, N., & Greer, R. D. (2008). The things we care to see: The effects of rotated protocol immersion on the emergence of early observing responses. *Journal of Early and Intensive Behavior Intervention*, 5(1), 23–39. <https://doi.org/10.1037/h0100408>

- Kern J. (2006). Evaluation of teleconsultation systems. *International Journal of Medical Informatics*, 75(3-4), 330–334. <https://doi.org/10.1016/j.ijmedinf.2005.08.001>
- Kim, J. Y., Fienup, D. M., Oh, A. E., & Wang, Y. (2021). Systematic review and meta-analysis of token economy practices in K-5 educational settings, 2000 to 2019. *Behavior Modification*. <https://doi.org/10.1177/01454455211058077>
- Kinder, D., & Carnine, D. (1991). Direct instruction: What it is and what it is becoming. *Journal of Behavioral Education*, 1(2), 193–213. <https://doi.org/10.1007/BF00957004>
- King-Sears, Margaret E. (2008). Using teacher and researcher data to evaluate the effects of self-management in an inclusive classroom. *Preventing School Failure*, 52(4), 25-36.
- Kirkpatrick, M., Akers, J., & Rivera, G. (2019). Use of behavioral skills training with teachers: A systematic review. *Journal of Behavioral Education*, 28(3), 344–361. <https://doi.org/10.1007/s10864-019-09322-z>
- Kleinert, W. L., Coddling, R. S., Minami, T., & Gould, K. (2018). A meta-analysis of the taped problems intervention. *Journal of Behavioral Education*, 27(1), 53–80. <https://doi.org/10.1007/s10864-017-9284-5>
- Kocdar, S., Karadeniz, A., & Goksel, N. (2018). Using Facebook for leveraging sense of community in self-paced open and distance learning courses. *International Journal of Emerging Technologies in Learning*, 13(5), 100–116. <https://doi.org/10.3991/IJET.V13I05.8192>
- Koegel, L. K., Harrower, J. K., & Koegel, R. L. (1999). Support for children with developmental disabilities in full inclusion classrooms through self-management. *Journal of Positive Behavior Interventions*, 1(1), 26–34. <https://doi.org/10.1177/109830079900100104>
- Kowalewicz, E. A., & Coffee, G. (2014). Mystery Motivator: a Tier 1 classroom behavioral intervention. *School psychology quarterly: The official journal of the division of school psychology. American Psychological Association*, 29(2), 138–156. <https://doi.org/10.1037/spq0000030>
- Kratochwill, T. R., Albers, C. A., & Steele Shernoff, E. (2004). School-based interventions. *Child and Adolescent Psychiatric Clinics of North America*, 13(4), 885–903. <https://doi.org/10.1016/j.chc.2004.05.003>

- Kubina, R. M. (2021). Precision teaching and behavior dynamics. *Behavior Analysis in Practice*, 14(3), 577–581. <https://doi.org/10.1007/s40617-020-00482-3>
- Kulik, C. L. C., Kulik, J. A., & Bangert-Drowns, R. L. (1990). Effectiveness of mastery learning programs: A Meta-Analysis. *Review of Educational Research*, 60(2), 265–299. <https://doi.org/10.3102/00346543060002265>
- Kulik, C.-L. C., Kulik, J. A., & Bangert-Drowns, R. L. (1990). Effectiveness of mastery learning programs: A meta-Analysis. *Review of Educational Research*, 60(2), 265–299. <https://doi.org/10.2307/1170612>
- Kulik, J. A., & Kulik, C.-I. C. (1988). Timing of feedback and verbal learning. *Review of Educational Research*, 58(1), 79–97. <https://doi.org/10.2307/1170349>
- Kulik, J. A., Kulik, C.-I. C., & Cohen, P. A. (1979). A meta-analysis of outcome studies of Keller's personalized system of instruction. *American Psychologist*, 34(4), 307–318. <https://doi.org/10.1037/0003-066X.34.4.307>
- Kumm, S., & Maggin, D.M. (2021). Intensifying goal-setting interventions for students with emotional and behavioral disorders. *Beyond Behavior*, 30, 14 - 23. <https://doi.org/10.1177/1074295621996616>
- Lane, K. L., Carter, E. W., Jenkins, A., Dwiggins, L., & Germer, K. (2015). Supporting comprehensive, integrated, three-tiered models of prevention in schools: Administrators' perspectives. *Journal of Positive Behavior Interventions*, 17, 209–222. <http://dx.doi.org/10.1177/1098300715578916>
- Lattal, K. A. (2010). Delayed reinforcement of operant behavior. *Journal of the Experimental Analysis of Behavior*, 93(1), 129–139. <https://doi.org/10.1901/jeab.2010.93-129>
- Lee, J. A. N. (2004). History of computing in education: An overview. *IFIP Advances in Information and Communication Technology*, 145(April 2006), 1–16. https://doi.org/10.1007/1-4020-8136-7_1
- Leung, K. C. (2015). Preliminary empirical model of crucial determinants of best practice for peer tutoring on academic achievement. *Journal of Educational Psychology*, 107(2), 558–579. <https://doi.org/10.1037/a0037698>

- Leung, K. C. (2019). An updated meta-analysis on the effect of peer tutoring on tutors' achievement. *School Psychology International*, 40(2), 200–214. <https://doi.org/10.1177/0143034318808832>
- Lindsley O. R. (1992). Precision teaching: Discoveries and effects. *Journal of Applied Behavior Analysis*, 25(1), 51–57. <https://doi.org/10.1901/jaba.1992.25-51>
- Lindström, B., Bellander, M., & Schultner, D.T. (2021). A computational reward learning account of social media engagement. *Nature Communications*, 12, 1311. <https://doi.org/10.1038/s41467-020-19607-x>
- Liu, W. (2021). Does teacher immediacy affect students? A systematic review of the association between teacher verbal and non-verbal immediacy and student motivation. *Frontiers in Psychology*, 12, 713978. <https://doi.org/10.3389/fpsyg.2021.713978>
- Lloveras, L. A., Tate, S. A., Vollmer, T. R., King, M., Jones, H., & Peters, K. P. (2022). Training behavior analysts to conduct functional analyses using a remote group behavioral skills training package. *Journal of Applied Behavior Analysis*, 55(1), 290–304. <https://doi.org/10.1002/JABA.893>
- Locke, E., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation. *American Psychologist*, 57, 705–717. <https://doi.org/10.1037//0003-066X.57.9.705>
- Lovitz, E. D., Cihon, T. M., & Eshleman, J. W. (2020). Exploring the Effects of Daily, Timed, and Typed Technical Term Definition Practice on Indicators of Fluency. *Behavior Analysis in Practice*, 14(3), 704–727. <https://doi.org/10.1007/s40617-020-00481-4>
- Ludwig, T. D., & Geller, E. S. (1991). Improving the driving practices of pizza deliverers: Response generalization and moderating effects of driving history. *Journal of Applied Behavior Analysis*, 24(1), 31–44. <https://doi.org/10.1901/jaba.1991.24-31>
- MacDonald, R., Sacramone, S., Mansfield, R., Wiltz, K., & Ahearn, W. H. (2009). Using video modeling to teach reciprocal pretend play to children with autism. *Journal of Applied Behavior Analysis*, 42(1), 43–55. <https://doi.org/10.1901/jaba.2009.42-43>

- Mace, F. C., & Critchfield, T. S. (2010). Translational research in behavior analysis: historical traditions and imperative for the future. *Journal of the Experimental Analysis of Behavior*, 93(3), 293–312. <https://doi.org/10.1901/jeab.2010.93-293>
- Mangiapanello, K. A., & Hemmes, N. S. (2015). An Analysis of feedback from a behavior analytic perspective. *The Behavior analyst*, 38(1), 51–75. <https://doi.org/10.1007/s40614-014-0026-x>
- Mangundayao, J., McLaughlin, T. F., Williams, R. L., & Toone, E. (2013). An evaluation of a direct instructions flashcard system on the acquisition and generalization of numerals, shapes, and colors for preschool-aged students with developmental delays. *Journal of Developmental and Physical Disabilities*, 25(4), 461–473. <https://doi.org/10.1007/s10882-012-9326-9>
- Mann, V. A., & Foy, J. G. (2003). Phonological awareness, speech development, and letter knowledge in preschool children. *Annals of Dyslexia*, 53, 149–173. <https://doi.org/10.1007/s11881-003-0008-2>
- Marchand-Martella, N. E., Martella, R. C., Modderman, S. L., Petersen, H. M., & Pan, S. (2013). Key areas of effective adolescent literacy programs. *Education and Treatment of Children*, 36(1), 161–184. <https://doi.org/10.1353/etc.2013.0005>
- Martin, T. L., Pear, J. J., & Martin, G. L. (2002a). Analysis of proctor marking accuracy in a computer-aided personalized system of instruction course. *Journal of Applied Behavior Analysis*, 35(3), 309–312. <https://doi.org/10.1901/jaba.2002.35-309>
- Martin, T. L., Pear, J. J., & Martin, G. L. (2002b). Feedback and its effectiveness in a computer-aided personalized system of instruction course. *Journal of Applied Behavior Analysis*, 35(4), 427–430. <https://doi.org/10.1901/jaba.2002.35-427>
- Mayfield, A. (2008). *What is social media?* iCrossing. Retrieved August 25, 2009, from http://crmexchange.com/uploadedFiles/White_Papers/PDF/What_is_Social_Media_iCrossing_ebook.pdf
- McCutchen, D. (1996). A capacity theory of writing: working memory in composition. *Educational Psychology Review*, 8(3), 299–325. <https://doi.org/10.1007/BF01464076>

- McDougall, D. (1998). Research on self-management techniques used by students with disabilities in general education settings. *Remedial and Special Education, 19*(5), 310–320. <https://doi.org/10.1177/074193259801900507>
- McDougall, D., Heine, R. C., Wiley, L. A., Sheehey, M. D., Sakanashi, K. K., Cook, B. G., & Cook, L. (2017). Meta-analysis of behavioral self-management techniques used by students with disabilities in inclusive settings. *Behavioral Interventions, 32*(4), 399–417. <https://doi.org/10.1002/bin.1491>
- McLay, L., Sutherland, D., Machalicek, W., & Sigafos, J. (2020). Systematic review of telehealth interventions for the treatment of sleep problems in children and adolescents. *Journal of Behavioral Education, 29*(2), 222–245. <https://doi.org/10.1007/s10864-020-09364-8>
- McTiernan, A., Holloway, J., Leonard, C., & Healy, O. (2018). Employing precision teaching, frequency-building, and the Morningside Math facts curriculum to increase fluency with addition and subtraction computation: A randomized-controlled trial. *European Journal of Behavior Analysis, 19*(1), 90–104. <https://doi.org/10.1080/15021149.2018.1438338>
- Messer, D., Dockrell, J. E., & Murphy, N. (2004). Relation between naming and literacy in children with word-finding difficulties. *Journal of Educational Psychology, 96*(3), 462–470. <https://doi.org/10.1037/0022-0663.96.3.462>
- Metsala, J. L. (1999). Young children's phonological awareness and nonword repetition as a function of vocabulary development. In *Journal of Educational Psychology* (Vol. 91, Issue 1, pp. 3–19). <https://doi.org/10.1037/0022-0663.91.1.3>
- Michael J. (1974). Statistical inference for individual organism research: some reactions to a suggestion by Gentile, Roden, and Klein. *Journal of Applied Behavior Analysis, 7*(4), 627–628. <https://doi.org/10.1901/jaba.1974.7-627>
- Michael J. (1974). Statistical inference for individual organism research: mixed blessing or curse? *Journal of Applied Behavior Analysis, 7*(4), 647–653. <https://doi.org/10.1901/jaba.1974.7-647>
- Milošević, I., Živković, D., Arsić, S., & Manasijević, D. (2015). Facebook as virtual classroom - Social networking in learning and teaching among Serbian students. *Telematics and Informatics, 32*(4), 576–585. <https://doi.org/10.1016/j.tele.2015.02.003>

- Miltenberger, R. G., & Olsen, L. A. (1996). Abduction Prevention Training: A Review of Findings and Issues for Future Research. *Education and Treatment of Children*, 19(1), 69–82. <http://www.jstor.org/stable/42899442>
- Miltenberger, R. G., Flessner, C., Gatheridge, B., Johnson, B., Satterlund, M., & Egemo, K. (2004). Evaluation of behavioral skills training to prevent gun play in children. *Journal of Applied Behavior Analysis*, 37(4), 513–516. <https://doi.org/10.1901/jaba.2004.37-513>
- Miltenberger, R. G., Roberts, J. A., Ellingson, S., Galensky, T., Rapp, J. T., Long, E. S., & Lumley, V. A. (1999). Training and generalization of sexual abuse prevention skills for women with mental retardation. *Journal of Applied Behavior Analysis*, 32(3), 385–388. <https://doi.org/10.1901/jaba.1999.32-385>
- Moeller. (2012). Goal Setting and Student Achievement: A Longitudinal Study. *The Modern Language Journal*, 96(2), 153–169. <https://doi.org/10.2307/41684067>
- Moore, M. (1999). Evolution of telemedicine. *Future Generation Computer Systems*, 15(2), 245–254. [https://doi.org/10.1016/S0167-739X\(98\)00067-3](https://doi.org/10.1016/S0167-739X(98)00067-3)
- Morgan, G. A., Kim, J. Y., & Fienup, D. M. (2020). Effects of nonreciprocal peer tutoring with pre-school students. *Behavior Analysis in Practice*, 13(4), 950–954. <https://doi.org/10.1007/s40617-020-00422-1>
- Morris, E. K., Smith, N. G., & Altus, D. E. (2005). B.F. Skinner's contributions to applied behavior analysis. *Behavior Analyst*, 28(2), 99–131. <https://doi.org/10.1007/BF03392108>
- Morton, W.L., Heward, W.L., & Alber, S.R. (1998). When to self-correct? a comparison of two procedures on spelling performance. *Journal of Behavioral Education*, 8, 321–335. <https://doi.org/10.1023/A:1022871230565>
- Murray, L., & Sefchik, G. (1992). Regulating behavior management practices in residential treatment facilities. *Children and Youth Services Review*, 14(6), 519–539. [https://doi.org/10.1016/0190-7409\(92\)90004-f](https://doi.org/10.1016/0190-7409(92)90004-f)
- Najdowski, A. C., Wallace, M. D., Penrod, B., Tarbox, J., Reagon, K., & Higbee, T. S. (2008). Caregiver-conducted experimental functional analyses of inappropriate mealtime behavior.

ior. *Journal of Applied Behavior Analysis*, 41(3), 459–465. <https://doi.org/10.1901/jaba.2008.41-459>

Narita Y. (2021). Incorporating ethics and welfare into randomized experiments. *Proceedings of the National Academy of Sciences of the United States of America*, 118(1), e2008740118. <https://doi.org/10.1073/pnas.2008740118>

Neely, L., Hong, E. R., Kawamini, S., Umana, I., & Kurz, I. (2020). Intercontinental telehealth to train Japanese interventionists in incidental teaching for children with autism. *Journal of Behavioral Education*, 29(2), 433–448. <https://doi.org/10.1007/s10864-020-09377-3>

Nicolai, L., Schmidbauer, M., Gradel, M., Ferch, S., Antón, S., Hoppe, B., Pander, T., Borch, P. V., Pinnilla, S., Fischer, M., & Dimitriadis, K. (2017). Facebook groups as a powerful and dynamic tool in medical education: Mixed-method study. *Journal of Medical Internet Research*, 19(12). <https://doi.org/10.2196/jmir.7990>

Nigro-Bruzzi, D., & Sturmey, P. (2010). The effects of behavioral skills training on mand training by staff and unprompted vocal mands by children. *Journal of Applied Behavior Analysis*, 43(4), 757–761. <https://doi.org/10.1901/jaba.2010.43-757>

Noell, G. H., & Gansle, K. A. (2014). The use of performance feedback to improve intervention implementation in schools. In L. M. Hagermoser Sanetti & T. R. Kratochwill (Eds.), *Treatment integrity: A foundation for evidence-based practice in applied psychology* (pp. 161–183). American Psychological Association.

Oliveira, M., Goyos, C., & Pear, J. (2012). A pilot investigation comparing instructional packages for MTS training: “Manual alone” vs. “manual-plus-computer-aided personalized system of instruction.” *The Behavior Analyst Today*, 13(3–4), 20–26. <https://doi.org/10.1037/h0100727>

Pantermuehl, R. M., & Lechago, S. A. (2015). A comparison of feedback provided in vivo versus an online platform on the treatment integrity of staff working with children with autism. *Behavior Analysis in Practice*, 8(2), 219–222. <https://doi.org/10.1007/s40617-015-0059-y>

Parsons, M. B., Rollyson, J. H., & Reid, D. H. (2013). Teaching practitioners to conduct behavioral skills training: A Pyramidal approach for training multiple human service staff. *Behavior Analysis in Practice*, 6(2), 4–16. <https://doi.org/10.1007/BF03391798>

- Partin, T. C. M., Robertson, R. E., Maggin, D. M., Oliver, R. M., & Wehby, J. H. (2009). Using Teacher Praise and Opportunities to Respond to Promote Appropriate Student Behavior. *Preventing School Failure: Alternative Education for Children and Youth*, 54(3), 172–178. <https://doi.org/10.1080/10459880903493179>
- Patall, E. A., Cooper, H., & Wynn, S. R. (2010). The effectiveness and relative importance of choice in the classroom. *Journal of Educational Psychology*, 102, 896–915. <https://doi.org/10.1037/a0019545>
- Pear, J. J. & Crone-Todd, D. E. (1999). Personalized system of instruction in cyberspace. *Journal of Applied Behavior Analysis*, 32, 205–209. <https://doi.org/10.1901/jaba.1999.32-205>
- Pear, J. J., & Kinsner, W. (1988). Computer-Aided Personalized System of Instruction: An effective and economical method for short- and long-distance education. *Machine-Mediated Learning*, 2(3), 213–237.
- Pear, J. J., & Novak, M. (1996). Computer-Aided Personalized System of Instruction: A program evaluation. *Teaching of Psychology*, 23(2), 119–123. https://doi.org/10.1207/s15328023top2302_14
- Pellegrino, A. J., & DiGennaro Reed, F. D. (2020). Using telehealth to teach valued skills to adults with intellectual and developmental disabilities. *Journal of Applied Behavior Analysis*, 53(3), 1276–1289. <https://doi.org/10.1002/jaba.734>
- Pence, S. T., St Peter, C. C., & Tetreault, A. S. (2012). Increasing accurate preference assessment implementation through pyramidal training. *Journal of Applied Behavior Analysis*, 45(2), 345–359. <https://doi.org/10.1901/jaba.2012.45-345>
- Perin, C. T. (1943). A quantitative investigation of the delay-of-reinforcement gradient. *Journal of Experimental Psychology*, 32(1), 37–51. <https://doi.org/10.1037/h0056738>
- Peterson N. (1982). Feedback is not a new principle of behavior. *The Behavior Analyst*, 5(1), 101–102. <https://doi.org/10.1007/BF03393144>
- Petry, N. M., Barry, D., Pescatello, L., & White, W. B. (2011). A low-cost reinforcement procedure improves short-term weight loss outcomes. *The American Journal of Medicine*, 124(11), 1082–1085. <https://doi.org/10.1016/j.amjmed.2011.04.016>

- Polirstok, S. R., & Greer, R. D. (1986). A replication of collateral effects and a component analysis of a successful tutoring package for inner-city adolescents. *Education and Treatment of Children*, 9(2), 101–121. <http://www.jstor.org/stable/42898957>
- Popkin, J., & Skinner, C. H. (2003). Enhancing academic performance in a classroom serving students with serious emotional disturbance: Interdependent group contingencies with randomly selected components. *School Psychology Review*, 32(2), 282–295. <https://doi.org/10.1080/02796015.2003.12086199>
- Pranay, K., & Chhetri, B. (2015). Evolution of World Wide Web: Journey from Web 1.0 to Web 4.0. *International Journal of Computer Science and Technology*, 6(1), 134–138.
- Prescott, J., Stodart, M., Becket, G., & Wilson, S. (2013). The Experience of using Facebook as an Educational Tool. *Health and Social Care Education*, December, 0–0. <https://doi.org/10.11120/hsce.2013.00033>
- Prescott, J., Stodart, M., Becket, G., & Wilson, S. (2015). The experience of using Facebook as an educational tool. *Health and Social Care Education* <https://doi.org/10.11120/hsce.2013.00033>
- Prue, D. M., & Fairbank, J. A. (1981). Performance feedback in organizational behavior management: A review. *Journal of Organizational Behavior Management*, 3(1), 1–16. https://doi.org/10.1300/J075v03n01_01
- Qi, C. H., Barton, E. E., Collier, M., & Lin, Y. L. (2018). A systematic review of single-case research studies on using video modeling interventions to improve social communication skills for individuals with autism spectrum disorder. *Focus on Autism and Other Developmental Disabilities*, 33(4), 249–257. <https://doi.org/10.1177/1088357617741282>
- Querol, B. I. D., Rosales, R., & Soldner, J. L. (2015). A comprehensive review of interteaching and its impact on student learning and satisfaction. *Scholarship of Teaching and Learning in Psychology*, 1(4), 390–411. <https://doi.org/10.1037/stl0000048>
- Qureshi, I. A., Raza, H., & Whitty, M. (2014). Facebook as e-learning tool for higher education institutes. *Knowledge Management & E-Learning*, 6(4), 440–448 <https://doi.org/10.34105/j.kmel.2014.06.029>

- Reed, D. D., & Kaplan, B. A. (2011). The matching law: a tutorial for practitioners. *Behavior analysis in practice*, 4(2), 15–24. <https://doi.org/10.1007/BF03391780>
- Reese, E. P. (2007). Teaching concepts and understanding the role of antecedents. In J. S. Vargas, *Behavior Analysis for Effective Teaching*, 2nd ed. (pp. 317-366). Routledge.
- Rescorla, L. (2002). Language and reading outcomes to age 9 in late- talking toddlers. *Journal of Speech, Language, and Hearing Research*, 45(2), 360-371. [https://doi.org/10.1044/1092-4388\(2002/028\)](https://doi.org/10.1044/1092-4388(2002/028))
- Reyna, V. F. (2004). Why scientific evidence? The importance of evidence in changing educational practice. In P. McCardle and V. Chhabra (Eds.), *The Voice of Evidence in Reading Research* (pp. 47-80). Paul H. Brookes.
- Reynolds, B. (2006). A review of delay-discounting research with humans: relations to drug use and gambling. *Behavioural Pharmacology*, 17(8), 651-667 <https://doi.org/10.1097/FBP.0b013e3280115f99>
- Rieken, C. J., Dotson, W. H., Carter, S. L., & Griffith, A. K. (2018). An evaluation of interteaching in an asynchronous online graduate-level behavior analysis course. *Teaching of Psychology*, 45(3), 264–269. <https://doi.org/10.1177/0098628318779275>
- Roberts, G. J., Mize, M., Reutebuch, C. K., Falcomata, T., Capin, P., & Steelman, B. L. (2019). Effects of a self-management with peer training intervention on academic engagement for high school students with autism spectrum disorder. *Journal of Behavioral Education*, 28(4), 456-478. <http://dx.doi.org/10.1007/s10864-018-09317-2>
- Roberts, M. S. & Semb, G. (1989). Student selection of deadline conditions in a personalized psychology course. *Teaching of Psychology*, 16, 128–130. https://doi.org/10.1207/s15328023top1603_6
- Roscoe, E. M., Fisher, W. W., Glover, A. C., & Volkert, V. M. (2006). Evaluating the relative effects of feedback and contingent money for staff training of stimulus preference assessments. *Journal of Applied Behavior Analysis*, 39(1), 63–77. <https://doi.org/10.1901/jaba.2006.7-05>

- Rosenbaum, M. S., & Drabman, R. S. (1979). Self-control training in the classroom: a review and critique. *Journal of Applied Behavior Analysis*, 12(3), 467–485. <https://doi.org/10.1901/jaba.1979.12-467>
- Rosenshine, B.V. (1986). Synthesis of Research on Explicit Teaching. *Educational Leadership*, 43, 60-69.
- Roseth, C. J., Johnson, D. W., & Johnson, R. T. (2008). Promoting Early Adolescents' Achievement and Peer Relationships: The Effects of Cooperative, Competitive, and Individualistic Goal Structures. *Psychological Bulletin*, 134(2), 223–246. <https://doi.org/10.1037/0033-2909.134.2.223>
- Royer, D. J., Lane, K. L., Cantwell, E. D., & Messenger, M. (2017). A systematic review of the evidence based for instructional choice in K–12 settings. *Behavioral Disorders*, 42, 89–107. <http://dx.doi.org/10.1177/0198742916688655>
- Salmoni, A. W., Schmidt, R. A., & Walter, C. B. (1984). Knowledge of results and motor learning: A review and critical reappraisal. *Psychological Bulletin*, 95(3), 355–386. <https://doi.org/10.1037/0033-2909.95.3.355>
- Samburgo, N. (September, 2017). *Token Economy Systems to Increase Appropriate Behaviors*. Classroom Management Series, National Association of Special Education Teachers. <https://www.naset.org/publications/classroom-management-series/token-economy-systems-to-increase-appropriate-behaviors>
- Saville, B. K., Bureau, A., Eckenrode, C., Fullerton, A., Herbert, R., Maley, M., Porter, A., & Zombakis, J. (2014). Interteaching and lecture: A comparison of long-term recognition memory. *Teaching of Psychology*, 41(4), 325-329. <https://doi.org/10.1177/0098628314549704>
- Saville, B. K., Cox, T., O'Brien, S., & Vanderveldt, A. (2011). Interteaching: the impact of lectures on student performance. *Journal of Applied Behavior Analysis*, 44(4), 937–941. <https://doi.org/10.1901/jaba.2011.44-937>
- Saville, B. K., Zinn, T. E., & Elliott, M. P. (2005). Interteaching Versus Traditional Methods of Instruction: A Preliminary Analysis. *Teaching of Psychology*, 32(3), 161–163. https://doi.org/10.1207/s15328023top3203_6

- Saville, B. K., Zinn, T. E., Neef, N. A., Norman, R. van, & Ferreri, S. J. (2006). A Comparison of Interteaching and Lecture in the College Classroom. *Journal of Applied Behavior Analysis*, 39(1), 49–61. <https://doi.org/10.1901/jaba.2006.42-05>
- Sawyer, M. R., Crosland, K. A., Miltenberger, R. G., & Rone, A. B. (2015). Using behavioral skills training to promote the generalization of parenting skills to problematic routines. *Child and Family Behavior Therapy*, 37(4), 261–284. <https://doi.org/10.1080/07317107.2015.1071971>
- Schieltz, K. M., & Wacker, D. P. (2020). Functional assessment and function-based treatment delivered via telehealth: A brief summary. *Journal of Applied Behavior Analysis*, 53(3), 1242–1258. <https://doi.org/10.1002/jaba.742>
- Selinske, J. E., Greer, R. D., & Lodhi, S. (1991). A functional analysis of the comprehensive application of behavior analysis to schooling. *Journal of Applied Behavior Analysis*, 24(1), 107–117. <https://doi.org/10.1901/jaba.1991.24-107>
- Sella, A. C., Ribeiro, D. M., & White, G. W. (2014). Effects of an online stimulus equivalence teaching procedure on research design open-ended questions performance of international undergraduate students. *The Psychological Record*, 64(1), 89–103. <https://doi.org/10.1007/s40732-014-0007-1>
- Sherman, J. G. (1992). Reflections on PSI: good news and bad. *Journal of Applied Behavior Analysis*, 25(1), 59–64. <https://doi.org/10.1901/jaba.1992.25-59>
- Shih, Y. & Reynolds, B. (2018). The effects of integrating goal setting and reading strategy instruction on English reading proficiency and learning motivation: A quasi-experimental study. *Applied Linguistics Review*, 9(1), 35-62. <https://doi.org/10.1515/applirev-2016-1022>
- Shimabukuro, S.M., Prater, M.A., Jenkins, A.A., & Edelen-Smith, P.J. (1999). The effects of self-monitoring of academic performance on students with learning disabilities and ADD/ADHD. *Education and Treatment of Children*, 22, 397-414.
- Shin, J., Guo, Q., & Gierl, M. J. (2019). Multiple-choice item distractor development using topic modeling approaches. *Frontiers in Psychology*, 10, 825. <https://doi.org/10.3389/fpsyg.2019.00825>

- Sides, J. D., & Cuevas, J. A. (2020). Effect of goal setting for motivation, self-efficacy, and performance in elementary mathematics. *International Journal of Instruction*, 13(4), 1–16. <https://doi.org/10.29333/iji.2020.1341a>
- Sidman M. (2009). Equivalence relations and behavior: An introductory tutorial. *The Analysis of Verbal Behavior*, 25(1), 5–17. <https://doi.org/10.1007/BF03393066>
- Sidman, M. (1971). Reading and auditory-visual equivalences. *Journal of Speech and Hearing Research*, 14,5–13. <https://doi.org/10.1044/jshr.1401.05>
- Sidman, M., & Cresson, O. (1973). Reading and crossmodal transfer of stimulus equivalences in severe retardation. *American Journal of Mental Deficiency*, 77(5), 515–523.
- Sidman, M., & Tailby, W. (1982). Conditional discrimination vs. matching to sample: An expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior*, 37(1), 5–22. <https://doi.org/10.1901/jeab.1982.37-5>
- Simonsen, B., Fairbanks, S., Briesch, A., Myers, D., & Sugai, G. (2008). Evidenced-based practices in classroom management: Considerations for research to practice. *Education and Treatment of Children*, 31, 351–380.
- Simonsen, B., Fairbanks, S., Briesch, A., Myers, D., & Sugai, G. (2008). Evidence- based practices in classroom management: Considerations for research to practice. *Education and Treatment of Children*, 31(3), 351–380. <https://doi.org/10.1353/etc.0.0007>
- Singer-Dudek, J., Speckman, J., & Nuzzolo, R. (2010). A comparative analysis of the CABAS® model of education at the Fred S. Keller School: A twenty-year review. *The Behavior Analyst Today*, 11(4), 253–265. <https://doi.org/10.1037/h0100705>
- Sivaraman, M., Virues-Ortega, J., Maes-Dhaeyere, L., Saelaert, E., & Roeyers, H. (2022). Social referencing training in children with autism spectrum disorder: A randomized controlled study. Accepted for publication at *Journal of Applied Behavior Analysis* 55(3):763-781. <https://doi.org/10.1002/jaba.935>
- Skerbetz, M. D., & Kostewicz, D. E. (2013). Academic choice for included students with emotional and behavioral disorders. *Preventing School Failure*, 57, 212–222. <http://dx.doi.org/10.1080/1045988X.2012.701252>

- Skinner, B. F. (1938). *The Behavior of Organism: An Experimental Analysis*. Appleton-Century-Crofts.
- Skinner, B. F. (1986). The Evolution of Verbal Behavior. *Journal of the Experimental Analysis of Behavior*, 45(1), 115–122. <https://doi.org/10.1901/jeab.1986.45-115>
- Slavin, R. E. (1995). A model of effective instruction. *Educational Forum*, 59(2), 166–176. <https://doi.org/10.1080/00131729509336383>
- Sleeman, M., Friesen, M., Tyler-Merrick, G., & Walker, L. (2021). The effects of precision teaching and self-regulated learning on early multiplication fluency. *Journal of Behavioral Education*, 30(2), 149–177. <https://doi.org/10.1007/s10864-019-09360-7>
- Slocum, T. A., & Rolf, K. R. (2021). Features of Direct Instruction: Content Analysis. *Behavior analysis in practice*, 14(3), 775–784. <https://doi.org/10.1007/s40617-021-00617-0>
- Southall, C. M., & Gast, D. L. (2011). Self-Management procedures: a comparison across the autism spectrum. *Education and Training in Autism and Developmental Disabilities*, 46(2), 155–171. <http://www.jstor.org/stable/23879688>
- Spring, B. (2007). Evidence-based practice in clinical psychology: What it is, why it matters; what you need to know. *J. Clin. Psychol.* 63, 611–631. doi: 10.1002/jclp.20373
- Springer, Crystal R, & Pear, Joseph J. (2008). Performance measures in courses using computer-aided personalized system of instruction. *Computers and Education*, 51(2), 829-835.
- States, J., Detrich, R. & Keyworth, R. (2019). Active Student Responding (ASR) Overview. Oakland, CA: The Wing Institute. <https://www.winginstitute.org/instructional-delivery-student-respond>
- Statista (2021) Average daily time spent by users worldwide on mobile social media apps from October 2020 to March 2021, by age group (in minutes). Retrieved from <https://www.statista.com/statistics/1272883/worldwide-social-apps-time-spent-daily-age/>
- Statista. (2022). Most popular social networks worldwide as of January 2022, ranked by number of monthly active users. Retrieved from <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>

- Stockard, J., Wood, T. W., Coughlin, C., & Rasplika Khoury, C. (2018). The effectiveness of direct instruction curricula: A meta-analysis of a half century of research. *Review of Educational Research, 88*(4), 479–507. <https://doi.org/10.3102/0034654317751919>
- Sturmey, P., Dalfen, S., & Fienup, D. M. (2015). Inter-teaching: A systematic review. *European Journal of Behavior Analysis, 16*(1), 121–130. <https://doi.org/10.1080/15021149.2015.1069655>
- Styles, B., & Torgerson, C. (2018). Randomised controlled trials (RCTs) in education research –methodological debates, questions, challenges. *Education Research, 60*(3), 255–264. <https://doi.org/10.1080/00131881.2018.1500194>
- Sugai, G., & Horner, R. (2002). The evolution of discipline practices: School-wide positive behavior supports. *Child & Family Behavior Therapy, 24*(1-2), 23–50. https://doi.org/10.1300/J019v24n01_03
- Sulzer-Azaroff, B., & Gillat, A. (1990). Trends in behavior analysis in education. *Journal of Applied Behavior Analysis, 23*(4), 491–495. <https://doi.org/10.1901/jaba.1990.23-491>
- Sutherland, K. S., & Wehby, J. H. (2001). Exploring the relationship between increased opportunities to respond to academic requests and the academic and behavioral outcomes of students with EBD: A Review. *Remedial and Special Education, 22*(2), 113–121. <https://doi.org/10.1177/074193250102200205>
- Sweeney, W. J., Salva, E., Cooper, J. O., & Talbert-Johnson, C. (1993). Using self-evaluation to improve difficult-to-read handwriting of secondary students. *Journal of Behavioral Education, 3*, 427–443. <https://doi.org/10.1007/BF00961545>
- Sweeney, W.J., Salva, E., Cooper, J.O., & Talbert-Johnson, C. (1993). Using self-evaluation to improve difficult-to-read handwriting of secondary students. *Journal of Behavioral Education, 3*, 427–443. <https://doi.org/10.1007/BF00961545>
- Tackett, J. L., Brandes, C. M., King, K. M., & Markon, K. E. (2019). Psychology's replication crisis and clinical psychological science. *Annual Review of Clinical Psychology, 15*, 579–604. <https://doi.org/10.1146/annurev-clinpsy-050718-095710>

- Taveggia, T. C. (1976). Personalized instruction: A summary of comparative research, 1967–1974. *American Journal of Physics*, 44(11), 1028–1033. <https://doi.org/10.1119/1.10579>
- Terenzini, P. T., & Pascarella, E. T. (1991). Twenty Years of Research on College Students: Lessons for Future Research. *Research in Higher Education*, 32(1), 83–92. <http://www.jstor.org/stable/40195990>
- Tobin, K. G. (2004). The effects of beginning reading instruction in the “horizons” reading program on the reading skills of third and fourth Graders. *Journal of Direct Instruction*, 4(2), 4(2): 129-137.
- Tomlinson, S.R., Gore, N., & McGill, P. (2018). Training individuals to implement applied behavior analytic procedures via telehealth: A systematic review of the literature. *Journal of Behavioral Education*, 27, 172-222. <https://doi.org/10.1007/s10864-018-9292-0>
- van Ryzin, M. J., Roseth, C. J., & Biglan, A. (2020). Mediators of effects of cooperative learning on prosocial behavior in middle school. *International Journal of Applied Positive Psychology*, 5(1–2), 37–52. <https://doi.org/10.1007/s41042-020-00026-8>
- Vanselow, N. R., & Bourret, J. C. (2012). Online interactive tutorials for creating graphs with excel 2007 or 2010. *Behavior Analysis in Practice*, 5(1), 40–46. <https://doi.org/10.1007/BF03391816>
- Viel-Ruma, K., Houchins, D. E., Jolivette, K., Fredrick, L. D., & Gama, R. (2010). Direct Instruction in Written Expression: The effects on English speakers and English language learners with disabilities. *Learning Disabilities Research & Practice*, 25(2), 97–108. <https://doi.org/10.1111/j.1540-5826.2010.00307.x>
- Virues-Ortega, J. (May, 2022). *From Standardized Measurements to Biomarkers: The Emerging Science of Treatment Outcomes in ABA and Autism Research* (Chair: Dr. Susan M. Wilczynski). Practice Board Invited Address, 48th Annual Convention of the Association for Behavior Analysis International, Boston.
- Virues-Ortega, J., Arias-Higuera, M., Hurtado-Parrado, C., & Iwata, B. A. (2021). Nathan H. Azrin: A case study in research translation in behavior science. *Perspectives on Behavior Science*, 44, 41–67. <https://doi.org/10.1007/s40614-020-00278-4>

- Virues-Ortega, J., Arias-Higuera, M., Hurtado-Parrado, C., & Iwata, B. A. (2021). Nathan H. Azrin: A case study in research translation in behavior science. *Perspectives on Behavior Science*, 44, 41-67 <https://doi.org/10.1007/s40614-020-00278-4>
- Virues-Ortega, J., Delgado Casas, C., Martin, N., Tarifa-Rodriguez, A., Reina Hidalgo, A. J., & Cox, A. D., & Navarro Guzmán, J. I. (2022b). Interobserver agreement in software-aided and paper-and-pencil behavioral observation. *Behavior Research Methods*. <https://doi.org/10.3758/s13428-022-01861-0>
- Virues-Ortega, J., Moeyaert, M., Sivaraman, M., Tarifa-Rodriguez, & Fernandez Castilla, B. (2022a). Quantifying Outcomes in Applied Behavior Analysis through Visual and Statistical Analyses: A Synthesis. In J. L. Matson, *Applied Behavior Analysis: A Comprehensive Handbook*. SpringerNature.
- Virues-Ortega, J., Rodríguez, V., & Yu, C. T. (2013). Prediction of treatment outcomes and longitudinal analysis in children with autism undergoing intensive behavioral treatment. *International Journal of Clinical and Health Psychology*, 13, 91-100. [https://doi.org/10.1016/S1697-2600\(13\)70012-7](https://doi.org/10.1016/S1697-2600(13)70012-7)
- Virues-Ortega, J., Taylor, S., Phillips, K., & McCormack, J. C. (2022). Functional assessment in applied behavior analysis. In J. L. Matson (Ed.), *Handbook of ABA for Children with Autism: Clinical Guide to Assessment and Treatment*. SpringerNature.
- Vostanis, A., Padden, C., Chiesa, M., Rizos, K., & Langdon, P. E. (2021). A precision teaching framework for improving mathematical skills of students with intellectual and developmental disabilities. *Journal of Behavioral Education*, 30(4), 513–533. <https://doi.org/10.1007/s10864-020-09394-2>
- Wacker, D. P., Lee, J. F., Padilla Dalmau, Y. C., Kopelman, T. G., Lindgren, S. D., Kuhle, J., Pelzel, K. E., Dyson, S., Schieltz, K. M., & Waldron, D. B. (2013). Conducting functional communication training via telehealth to reduce the problem behavior of young children with autism. *Journal of Developmental and Physical Disabilities*, 25(1), 35–48. <https://doi.org/10.1007/s10882-012-9314-0>
- Wacker, D. P., Schieltz, K. M., Berg, W. K., Harding, J. W., Dalmau, Y., & Lee, J. F. (2017). The long-term effects of functional communication training conducted in young children's

- home settings. *Education & treatment of children*, 40(1), 43–56. <https://doi.org/10.1353/etc.2017.0003>
- Wainer, A. L., & Ingersoll, B. R. (2015). Increasing access to an ASD imitation intervention via a telehealth parent training program. *Journal of Autism and Developmental Disorders*, 45(12), 3877–3890. <https://doi.org/10.1007/s10803-014-2186-7>
- Walker, B. D., & Rehfeldt, R. A. (2012). An evaluation of the stimulus equivalence paradigm to teach single-subject design to distance education students via Blackboard. *Journal of Applied Behavior Analysis*, 45(2), 329–344. <https://doi.org/10.1901/jaba.2012.45-329>
- Walker, B. D., Shippen, M. E., Houchins, D. E., & Cihak, D. F. (2007). Improving the writing skills of high school students with learning disabilities using the Expressive Writing program. *International Journal of Special Education*, 22(2), 67–77. <https://doi.org/10.1111/j.1540-5826.2005.00131.x>
- Ward-Horner, J., & Sturmey, P. (2010). Component analyses using single-subject experimental designs: a review. *Journal of Applied Behavior Analysis*, 43(4), 685–704. <https://doi.org/10.1901/jaba.2010.43-685>
- Ward-Horner, J., & Sturmey, P. (2010). Component analyses using single-subject experimental designs: a review. *Journal of Applied Behavior Analysis*, 43(4), 685–704. <https://doi.org/10.1901/jaba.2010.43-685>
- Ward-horner, John & Sturmey, & Peter. (2012). Component Analysis of Behavior Skills. *Behavioral Interventions*, 27, 75–92. <https://doi.org/10.1002/bin>
- Warzak, W. J., & Page, T. J. (1990). Teaching refusal skills to sexually active adolescents. *Journal of Behavior Therapy and Experimental Psychiatry*, 21(2), 133–139. [https://doi.org/10.1016/0005-7916\(90\)90018-G](https://doi.org/10.1016/0005-7916(90)90018-G)
- Watkins, C. L., & Slocum, T. A. (2004). The Components of Direct Instruction. *Journal of Direct Instruction*, 3(2), 75–110.
- Watkins, C.L. (1997). *Project Follow Through: A case study of the contingencies influencing instructional practices of the educational establishment*. (Monograph). Concord, MA: Cambridge Center for Behavioral Studies.

- Weatherly, N. L., & Malott, R. W. (2008). An analysis of organizational behavior management research in terms of the three-Contingency model of performance management. *Journal of Organizational Behavior Management*, 28(4), 260–285. <https://doi.org/10.1080/01608060802454643>
- Weigandt, S., (2021). *To Write or Type? A Comparison of Flashcard Creation Methods on College Students' Exam Scores (29096020)* [Doctoral Dissertation, Western Michigan University]. ProQuest. <https://www.proquest.com/openview/99ef7e242faefa03791e51a7e-a9bb4db/1?pq-origsite=gscholar&cbl=18750&diss=y>
- West, R. P., Young, K. R., & Spooner, F. (1990). Precision Teaching: An Introduction. *Teaching Exceptional Children*, 22(3), 4-9. <https://doi.org/10.1177/004005999002200301>
- What Works Clearinghouse (2017). *Key Criteria Used in WWC Reviews of Single-Case Design Research [version 1.0]*. Author. <https://ies.ed.gov/ncee/wwc/Document/264>
- What Works Clearinghouse (2020). *Standards handbook (version 4.1)*. Washington, DC: Institute of Education Sciences.
- Wheeler, J. J., Mayton, M. R., Carter, S. L., Chitiyo, M., Menendez, A. L., & Huang, A. (2009). An assessment of treatment integrity in behavioral intervention studies conducted with persons with mental retardation. *Education and Training in Developmental Disabilities*, 44(2), 187-195. <https://www.jstor.org/stable/24233493>
- Wirth, K. M. (2004). *A Study of Rules Designed to Increase Peer-Review Accuracy in a Computer Aided Personalized System of Instruction Course* [Master's thesis, University of Manitoba]. https://mspace.lib.umanitoba.ca/bitstream/handle/1993/17920/Wirth_A_study.pdf?sequence=1
- Wolniczak, I., Cáceres-DelAguila, J. A., Palma-Ardiles, G., Arroyo, K. J., Solís-Visscher, R., Paredes-Yauri, S., Mego-Aquije, K., & Bernabe-Ortiz, A. (2013). Association between Facebook dependence and poor sleep quality: a study in a sample of undergraduate students in Peru. *PLoS one*, 8(3), e59087. <https://doi.org/10.1371/journal.pone.0059087>
- Woodley, C., & Meredith, C. A. (2012). Supporting student transition through social media. *American Journal of Distance Education*, 26(2), 86–95. <https://doi.org/10.1080/08923647.2012.655055>

- Yakubova, G., Hughes, E. M., & Chen, B. B. (2020). Teaching students with ASD to solve fraction computations using a video modeling instructional package. *Research in developmental disabilities, 101*, 103637. <https://doi.org/10.1016/j.ridd.2020.103637>
- Zamzow, R. (2022). Why autism therapies have an evidence problem. *Spectrum*. <https://doi.org/10.53053/SJUS7289>
- Zaragoza Scherman, A., Thomson, K., Boris, A., Dodson, L., Pear, J. J., & Martin, G. (2015). Online training of Discrete-Trials Teaching for educating children with autism spectrum disorders: A preliminary study. *Journal on Developmental Disabilities, 21*(1), 22–33.
- Zayac, R., & Paulk, A. L. (2013). Interteaching: Its effects on exam scores in a compressed-schedule format. *Journal of the Scholarship of Teaching and Learning, 14*(1), 1–12. <https://doi.org/10.14434/josotl.v14i1.3649>
- Zhu, J., Bruhn, A., Yuan, C., & Wang, L. (2021). Comparing the effects of videoconference and email feedback on treatment integrity. *Journal of Applied Behavior Analysis, 54*(2), 618–635. <https://doi.org/10.1002/jaba.810>

ANEXOS

Anexo A. Modelo de consentimiento informado



HOJA DE INFORMACIÓN PARA EL PARTICIPANTE

RESPONSABLES

Dr. Javier Virués Ortega (director)
javier.virues-ortega@uam.es

Da Aida Tarifa Rodríguez (estudiante de doctorado responsable)
aida.tarifa@estudiante.uam.es

DATOS DE LA INVESTIGACIÓN

Título del proyecto: Análisis experimental de la comunicación mediada por ordenador en la enseñanza de habilidades terapéuticas mediante procedimientos analítico-conductuales.

Introducción

Nos dirigimos a usted para informarle sobre un estudio de investigación en el que se le invita a participar. En esta hoja puede encontrar información sobre el estudio para que pueda evaluar y juzgar si quiere o no participar. Le aclararemos cualquier duda que pueda tener sobre el mismo y quedamos a su disposición para que en cualquier momento nos pueda contactar.

Duración y naturaleza de la participación

La participación en el estudio se prolongará por idéntica duración que la del programa de especialización profesional online en el que participa (un año). Si confirma su participación es posible

que le invitamos a algunas sesiones adicionales de formación para facilitar una participación más activa en el curso, una mejor elaboración y memorización del material de estudio. No todos los participantes formarán parte de estas sesiones y los que lo harán podrán hacerlo en distinto número de ellas según las demandas del diseño del estudio. Estimamos que los participantes podrán participar en una media de 5 horas de sesiones adicionales. Estas sesiones serán personalizadas o en grupo y las llevará un instructor bien formado. Esperamos que estas sesiones tengan un impacto positivo en tu rendimiento en el curso online de especialización profesional en el que participas.

Participación voluntaria y compensación económica

Debe saber que su participación en este estudio es voluntaria y que puede decidir no participar en cualquier momento, sin que por ello se altere la relación con el centro de formación clínica del que está recibiendo servicios, ni se produzca perjuicio alguno. Si en algún momento decide cesar su participación, le rogamos que nos lo comunique. En ese momento se dejarán de recopilarse y utilizarse sus datos con fines de investigación. Además, debe saber que no recibirá compensación económica por su participación y que ésta no le supondrá ningún gasto adicional con respecto de los servicios de formación que recibe.

Descripción general del estudio

Como parte del presente proyecto de investigación queremos evaluar durante todo el curso académico las diferentes estrategias de aprendizaje con estudiantes de especialización profesional vinculados a cursos reglados durante la enseñanza online. Los estudios propuestos se realizarán con la participación de estudiantes que son partícipes en cursos de especialización profesional en el área del análisis aplicado de conducta. Tan sólo tendrán que incorporarse mediante un enlace a un grupo de estudio cerrado creado a través de la plataforma de Facebook Groups.

El objetivo de la línea de investigación es evaluar los elementos de interactividad de estudiantes en un grupo de especialización profesional de enseñanza online y su impacto en el rendimiento. En concreto, exploraremos los siguientes objetivos:

1. Análisis de la interactividad entre los participantes del grupo de Facebook y su relación con el rendimiento de pruebas objetivas semanales en programas de formación online.

2. Análisis de la interactividad entre los miembros de un grupo cerrado de Facebook como un elemento activo en la efectividad docente de la enseñanza online.
3. Análisis de los componentes educativos empleados y el impacto en el rendimiento de los participantes.

Para que este proyecto pueda llevarse a cabo, se hace imprescindible la colaboración de personas que quieran participar en el programa y, al mismo tiempo, nos permitan analizar los resultados de su participación en este.

Este proyecto se llevará a cabo de acuerdo con los principios éticos que tienen su origen en la Declaración de Helsinki y siguiendo las buenas prácticas clínicas según la Ley 14/2007 de Investigación Biomédica.

Uso de datos personales

Usted tiene derecho a ser informado de cualquier registro que se realice en su participación del curso online _____. Este documento desea ayudarle a comprender todas las cuestiones relativas a los registros que, dado el caso, se realizarían de su actividad mientras participa en el curso. Léalo atentamente y consulte con los investigadores todas las dudas que se le planteen. Le recordamos que, por imperativo legal, usted tendrá que firmar el consentimiento informado para que podamos realizar dichos registros y, en caso de que no dé su consentimiento, no se obtendrán dichos datos con fines de investigación en el contexto del presente proyecto. La participación en este estudio es completamente voluntaria y su negativa a participar no conllevará la pérdida de ningún servicio. No utilizaremos datos personales durante el estudio. Cada participante estará identificado mediante un código alfanumérico. Solo los investigadores responsables tendrán acceso a un fichero que identifique cada código con el nombre del participante.

Procedimiento de toma de datos

Los datos se tomarán de forma automática de aquellos estudiantes que hayan expresado su consentimiento extrayendo los logs correspondientes que quedan registrados de su actividad en la plataforma del curso y sus grupos de estudio de red social asociados. No debe hacer nada diferente a lo que hace en su participación en el curso para permitir la toma de estos datos.

Cómo se preserva la confidencialidad

Los datos que se obtengan de su participación serán utilizados únicamente con fines de investigación y se utilizará un pseudónimo con el fin de garantizar su privacidad. Solamente el investigador principal guardará sus datos personales en un lugar seguro (disco duro local con clave de acceso) de tal manera que ninguna persona ajena pueda acceder a esta información y atendiendo a un estricto cumplimiento de la Ley Orgánica 15/1999 sobre la Protección de Datos de Carácter Personal.

Al realizar el análisis de datos, únicamente se utilizará la información sobre sus datos socio - demográficos (edad, sexo, nivel educativo y situación laboral), de rendimiento académico y de interacción con funcionalidades online del programa y en ningún caso se harán públicos sus datos personales, siempre garantizando la plena confidencialidad de los datos y el riguroso cumplimiento del secreto profesional en el uso y manejo de la información y el material obtenidos. Al transcurrir un periodo de cuatro años, los datos personales serán completamente destruidos.

Modo de reclutamiento

Hemos contactado con el director del curso de especialización profesional que está realizando, quien gentilmente ha accedido a facilitarle la presente información sobre el mismo. Una vez Vd. nos haga llegar el presente documento firmado por correo electrónico nos pondremos en contacto con Vd. para confirmarle su participación en el proyecto y darle las instrucciones que procedan.

Revocación del consentimiento

Si, en el caso de consentir inicialmente en participar, en algún momento del estudio usted desea dejar de colaborar en nuestra investigación, le rogamos que nos lo comunique e inmediatamente se dejarán de registrar sus datos y se destruirán los ficheros del estudio existentes que contengan información sobre su actividad. Esto no supondrá en ningún caso perjuicio alguno para su correcto seguimiento del curso.

Con quién puedo contactar para cualquier duda

Director del proyecto:

Dr. Javier Virués Ortega

Investigador Ramón y Cajál adscrito al Dpto. de Psicología Biológica y de la Salud Universidad Autónoma de Madrid

javier.virues-ortega@uam.es

Estudiante de doctorado responsable:

D^a Aida Tarifa Rodríguez

Universidad Autónoma de Madrid

aida.tarifa@estudiante.uam.es

HOJA DE CONSENTIMIENTO**Declaro:**

1. Que he leído y entendido toda la información oral y escrita en relación con la participación en el citado proyecto.
2. Se me ha entregado una copia de la Hoja de Información al Participante y una copia de este Consentimiento Informado, fechado y firmado. Se me han explicado las características y el objetivo del estudio.
3. He tenido la oportunidad de debatir y preguntar sobre dicha información y he recibido las respuestas adecuadas por parte de alguno de los miembros del equipo investigador encargado de este estudio.
4. Se me ha asegurado que se mantendrá la confidencialidad de mis datos personales y que, transcurridos cuatro años, éstos serán destruidos.
5. El consentimiento lo otorgo de manera voluntaria, consciente de que soy libre de retirarme del estudio en cualquier momento de este, por cualquier razón y sin que tenga que dar explicación alguna o me atenga a perjuicio alguno por ello.
6. Doy mi consentimiento a la utilización de los datos obtenidos en congresos y/o publicaciones. La Dirección de la investigación garantiza el cumplimiento de la Ley Orgánica 15/1999 de 13 de diciembre que toda la información que se obtenga será utilizada exclusivamente para la consecución de los objetivos del proyecto de investigación,

asegurándose la más absoluta confidencialidad de los datos de los participantes en el estudio.

En Madrid, a _____ de _____ de 202_____

Consiento:

No consiento:

Don/Dña. _____

Don/Dña. _____

(El/la participante)

(El/la participante)

Yo, Don/Dña. _____, investigador del proyecto, declaro que he discutido el contenido de este documento de consentimiento informado con el /la arriba firmante.

Firma del investigador/a:

Don/Dña. _____

Anexo B. Accuracy of paper-and-pencil Systematic Observation versus Computer-Aided systems.

SUPPLEMENTARY ONLINE MATERIAL

Virues-Ortega et al.

Table A Common Attributes and Functionalities of Software-Based Observation Applications

	The Observer XT	Big Eye Observer	Countee	Solomon Coder
Platform	PC	iOSX	Android, iOSX	PC
Behavior dimensions	Frequency, duration	Frequency, duration, partial interval, total interval	Frequency, duration	Frequency, durations, percentage, latencies, alternations
Main application	In video	In vivo	In vivo	In video
Clutter-free inter-phase	No	Yes	Yes	Yes
Video playback functions	Yes	No	No	Yes
In-session error correction	Yes	Yes	No	Yes
Session pausing	Yes	Yes	Yes	Yes
One-key/tap recording	Yes ¹	Yes	Yes	Yes
One-screen recording	Yes	Yes	Yes	Yes

Note. (1) Depending on the program configuration more than one keystroke may be needed to record an event.

Table B Presence of Time-Dependent Trends in in the Three Observation Methods

	Dickey-Fuller Test			Trend Factor		
	Z(t)	5% Critical value	p value	Coefficient ±SE	t	p value
P&P	-3.26	-3.60	.072	0.001±0.001	0.56	.584
OXT	-2.76	-3.60	.211	0.004±0.003	1.55	.144
BEO	-3.43	-3.60	.005	-0.002±0.001	-1.22	.243

Notes. All regression analyses assume an autoregressive factor of 1. BEO = Big Eye Observer, OXT = The Observer XT, P&P = paper-and-pencil; SE = Standard error.

Anexo C. Quantitative Indices of Student Social Media Engagement in Tertiary Education: A Systematic Review and a Taxonomy

Section and Topic	Item #	Checklist item	Location where item is reported
TITLE			
Title	1	Identify the report as a systematic review.	Y
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	Y
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	Y
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	Y
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	Y
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	Y
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	Y
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	Y
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	Y
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	Y
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	Y

Section and Topic	Item #	Checklist item	Location where item is reported
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	Y
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	N/A
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	Y
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	Y
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	Y
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	N/A
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	N/A
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	N/A
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	N/A
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	N/A
RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	Y
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	Suppl. files
Study characteristics	17	Cite each included study and present its characteristics.	Y
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	N/A
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	Y

Section and Topic	Item #	Checklist item	Location where item is reported
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	N/A
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	N/A
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	N/A
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	N/A
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	N/A
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	N/A
DISCUSSION			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	Y
	23b	Discuss any limitations of the evidence included in the review.	Y
	23c	Discuss any limitations of the review processes used.	Y
	23d	Discuss implications of the results for practice, policy, and future research.	Y
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information, including register name and registration number, or state that the review was not registered.	Not registered (1)
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	Not registered (1)
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	Not registered (1)
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	Title page
Conflict of interest	26	Declare any competing interests of review authors.	Title page
Availability of data	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	Figshare (2) see also suppl. files

Notes. (1) Existing systematic review registries like PROSPERO do not accept scoping reviews, literature reviews or mapping reviews. (2) The dataset will be made available via Figshare once the manuscript is accepted for publication.

Section and Topic	Item #	ABSTRACT Checklist Items	Reported (Yes/No)
TITLE			
Title	1	Identify the report as a systematic review.	Y
BACKGROUND			
Objectives	2	Provide an explicit statement of the main objective(s) or question(s) the review addresses.	Y
METHODS			
Eligibility criteria	3	Specify the inclusion and exclusion criteria for the review.	Y
Information sources	4	Specify the information sources (e.g. databases, registers) used to identify studies and the date when each was last searched.	Y
Risk of bias	5	Specify the methods used to assess risk of bias in the included studies.	Y
Synthesis of results	6	Specify the methods used to present and synthesise results.	N/A
RESULTS			
Included studies	7	Give the total number of included studies and participants and summarise relevant characteristics of studies.	Y
Synthesis of results	8	Present results for main outcomes, preferably indicating the number of included studies and participants for each. If meta-analysis was done, report the summary estimate and confidence/credible interval. If comparing groups, indicate the direction of the effect (i.e. which group is favoured).	N/A
DISCUSSION			
Limitations of evidence	9	Provide a brief summary of the limitations of the evidence included in the review (e.g. study risk of bias, inconsistency and imprecision).	Y
Interpretation	10	Provide a general interpretation of the results and important implications.	Y
OTHER			
Funding	11	Specify the primary source of funding for the review.	Y
Registration	12	Provide the register name and registration number.	Not registered (1)

Note. (1) Existing systematic review registries like PROSPERO do not accept scoping reviews, literature reviews or mapping reviews.

SUPPLEMENTARY ONLINE MATERIAL

Table A Implementation of Inclusion Criteria

Reference	Archival Reference	Criterion 1	Criterion 2	Criterion 3	Status
Abd-El-Aal et al. (2017)	File 1, no. 99	Met	Met	Met	Included
Abdullah et al. (2019)	File 2, no. 83	Met	Not met	Met	Not included
Abe et al. (2013)	File 8, no. 31	Not met	Not met	Not met	Not included
Abney et al. (2019)	File 3, no. 49	Met	Met	Met	Included
Abramo (2016)	File 7, no. 47	Not met	Not met	Not met	Not included
Abu-Alruz (2014)	File 4, no. 38	Met	Not met	Not met	Not included
Abulibdeh (2013)	File 6, no. 91	Met	Met	Not met	Not included
AbuSa'aleek (2015)	File 2, no. 72	Met	Not met	Met	Not included
Acun (2020)	File 1, no. 83	Met	Not met	Met	Not included
Acun (2020)	File 1, no. 90	Met	Not met	Met	Not included
Adhikary et al. (2018)	File 2, no. 3	Not met	Not met	Not met	Not included
Agbo et al. (2020)	File 1, no. 67	Met	Not met	Met	Not included
Agozzino et al. (2014)	File 8, no. 60	Not met	Not met	Not met	Not included
Aharony (2014)	File 4, no. 42	Met	Not met	Not met	Not included
Ahmed et al. (2017)	File 4, no. 6	Met	Not met	Not met	Not included
Akcaoglu et al. (2018)	File 2, no. 94	Met	Met	Met	Included
Akman et al. (2018)	File 6, no. 47	Met	Not met	Not met	Not included
Aksoy (2018)	File 4, no. 5	Met	Not met	Not met	Not included
Al-Azawei (2019)	File 1, no. 100	Met	Met	Met	Included
Al-Bahrani et al. (2017)	File 7, no. 39	Not met	Not met	Not met	Not included
Al-Dheleai et al. (2017)	File 1, no. 94	Met	Not met	Met	Not included
Al-marayeah et al. (2017)	File 4, no. 82	Met	Not met	Not met	Not included
Al-Qaysi et al. (2020)	File 2, no. 100	Met	Not met	Met	Not included
Alam et al. (2019)	File 1, no. 23	Met	Met	Not met	Not included
Albayrak et al. (2015)	File 4, no. 97	Met	Met	Met	Included
Alberth (2019)	File 2, no. 91	Met	Met	Met	Included
Aldahdouh et al. (2020)	File 3, no. 79	Not met	Not met	Not met	Not included
Alden (2017)	File 7, no. 14	Not met	Not met	Not met	Not included
AlGhamdi et al. (2018)	File 3, no. 83	Met	Not met	Not met	Not included
Alghamdi et al. (2019)	File 4, no. 19	Met	Met	Not met	Not included
Alghazo et al. (2018)	File 4, no. 77	Met	Met	Met	Included
Alhadabi et al. (2020)	File 1, no. 15	Met	Not met	Met	Not included

Reference	Archival Reference	Criterion 1	Criterion 2	Criterion 3	Status
Alharbi et al. (2018)	File 3, no. 73	Met	Not met	Not met	Not included
Alillaiti (2018)	File 3, no. 56	Not met	Not met	Not met	Not included
Aljehani (2019)	File 2, no. 37	Met	Not met	Met	Not included
Allen (2013)	File 6, no. 97	Not met	Not met	Not met	Not included
Alm (2015)	File 2, no. 31	Met	Not met	Met	Not included
Alon et al. (2014)	File 8, no. 79	Met	Met	Not met	Not included
Alruwaili et al. (2020)	File 1, no. 57	Met	Not met	Not met	Not included
Alsaleem (2018)	File 1, no. 78	Met	Met	Met	Included
Alt (2017a)	File 6, no. 17	Met	Not met	Not met	Not included
Alt (2017b)	File 6, no. 73	Met	Not met	Not met	Not included
Altin et al. (2018)	File 2, no. 85	Met	Not met	Met	Not included
Altunkaya et al. (2018)	File 1, no. 79	Met	Met	Met	Included
Amador et al. (2017)	File 3, no. 8	Met	Not met	Not met	Not included
Amadu et al. (2018)	File 3, no. 98	Met	Not met	Not met	Not included
Amin et al. (2016)	File 1, no. 82	Met	Not met	Met	Not included
Anderson (2019)	File 2, no. 15	Not met	Not met	Not met	Not included
Anderson et al. (2014)	File 8, no. 54	Met	Met	Not met	Not included
Apostel (2015)	File 7, no. 3	Met	Not met	Not met	Not included
Arabacioglu (2014)	File 2, no. 92	Met	Met	Met	Included
Aragon et al. (2012)	File 3, no. 65	Met	Not met	Not met	Not included
Arendale (2017)	File 4, no. 96	Not met	Met	Not met	Not included
Arias Chávez et al. (2018)	File 1, no. 14	Not met	Not met	Not met	Not included
Armstrong et al. (2017)	File 6, no. 16	Met	Not met	Not met	Not included
Arquero et al. (2017)	File 6, no. 99	Met	Met	Not met	Not included
Arshad et al. (2018)	File 3, no. 69	Met	Not met	Not met	Not included
Arslantas et al. (2019)	File 2, no. 70	Met	Not met	Met	Not included
Arteaga et al. (2019)	File 1, no. 35	Met	Not met	Met	Not included
Aslan et al. (2018)	File 3, no. 86	Met	Not met	Not met	Not included
Atkins et al. (2017)	File 1, no. 9	Not met	Not met	Met	Not included
Atwong (2015)	File 7, no. 1	Not met	Met	Not met	Not included
Aubry (2013)	File 4, no. 40	Met	Not met	Not met	Not included
Auxier (2020)	File 2, no. 22	Not met	Not met	Met	Not included
Bacile (2013)	File 8, no. 89	Met	Met	Met	Included
Bagarukayo (2018)	File 4, no. 66	Met	Met	Not met	Not included
Bagarukayo et al. (2016)	File 2, no. 5	Met	Met	Not met	Not included

Reference	Archival Reference	Criterion 1	Criterion 2	Criterion 3	Status
Bagci et al. (2020)	File 1, no. 70	Met	Not met	Not met	Not included
Bahati (2015)	File 2, no. 24	Met	Met	Met	Included
Baim (2016)	File 7, no. 9	Met	Met	Not met	Not included
Baim (2017)	File 4, no. 31	Not met	Not met	Not met	Not included
Bajko et al. (2016)	File 4, no. 26	Met	Met	Met	Included
Baker (2013)	File 5, no. 15	Met	Not met	Not met	Not included
Baker et al. (2016)	File 3, no. 52	Met	Not met	Not met	Not included
Bal et al. (2015)	File 7, no. 30	Met	Met	Not met	Not included
Balakrishnan et al. (2017)	File 7, no. 65	Met	Not met	Not met	Not included
Balcikanli (2015)	File 2, no. 56	Met	Met	Met	Included
Baltaci (2019)	File 2, no. 81	Met	Not met	Met	Not included
Balzhiser et al. (2011)	File 8, no. 86	Met	Met	Not met	Not included
Bamberger et al. (2020)	File 3, no. 54	Not met	Not met	Not met	Not included
Barczyk et al. (2012)	File 8, no. 71	Met	Met	Not met	Not included
Barczyk et al. (2017)	File 1, no. 25	Met	Met	Met	Included
Barden (2012)	File 5, no. 1	Met	Met	Not met	Not included
Barnes et al. (2015)	File 4, no. 3	Met	Not met	Not met	Not included
Barnett et al. (2011)	File 7, no. 75	Met	Not met	Not met	Not included
Barquero et al. (2018)	File 2, no. 51	Met	Met	No Met	Not Included
Barrot (2016)	File 3, no. 97	Met	Met	Not met	Not included
Barrot (2018)	File 2, no. 29	Not met	Not met	Not met	Not included
Bassili (2008)	File 9, no. 115	Met	Not met	Not met	Not included
Battles et al. (2011)	File 8, no. 56	Not met	Not met	Not met	Not included
Bedford (2019)	File 2, no. 97	Not met	Met	Not met	Not included
Belanger et al. (2014)	File 8, no. 26	Met	Not met	Not met	Not included
Belkofer et al. (2011)	File 8, no. 95	Not met	Not met	Not met	Not included
Belluci et al. (2019)	File 5, no. 20	Not met	Not met	Not met	Not included
Beltran-Cruz et al. (2013)	File 6, no. 82	Met	Met	Not met	Not included
Benko et al. (2016)	File 7, no. 50	Not met	Met	Not met	Not included
Bennett et al. (2014)	File 8, no. 3	Met	Not met	Not met	Not included
Bennett et al. (2017)	File 7, no. 76	Not met	Not met	Not met	Not included
Bennett et al. (2020a)	File 3, no. 32	Met	Not met	Not met	Not included
Bennett et al. (2020b)	File 3, no. 77	Not met	Not met	Not met	Not included
Berdanier et al. (2020)	File 3, no. 11	Not met	Not met	Not met	Not included
Bessenoff (2006)	File 9, no. 103	Met	Not met	Not met	Not included
Best et al. (2011)	File 3, no. 15	Not met	Met	Not met	Not included

Reference	Archival Reference	Criterion 1	Criterion 2	Criterion 3	Status
Best et al. (2018)	File 6, no. 5	Met	Not met	Not met	Not included
Bharucha (2018)	File 6, no. 29	Met	Not met	Not met	Not included
Bhatnagar et al. (2020)	File 1, no. 71	Met	Not met	Met	Not included
Bigelow et al. (2016)	File 7, no. 63	Met	Not met	Not met	Not included
Bilgin et al. (2018)	File 3, no. 26	Met	Not met	Not met	Not included
Bindra et al. (2020)	File 1, no. 53	Met	Not met	Met	Not included
Black et al. (2020)	File 3, no. 13	Not met	Not met	Not met	Not included
Blankenship (2011)	File 8, no. 28	Met	Not met	Not met	Not included
Bodnar et al. (2011)	File 1, no. 5	Not met	Not met	Not met	Not included
Bolat (2018)	File 3, no. 6	Met	Not met	Met	Not included
Boon et al. (2009)	File 6, no. 4	Not met	Not met	Not met	Not included
Bor (2014)	File 8, no. 82	Met	Met	Not met	Not included
Borch (2019)	File 2, no. 20	Met	Met	Not met	Not included
Bowen et al. (2017)	File 4, no. 11	Met	Not met	Not met	Not included
Bozanta et al. (2017)	File 4, no. 78	Met	Not met	Not met	Not included
Brech et al. (2017)	File 1, no. 19	Not met	Not met	Met	Not included
Bresmer et al. (2015)	File 8, no. 34	Not met	Met	Not met	Not included
Brew et al. (2013)	File 3, no. 60	Met	Not met	Not met	Not included
Briggs (2013)	File 7, no. 100	Not met	Not met	Not met	Not included
Brocato et al. (2015)	File 7, no. 64	Not met	Not met	Not met	Not included
Brookbank (2015)	File 7, no. 61	Met	Not met	Not met	Not included
Brown et al. (2016)	File 7, no. 52	Met	Not met	Not met	Not included
Brown et al. (2020)	File 3, no. 53	Met	Met	Not met	Not included
Bruneel et al. (2013)	File 5, no. 36	Met	Not met	Not met	Not included
Bryant et al. (2014)	File 6, no. 10	Met	Not met	Not met	Not included
Bugeja (2006)	File 7, no. 33	Not met	Not met	Not met	Not included
Bukhari et al. (2018)	File 2, no. 54	Met	Not met	Met	Not included
Bulut et al. (2019)	File 2, no. 88	Met	Not met	Not met	Not included
Burkhart et al. (2017)	File 3, no. 100	Met	Not met	Not met	Not included
Buus (2012)	File 7, no. 58	Met	Met	Not met	Not included
Buzzetto-More et al. (2015)	File 7, no. 13	Met	Not met	Not met	Not included
Cabrera et al. (2017)	File 6, no. 7	Not met	Not met	Not met	Not included
Callaghan et al. (2016)	File 3, no. 88	Not met	Not met	Met	Not included
Cameron (2010)	File 9, no. 110	Not met	Not met	Not met	Not included
Campbell et al. (2019)	File 4, no. 22	Met	Met	Not met	Not included

Reference	Archival Reference	Criterion 1	Criterion 2	Criterion 3	Status
Cao et al 2019	File 3, no. 87	Met	Not met	Not met	Not included
Cao et al. (2013)	File 8, no. 97	Not met	Met	Not met	Not included
Carmichael et al. (2014)	File 6, no. 2	Not met	Not met	Not met	Not included
Carnevale (2006)	File 9, no. 106	Not met	Not met	Not met	Not included
Carreon et al. (2016)	File 5, no. 6	Not met	Not met	Not met	Not included
Cartner et al. (2017)	File 4, no. 15	Not met	Met	Not met	Not included
Carver (2019)	File 2, no. 41	Met	Met	Met	Included
Celebi (2018)	File 1, no. 85	Met	Met	Not met	Not included
Celik (2020)	File 2, no. 64	Met	Not met	Met	Not included
Celik et al. (2015)	File 4, no. 4	Met	Not met	Not met	Not included
Cevik et al. (2014)	File 4, no. 73	Met	Met	Not met	Not included
Chai et al. (2016)	File 7, no. 49	Met	Met	Not met	Not included
Chaka et al. (2020)	File 1, no. 41	Met	Met	Met	Included
Chalupa (2015)	File 7, no. 4	Not met	Not met	Not met	Not included
Chan (2017)	File 7, no. 56	Met	Not met	Not met	Not included
Charlton et al. (2009)	File 6, no. 89	Met	Not met	Not met	Not included
Charteris et al. (2018)	File 2, no. 68	Met	Met	Not met	Not included
Chatten et al. (2016)	File 7, no. 20	Not met	Not met	Not met	Not included
Chen (2015)	File 2, no. 50	Met	Met	Met	Included
Chen et al. (2012)	File 7, no. 42	Not met	Not met	Not met	Not included
Chen et al. (2014)	File 8, no. 42	Not met	Not met	Not met	Not included
Cheng et al. (2016)	File 3, no. 25	Met	Met	Not met	Not included
Childers (2017)	File 3, no. 96	Met	Met	Not met	Not included
Childers et al. (2016)	File 7, no. 8	Not met	Not met	Not met	Not included
Chromey et al. (2016)	File 7, no. 27	Met	Not met	Not met	Not included
Chugh et al. (2018)	File 1, no. 63	Not met	Not met	Not met	Not included
Churcher et al. (2014)	File 6, no. 38	Met	Met	Not met	Not included
Clark et al. (2017)	File 7, no. 57	Met	Not met	Not met	Not included
Clements (2015)	File 2, no. 96	Met	Met	Met	Included
Cochrane et al. (2013a)	File 8, no. 29	Met	Met	Not met	Not included
Cochrane et al. (2013b)	File 8, no. 38	Met	Met	Not met	Not included
Cochrane et al. (2013c)	File 8, no. 50	Met	Met	Not met	Not included
Coetzee et al. (2019)	File 4, no. 9	Met	Not met	Not met	Not included
Connell (2009)	File 6, no. 6	Met	Not met	Not met	Not included
Connolly (2014)	File 8, no. 11	Not met	Not met	Not met	Not included
Connolly et al. (2019)	File 2, no. 21	Met	Not met	Not met	Not included

Reference	Archival Reference	Criterion 1	Criterion 2	Criterion 3	Status
Constantinides (2011)	File 9, no. 117	Not met	Not met	Not met	Not included
Cooke (2017)	File 7, no. 78	Met	Not met	Not met	Not included
Cowley (2017)	File 7, no. 84	Not met	Met	Not met	Not included
Crews et al. (2012)	File 5, no. 42	Not met	Not met	Not met	Not included
Cru (2015)	File 4, no. 51	Met	Not met	Not met	Not included
Cuesta et al. (2016)	File 3, no. 94	Met	Not met	Not met	Not included
Cumberledge (2020)	File 2, no. 80	Not met	Not met	Met	Not included
Curro et al. (2018)	File 3, no. 70	Not met	Not met	Not met	Not included
Curtis et al. (2019)	File 1, no. 39	Met	Not met	Not met	Not included
da Silva et al. (2014)	File 4, no. 84	Met	Met	Not met	Not included
Dabbagh et al. (2012)	File 8, no. 48	Not met	Not met	Not met	Not included
Dabner (2012)	File 1, no. 7	Not met	Not met	Met	Not included
Dahdal (2020)	File 2, no. 98	Met	Not met	Met	Not included
Daniels et al. (2014)	File 1, no. 34	Met	Met	Met	Included
Datko (2019)	File 1, no. 81	Met	Not met	Not met	Not included
Daugird et al. (2015)	File 7, no. 26	Not met	Not met	Not met	Not included
Davidovitch et al. (2018)	File 1, no. 31	Met	Met	Met	Included
Davis et al. (2015)	File 7, no. 40	Met	Met	Not met	Not included
Davis et al. (2015)b	File 8, no. 68	Not met	Not met	Not met	Not included
DeAndrea et al. (2012)	File 8, no. 58	Met	Not met	Not met	Not included
Deans (2012)	File 8, no. 33	Met	Met	Not met	Not included
del Rocio Bonilla et al. (2020)	File 3, no. 48	Not met	Not met	Met	Not included
Delello et al. (2015)	File 8, no. 98	Met	Met	Not met	Not included
Demir (2018)	File 6, no. 90	Met	Met	Not met	Not included
Demirbilek (2015)	File 1, no. 59	Met	Met	Met	Included
Delen (2017)	File 1, no. 75	Met	Met	Met	Included
Dennen et al. (2017)	File 3, no. 45	Met	Not met	Not met	Not included
Dennen et al. (2020)	File 1, no. 69	Not met	Not met	Not met	Not included
DePew (2011)	File 1, no. 61	Met	Not met	Not met	Not included
Dincer et al. (2019)	File 2, no. 6	Met	Not met	Met	Not included
Dizon (2010)	File 9, no. 102	Not met	Met	Not met	Not included
Dizon (2016)	File 3, no. 2	Met	Met	Not met	Not included
Dizon et al. (2016)	File 2, no. 99	Met	Met	Met	Included
Doak (2011)	File 8, no. 51	Not met	Not met	Not met	Not included
Dogan et al. (2018)	File 1, no. 93	Met	Met	Not met	Not included

Reference	Archival Reference	Criterion 1	Criterion 2	Criterion 3	Status
Dogari et al. (2019)	File 2, no. 82	Met	Not met	Not met	Not included
Donelan (2016)	File 7, no. 68	Not met	Not met	Not met	Not included
Donlan (2014)	File 4, no. 35	Met	Not met	Not met	Not included
Dougherty et al. (2014)	File 4, no. 91	Met	Met	Met	Included
Downes (2007)	File 6, no. 55	Not met	Not met	Not met	Not included
Dragseth (2020)	File 3, no. 17	Not met	Not met	Not met	Not included
Duncan et al. (2016)	File 2, no. 34	Met	Met	Met	Included
Dyson et al. (2015)	File 4, no. 29	Met	Met	Not met	Not included
Eaton et al. (2012)	File 7, no. 5	Not met	Not met	Not met	Not included
Eberhardt (2007)	File 6, no. 35	Not met	Not met	Not met	Not included
Edmiston (2016)	File 7, no. 21	Not met	Not met	Not met	Not included
Ekoc (2014)	File 2, no. 32	Met	Met	Not met	Not included
Elavsky (2013)	File 8, no. 5	Not met	Not met	Not met	Not included
Ellefsen (2016)	File 2, no. 8	Met	Not met	Not met	Not included
Enskat et al. (2017)	File 3, no. 4	Met	Met	Not met	Not included
Eraslan et al. (2019)	File 2, no. 65	Met	Not met	Not met	Not included
Ercoskun et al. (2019)	File 2, no. 45	Met	Met	Met	Included
Erdem et al. (2014)	File 2, no. 71	Met	Met	Met	Included
Erdogdu (2016)	File 2, no. 30	Met	Not met	Met	Not included
Eroglu (2016)	File 2, no. 44	Met	Not met	Met	Not included
Ersoz et al. (2018)	File 1, no. 24	Not met	Met	Not met	Not included
Escobar-Rodriguez et al. (2014)	File 4, no. 43	Met	Not met	Not met	Not included
Esgi (2016)	File 5, no. 19	Met	Not met	Not met	Not included
Estrella ibarra (2018)	File 1, no. 45	Met	Met	Met	Included
Evans (2014)	File 8, no. 93	Met	Met	Met	Included
Fagioli et al. (2015)	File 6, no. 15	Met	Not met	Not met	Not included
Fajardo Vizquerra et al. (2019)	File 1, no. 22	Not met	Met	Met	Not included
Faulds et al. (2014)	File 8, no. 15	Met	Met	Not met	Not included
Fenwick (2016)	File 7, no. 77	Not met	Not met	Not met	Not included
Feria-Galicia (2011)	File 8, no. 41	Not met	Not met	Not met	Not included
Fidan (2019)	File 1, no. 16	Met	Not met	Met	Not included
Field et al. (2012)	File 7, no. 90	Not met	Not met	Not met	Not included
Fife (2017)	File 4, no. 16	Not met	Met	Not met	Not included
Filimowicz et al. (2017)	File 6, no. 22	Met	Not met	Not met	Not included

Reference	Archival Reference	Criterion 1	Criterion 2	Criterion 3	Status
Firat (2017)	File 1, no. 54	Met	Not met	Met	Not included
Firat et al. (2017)	File 4, no. 24	Not met	Not met	Not met	Not included
Fischbach et al. (2018)	File 6, no. 12	Not met	Not met	Not met	Not included
Flores et al. (2020)	File 3, no. 67	Met	Not met	Not met	Not included
Foos (2020)	File 1, no. 76	Met	Not met	Not met	Not included
Fornara et al. (2019)	File 4, no. 98	Met	Met	Not met	Not included
Friedman et al. (2013)	File 6, no. 93	Not met	Not met	Not met	Not included
Friesen et al. (2012)	File 8, no. 90	Not met	Not met	Not met	Not included
Frimming et al. (2011)	File 7, no. 31	Met	Met	Not met	Not included
Frisby et al. (2016)	File 3, no. 57	Met	Met	Not met	Not included
Fujita et al. (2017)	File 6, no. 42	Met	Not met	Not met	Not included
Galan et al. (2015)	File 8, no. 70	Met	Not met	Not met	Not included
Gamble et al. (2014)	File 2, no. 66	Met	Met	Met	Included
Gammon et al. (2015)	File 7, no. 91	Not met	Not met	Not met	Not included
Ganster et al. (2009)	File 6, no. 28	Not met	Not met	Not met	Not included
Gavrin et al. (2017)	File 7, no. 28	Met	Met	Not met	Not included
George (2011)	File 1, no. 37	Not met	Not met	Not met	Not included
George et al. (2018)	File 5, no. 14	Not met	Not met	Not met	Not included
Gerlich et al. (2010)	File 7, no. 87	Met	Not met	Not met	Not included
Gettman et al. (2015)	File 2, no. 46	Met	Not met	Met	Not included
Gharis et al. (2017)	File 6, no. 3	Not met	Not met	Not met	Not included
Giannikas (2020)	File 1, no. 27	Met	Met	Met	Included
Gilfoil et al. (2015)	File 5, no. 18	Not met	Not met	Not met	Not included
Gilster et al. (2020)	File 3, no. 78	Met	Met	Not met	Not included
Gin et al. (2016)	File 7, no. 18	Met	Not met	Not met	Not included
Glenn (2015)	File 8, no. 6	Not met	Not met	Not met	Not included
Glider et al., 2001	File 9, no. 104	Met	Not met	Not met	Not included
Goh et al. (2015)	File 8, no. 80	Met	Not met	Not met	Not included
Goktalay (2015)	File 2, no. 78	Met	Met	Met	Included
Gopalakrishnan et al. (2019)	File 3, no. 66	Met	Not met	Not met	Not included
Gordon (2016)	File 3, no. 44	Met	Met	Met	Included
Gorham et al (2012)	File 9, no. 119	Not met	Not met	Not met	Not included
Green (2018)	File 3, no. 74	Met	Not met	Not met	Not included
Greenwood (2012)	File 7, no. 32	Not met	Not met	Not met	Not included
Grefory et al. (2016)	File 3, no. 39	Met	Met	Met	Included

Reference	Archival Reference	Criterion 1	Criterion 2	Criterion 3	Status
Gregory et al. (2014)	File 4, no. 79	Met	Met	Met	Included
Griffin et al. (2020)	File 1, no. 38	Met	Not met	Met	Not included
Gronbeck, 2000	File 9, no. 108	Not met	Not met	Not met	Not included
Guerin et al. (2020)	File 3, no. 22	Not met	Not met	Not met	Not included
Guillaume et al. (2019)	File 4, no. 67	Not met	Not met	Not met	Not included
Gulbahar et al. (2017)	File 4, no. 27	Met	Met	Not met	Not included
Gundlach et al. (2015)	File 7, no. 79	Met	Met	Not met	Not included
Guo et al. (2018)	File 6, no. 94	Met	Met	Met	Included
Gupta et al. (2013)	File 6, no. 60	Met	Not met	Not met	Not included
Halligan (2010)	File 9, no. 122	Not met	Not met	Not met	Not included
Halliwel (2020)	File 3, no. 12	Not met	Not met	Not met	Not included
Hamid et al. (2017)	File 7, no. 67	Not met	Not met	Not met	Not included
Hamutoglu et al. (2020)	File 1, no. 43	Met	Not met	Met	Not included
Hanson et al. (2011)	File 6, no. 14	Met	Not met	Not met	Not included
Harting (2017)	File 1, no. 96	Met	Met	Met	Included
Hashim et al. (2013)	File 3, no. 38	Met	Met	Not met	Not included
Hassell et al. (2016)	File 4, no. 2	Met	Not met	Not met	Not included
He et al. (2017)	File 6, no. 66	Met	Not met	Not met	Not included
Heilig et al. (2019)	File 2, no. 53	Not met	Not met	Not met	Not included
Hennessy et al. (2016)	File 7, no. 62	Met	Met	Met	Included
Hermes (2008)	File 9, no. 107	Not met	Not met	Not met	Not included
Hetz et al. (2015)	File 8, no. 66	Met	Not met	Not met	Not included
Hickerson et al. (2017)	File 6, no. 44	Met	Not met	Not met	Not included
Hoffman et al. (2014)	File 8, no. 19	Met	Not met	Not met	Not included
Hong et al. (2015)	File 1, no. 87	Met	Not met	Met	Not included
Hope (2016)	File 3, no. 23	Met	Not met	Not met	Not included
Hottell et al. (2014)	File 8, no. 75	Met	Not met	Not met	Not included
Hou et al. (2015)	File 4, no. 32	Met	Met	Met	Included
Howard et al. (2017)	File 7, no. 85	Met	Not met	Not met	Not included
Hughes et al. (2020)	File 3, no. 20	Met	Not met	Not met	Not included
Hurt et al. (2012)	File 3, no. 81	Met	Met	Not met	Not included
Hussain et al. (2018)	File 3, no. 72	Met	Not met	Not met	Not included
Hutchens et al. (2014)	File 4, no. 46	Met	Not met	Not met	Not included
Huwe (2011)	File 9, no. 113	Not met	Not met	Not met	Not included
Huwe (2012)	File 6, no. 86	Not met	Not met	Not met	Not included
Iqbal et al. (2016)	File 1, no. 1	Met	Not met	Not met	Not included

Reference	Archival Reference	Criterion 1	Criterion 2	Criterion 3	Status
Iredale et al. (2020)	File 3, no. 7	Not met	Not met	Not met	Not included
Isbulan et al. (2020)	File 1, no. 4	Met	Met	Met	Included
Islam (2018)	File 1, no. 21	Met	Not met	Not met	Not included
Ismail et al. (2017)	File 4, no. 7	Met	Not met	Not met	Not included
Ivala et al. (2012)	File 1, no. 62	Met	Met	Not met	Not included
Iwamoto et al. (2020)	File 1, no. 80	Met	Not met	Met	Not included
Izmirli (2017)	File 1, no. 8	Met	Met	Not met	Not included
Jaffar (2014)	File 4, no. 33	Met	Met	Not met	Not included
Jang (2015)	File 8, no. 13	Met	Not met	Not met	Not included
Jebba et al. (2019)	File 1, no. 84	Met	Not met	Met	Not included
Jin (2015)	File 4, no. 92	Met	Met	Not met	Not included
Joan (2015)	File 2, no. 13	Met	Not met	Met	Not included
Jocson (2015)	File 8, no. 45	Met	Not met	Not met	Not included
Johannesen et al. (2019)	File 1, no. 40	Met	Met	Not met	Not included
Johnson et al. (2017)	File 6, no. 100	Met	Not met	Not met	Not included
Jones (2011)	File 7, no. 99	Not met	Not met	Not met	Not included
Jones et al. (2019)	File 4, no. 14	Met	Not met	Not met	Not included
Joseffson et al. (2016)	File 7, no. 88	Met	Not met	Not met	Not included
Julien et al. (2018)	File 5, no. 39	Not met	Not met	Not met	Not included
Jumaat et al. (2016)	File 3, no. 3	Met	Met	Not met	Not included
Junco et al. (2010)	File 9, no. 105	Not met	Not met	Not met	Not included
Kabilan (2016)	File 3, no. 95	Met	Met	Not met	Not included
Kabilan et al. (2010)	File 6, no. 32	Met	Not met	Not met	Not included
Kahveci (2015)	File 7, no. 54	Met	Not met	Not met	Not included
Kalelioglu (2017)	File 1, no. 92	Met	Met	Not met	Not included
Kaler et al. (2020)	File 3, no. 50	Met	Not met	Not met	Not included
Karaa et al. (2016)	File 4, no. 23	Met	Not met	Not met	Not included
Karademir et al. (2020)	File 1, no. 73	Met	Not met	Met	Not included
Karaman (2019)	File 3, no. 30	Met	Not met	Not met	Not included
Karanjakwut (2018)	File 1, no. 3	Met	Met	Not met	Not included
Karl et al. (2011)	File 6, no. 37	Met	Not met	Not met	Not included
Karpman et al. (2016)	File 7, no. 72	Not met	Not met	Not met	Not included
Karsak (2016)	File 2, no. 86	Met	Met	Not met	Not included
Kasuma (2017)	File 1, no. 36	Met	Met	Not met	Not included
Kasuma et al. (2018)	File 6, no. 46	Met	Not met	Not met	Not included
Katy (2020)	File 3, no. 46	Met	Not met	Not met	Not included

Reference	Archival Reference	Criterion 1	Criterion 2	Criterion 3	Status
Kelly (2018)	File 2, no. 67	Met	Met	Not met	Not included
Kelm (2011)	File 8, no. 69	Met	Met	Not met	Not included
Kettunen et al. (2013)	File 8, no. 9	Not met	Not met	Not met	Not included
Khan et al. (2016)	File 7, no. 94	Met	Not met	Not met	Not included
Khosrovani et al. (2016)	File 7, no. 41	Met	Not met	Not met	Not included
Kidd et al. (2014)	File 8, no. 55	Not met	Not met	Not met	Not included
Kilis et al. (2016)	File 5, no. 27	Not met	Not met	Not met	Not included
Kim (2019)	File 4, no. 25	Met	Met	Not met	Not included
Kim et al. (2014)	File 8, no. 94	Met	Not met	Not met	Not included
Kim et al. (2016)	File 1, no. 13	Met	Not met	Met	Not included
Kinchin et al. (2015)	File 8, no. 99	Not met	Met	Not met	Not included
Kinsky (2015)	File 7, no. 92	Met	Met	Not met	Not included
Kitchakarn (2016)	File 5, no. 41	Met	Met	Not met	Not included
Ko (2019)	File 4, no. 71	Met	Met	Not met	Not included
Kolek et al. (2008)	File 6, no. 51	Met	Not met	Not met	Not included
Koles et al. (2012)	File 5, no. 32	Met	Not met	Not met	Not included
Kooy (2016)	File 7, no. 10	Not met	Not met	Not met	Not included
Koseoglu et al. (2016)	File 2, no. 74	Met	Met	Not met	Not included
Koshkin et al. (2017)	File 6, no. 49	Met	Not met	Not met	Not included
Kothari et al. (2016)	File 7, no. 74	Met	Not met	Not met	Not included
Kramer et al. (2015)	File 4, no. 57	Met	Not met	Not met	Not included
Krutka et al. (2017)	File 6, no. 85	Met	Met	Not met	Not included
Kucuk et al. (2013)	File 3, no. 41	Met	Met	Not met	Not included
Kumar et al. (2019)	File 4, no. 63	Not met	Not met	Not met	Not included
Kurtz (2014)	File 4, no. 48	Met	Met	Not met	Not included
Kwon et al. (2013)	File 5, no. 33	Met	Not met	Not met	Not included
Lam (2012)	File 3, no. 10	Met	Met	Not met	Not included
Lampe et al. (2011)	File 6, no. 71	Met	Not met	Not met	Not included
Lansigan et al. (2016)	File 7, no. 60	Met	Not met	Not met	Not included
LaRiviere et al. (2012)	File 8, no. 52	Met	Not met	Not met	Not included
LaRoche et al. (2009)	File 6, no. 33	Met	Not met	Not met	Not included
Larson (2018)	File 6, no. 75	Not met	Not met	Not met	Not included
Lauricella (2019)	File 4, no. 20	Not met	Not met	Not met	Not included
Lawson et al. (2011)	File 6, no. 88	Met	Not met	Not met	Not included
Le et al. (2019)	File 5, no. 43	Not met	Not met	Not met	Not included
Leafman et al. (2013)	File 6, no. 70	Met	Not met	Not met	Not included

Reference	Archival Reference	Criterion 1	Criterion 2	Criterion 3	Status
Lee (2018)	File 1, no. 18	Met	Not met	Met	Not included
Lee et al. (2016)	File 2, no. 16	Met	Met	Met	Included
Lemon et al. (2019)	File 1, no. 68	Met	Met	Not met	Not included
Leppisaari et al. (2014)	File 8, no. 16	Not met	Met	Not met	Not included
Levesque (2016)	File 7, no. 70	Not met	Not met	Not met	Not included
Leyrer-Jackson et al. (2018)	File 6, no. 77	Met	Not met	Not met	Not included
Li (2012)	File 8, no. 100	Met	Met	Not met	Not included
Li et al. (2015)	File 8, no. 57	Met	Not met	Not met	Not included
Lie (2013)	File 8, no. 62	Met	Met	Not met	Not included
Liggett (2012)	File 8, no. 23	Not met	Not met	Not met	Not included
Lilburn (2012)	File 8, no. 10	Not met	Not met	Not met	Not included
Lim (2010)	File 4, no. 85	Met	Met	Met	Included
Lin (2016)	File 3, no. 27	Met	Met	Not met	Not included
Linder et al. (2016)	File 7, no. 35	Met	Not met	Not met	Not included
Linh et al. (2016)	File 2, no. 73	Not met	Met	Met	Not included
Lint (2013)	File 8, no. 4	Met	Not met	Not met	Not included
Lint et al. (2013)	File 8, no. 35	Met	Met	Met	Included
Linville (2019)	File 4, no. 10	Not met	Not met	Not met	Not included
Locatelli et al. (2012)	File 5, no. 34	Met	Not met	Not met	Not included
Lohnes et al. (2016)	File 3, no. 31	Met	Not met	Not met	Not included
Lou (2017)	File 6, no. 9	Met	Met	Not met	Not included
Lou et al. (2012)	File 5, no. 7	Met	Not met	Not met	Not included
Lovell et al. (2013)	File 5, no. 23	Not met	Met	Not met	Not included
Lu et al. (2014)	File 6, no. 21	Met	Not met	Not met	Not included
Lund (2019)	File 1, no. 89	Not met	Not met	Met	Not included
Luo (2018)	File 6, no. 24	Met	Met	Met	Included
Luo et al. (2020)	File 1, no. 46	Not met	Not met	Not met	Not included
Madden et al. (2016)	File 7, no. 82	Met	Met	Not met	Not included
Magde (2009)	File 6, no. 31	Met	Not met	Not met	Not included
Maglunog et al. (2019)	File 1, no. 32	Met	Not met	Met	Not included
Magolis et al. (2016)	File 5, no. 2	Met	Not met	Not met	Not included
Mahdiun et al. (2020)	File 2, no. 17	Met	Not met	Met	Not included
Mahmood et al. (2018)	File 2, no. 18	Met	Not met	Met	Not included
Maloney et al. (2014)	File 8, no. 64	Met	Not met	Not met	Not included
Manca et al. (2017)	File 4, no. 56	Met	Not met	Not met	Not included

Reference	Archival Reference	Criterion 1	Criterion 2	Criterion 3	Status
Mandavgane (2016)	File 3, no. 92	Met	Met	Not met	Not included
Mandviwalla et al. (2013)	File 8, no. 36	Met	Met	Not met	Not included
Manlow et al. (2010)	File 9, no. 114	Not met	Not met	Not met	Not included
Manson et al. (2018)	File 6, no. 25	Not met	Not met	Not met	Not included
Manuel et al (2016)	File 3, no. 59	Not met	Met	Not met	Not included
Marek et al. (2017)	File 6, no. 57	Not met	Not met	Not met	Not included
Maresova et al. (2020)	File 1, no. 66	Not met	Not met	Met	Not included
Marie Condie et al. (2018)	File 6, no. 53	Met	Not met	Not met	Not included
Marshalsey et al. (2020)	File 3, no. 14	Met	Met	Not met	Not included
Marsilio (2017)	File 4, no. 69	Met	Not met	Not met	Not included
Martinez-Alemán (2014)	File 8, no. 61	Not met	Not met	Not met	Not included
Martinez-Arbelaiz et al. (2017)	File 7, no. 7	Met	Not met	Not met	Not included
Martinez-Cardama et al. (2019)	File 4, no. 60	Met	Met	Met	Included
Marzo et al. (2017)	File 3, no. 63	Met	Not met	Not met	Not included
Mastrodicasa et al. (2013)	File 8, no. 87	Not met	Not met	Not met	Not included
Mazer et al. (2007)	File 6, no. 40	Met	Not met	Not met	Not included
Mazer et al. (2009)	File 6, no. 79	Met	Not met	Not met	Not included
Mazer et al. (2017)	File 7, no. 71	Not met	Not met	Not met	Not included
McCarthy (2015)	File 4, no. 52	Met	Met	Not met	Not included
McCarthy et al. (2014)	File 6, no. 69	Met	Not met	Not met	Not included
McCorkle et al. (2012)	File 8, no. 96	Met	Met	Not met	Not included
Mccorkle et al. (2017)	File 6, no. 96	Met	Met	Not met	Not included
McEachern (2011)	File 5, no. 26	Not met	Not met	Not met	Not included
McHaney et al. (2015)	File 8, no. 81	Not met	Not met	Not met	Not included
Mclain (2019)	File 2, no. 62	Met	Met	Met	Included
McNeill (2012)	File 8, no. 17	Not met	Not met	Not met	Not included
Megele (2015)	File 8, no. 18	Met	Met	Not met	Not included
Meishar-Tal et al. (2012)	File 3, no. 34	Met	Met	Not met	Not included
Melchiorre et al. (2017)	File 7, no. 34	Not met	Not met	Not met	Not included
Melton et al. (2011)	File 8, no. 30	Not met	Met	Not met	Not included
Melton et al. (2018)	File 2, no. 89	Met	Not met	Met	Not included
Mendez et al. (2014)	File 2, no. 42	Not met	Not met	Not met	Not included
Menevse (2019)	File 2, no. 36	Met	Not met	Met	Not included
Menzies et al. (2017)	File 3, no. 1	Met	Not met	Not met	Not included

Reference	Archival Reference	Criterion 1	Criterion 2	Criterion 3	Status
Meredith (2012)	File 8, no. 72	Not met	Not met	Not met	Not included
Messner et al. (2016)	File 7, no. 38	Not met	Met	Not met	Not included
Metzger (2015)	File 5, no. 17	Not met	Not met	Not met	Not included
Mikum et al. (2018)	File 6, no. 72	Met	Met	Not met	Not included
Milburn et al. (2014)	File 4, no. 49	Met	Met	Not met	Not included
Miller (2013)	File 4, no. 47	Met	Met	Met	Included
Miller (2017)	File 6, no. 50	Met	Not met	Not met	Not included
Miller et al. (2007)	File 6, no. 20	Not met	Not met	Not met	Not included
Mirembe et al. (2019)	File 2, no. 47	Met	Not met	Met	Not included
Miron et al. (2015)	File 4, no. 37	Met	Met	Not met	Not included
Mitchell (2012)	File 5, no. 3	Met	Not met	Not met	Not included
Mnkandla et al. (2017)	File 4, no. 87	Not met	Not met	Not met	Not included
Moghavvemi et al. (2018)	File 2, no. 40	Met	Met	Met	Included
Molina et al. (2015)	File 8, no. 24	Not met	Not met	Not met	Not included
Moll et al. (2017)	File 7, no. 22	Met	Not met	Not met	Not included
Mondahl et al. (2014)	File 6, no. 67	Met	Met	Not met	Not included
Monseau et al. (2017)	File 6, no. 43	Not met	Not met	Not met	Not included
Montoneri (2015)	File 2, no. 39	Met	Met	Met	Included
Montoneri (2017)	File 1, no. 30	Met	Met	Met	Included
Moore (2009)	File 9, no. 121	Not met	Not met	Not met	Not included
Moore (2012)	File 5, no. 30	Not met	Not met	Not met	Not included
Moreno et al. (2012)	File 5, no. 12	Met	Not met	Not met	Not included
Morris et al. (2009)	File 6, no. 30	Met	Not met	Not met	Not included
Mostafa (2015)	File 7, no. 29	Met	Not met	Not met	Not included
Muhametjanova et al. (2019)	File 4, no. 70	Met	Not met	Not met	Not included
Munoz et al. (2014)	File 6, no. 92	Met	Met	Not met	Not included
Muñoz et al. (2015)	File 7, no. 93	Not met	Not met	Not met	Not included
Murray et al. (2019)	File 5, no. 10	Met	Not met	Not met	Not included
Murugaiah et al. (2019)	File 4, no. 55	Met	Met	Not met	Not included
Nadelson et al. (2017)	File 4, no. 58	Met	Not met	Not met	Not included
Nagel et al. (2018)	File 3, no. 40	Met	Not met	Met	Not included
Naghdipour et al. (2016)	File 3, no. 47	Met	Met	Met	Included
Nairn et al. (2018)	File 2, no. 35	Met	Not met	Not met	Not included
Nantwi et al. (2017)	File 7, no. 69	Not met	Not met	Not met	Not included
Narayan et al. (2019)	File 4, no. 21	Met	Met	Met	Included

Reference	Archival Reference	Criterion 1	Criterion 2	Criterion 3	Status
Nathan et al. (2014)	File 6, no. 39	Met	Met	Not met	Not included
Nazir et al. (2019)	File 1, no. 11	Met	Met	Met	Included
Nehls (2014)	File 4, no. 81	Met	Not met	Not met	Not included
Neier et al. (2015)	File 7, no. 80	Met	Not met	Not met	Not included
Nekritz (2011)	File 8, no. 40	Not met	Not met	Not met	Not included
Neville et al. (2013)	File 6, no. 95	Met	Met	Not met	Not included
Nkhoma et al. (2015)	File 4, no. 36	Met	Met	Met	Included
Norman et al. (2015)	File 5, no. 29	Met	Not met	Not met	Not included
Normore et al. (2007)	File 9, no. 124	Not met	Not met	Not met	Not included
Nwangwa et al. (2014)	File 6, no. 87	Met	Not met	Not met	Not included
O'Bannon et al. (2014)	File 4, no. 89	Met	Met	Met	Included
O'Brien et al. (2018)	File 3, no. 51	Met	Met	Not met	Not included
Odewumi et al. (2018)	File 3, no. 99	Met	Not met	Not met	Not included
Odom et al. (2013)	File 8, no. 65	Met	Not met	Not met	Not included
Odom et al. (2019)	File 4, no. 95	Met	Met	Not met	Not included
Oelze (2019)	File 4, no. 44	Not met	Not met	Not met	Not included
Oh et al. (2014)	File 5, no. 13	Met	Not met	Not met	Not included
Okoro (2012)	File 8, no. 32	Not met	Not met	Not met	Not included
Okoroma et al. (2018)	File 3, no. 84	Met	Not met	Not met	Not included
Olson et al. (2009)	File 6, no. 58	Met	Not met	Not met	Not included
Olutola et al. (2016)	File 5, no. 9	Met	Not met	Not met	Not included
Omoera et al. (2018)	File 4, no. 45	Met	Not met	Not met	Not included
Oravec (2012)	File 7, no. 11	Not met	Not met	Not met	Not included
Orawiwatnakul et al. (2016)	File 2, no. 4	Met	Met	Met	Included
Oregon et al. (2019)	File 3, no. 18	Not met	Not met	Not met	Not included
Owan et al. (2020)	File 1, no. 88	Met	Not met	Met	Not included
Owens et al. (2017)	File 1, no. 86	Met	Met	Met	Included
Ozdemir. 2017	File 3, no. 64	Met	Met	Not met	Not included
Pai et al. (2017)	File 2, no. 10	Met	Met	Met	Included
Pai et al. (2017)	File 2, no. 93	Met	Met	Met	Included
Palen (2008)	File 9, no. 116	Not met	Not met	Not met	Not included
Paris et al. (2015)	File 2, no. 28	Met	Met	Not met	Not included
Parrot et al. (2010)	File 9, no. 125	Met	Not met	Not met	Not included
Pasquini et al. (2017)	File 6, no. 68	Not met	Not met	Not met	Not included
Pavlik et al. (2015)	File 5, no. 38	Not met	Not met	Not met	Not included

Reference	Archival Reference	Criterion 1	Criterion 2	Criterion 3	Status
Pearson (2010)	File 9, no. 118	Not met	Not met	Not met	Not included
Peeters et al. (2020)	File 1, no. 29	Met	Met	Met	Included
Peluchette et al. (2010)	File 6, no. 27	Met	Not met	Not met	Not included
Pempek et al. (2009)	File 6, no. 18	Met	Not met	Not met	Not included
Persson et al. (2016)	File 7, no. 45	Not met	Not met	Not met	Not included
Peruta et al. (2017)	File 1, no. 64	Not met	Not met	Met	Not included
Peruta et al. (2018)	File 1, no. 47	Not met	Not met	Met	Not included
Petrovic et al. (2014)	File 4, no. 39	Met	Met	Not met	Not included
Pham (2014)	File 8, no. 43	Not met	Not met	Not met	Not included
Phillips (2012)	File 5, no. 4	Not met	Not met	Not met	Not included
Pickering et al. (2017)	File 3, no. 58	Met	Not met	Met	Not included
Pikalek (2010)	File 8, no. 44	Not met	Not met	Not met	Not included
Pimmer et al. (2012)	File 5, no. 31	Met	Not met	Not met	Not included
Pimmer et al. (2017)	File 1, no. 26	Not met	Met	Not met	Not included
Ping et al. (2015)	File 2, no. 76	Met	Met	Met	Included
Piotrowski (2015a)	File 5, no. 22	Not met	Not met	Not met	Not included
Piotrowski (2015b)	File 7, no. 53	Not met	Not met	Not met	Not included
Piotrowski (2015c)	File 7, no. 59	Not met	Not met	Not met	Not included
Pitiporntapin et al. (2015)	File 7, no. 97	Met	Met	Not met	Not included
Platts (2019)	File 4, no. 12	Met	Not met	Not met	Not included
Poelhuber et al. (2011)	File 7, no. 25	Met	Not met	Not met	Not included
Polsgrove et al. (2013)	File 8, no. 2	Met	Not met	Not met	Not included
Popescu et al. (2020)	File 2, no. 23	Met	Met	Met	Included
Poth et al. (2016)	File 3, no. 76	Met	Not met	Not met	Not included
Powers et al. (2012)	File 8, no. 49	Not met	Not met	Not met	Not included
Premadasa (2019)	File 1, no. 55	Met	Not met	Met	Not included
Prescott (2014)	File 4, no. 74	Not met	Not met	Not met	Not included
Prescott et al. (2013)	File 5, no. 35	Met	Not met	Not met	Not included
Pringle et al. (2019)	File 4, no. 83	Not met	Not met	Not met	Not included
Purnamasari (2019)	File 1, no. 52	Met	Met	Met	Included
Qi (2019)	File 4, no. 64	Met	Not met	Not met	Not included
Quadri et al. (2016)	File 7, no. 73	Not met	Not met	Not met	Not included
Quan-Haase et al. (2010)	File 1, no. 91	Met	Not met	Met	Not included
Rahman et al. (2020)	File 2, no. 38	Met	Met	Met	Included
Ramadan (2017)	File 3, no. 91	Met	Not met	Not met	Not included

Reference	Archival Reference	Criterion 1	Criterion 2	Criterion 3	Status
Ramazanoglu et al. (2018)	File 1, no. 74	Met	Not met	Met	Not included
Rambe (2012a)	File 5, no. 5	Met	Met	Not met	Not included
Rambe (2012b)	File 5, no. 16	Met	Met	Not met	Not included
Rambe (2012c)	File 7, no. 16	Met	Met	Not met	Not included
Rambe (2013)	File 1, no. 12	Met	Not met	Not met	Not included
Rambe et al. (2014)	File 4, no. 53	Met	Not met	Not met	Not included
Rambe et al. (2015)	File 7, no. 83	Not met	Not met	Not met	Not included
Ramdani et al. (2020)	File 1, no. 72	Met	Met	Not met	Not included
Ramspott (2013)	File 8, no. 84	Not met	Not met	Not met	Not included
Raza et al. (2017)	File 3, no. 35	Met	Not met	Not met	Not included
Razak et al. (2015)	File 2, no. 19	Met	Met	Not met	Not included
Reich (2010)	File 6, no. 8	Met	Not met	Not met	Not included
Remund (2015)	File 7, no. 43	Met	Met	Not met	Not included
Reynolds et al. (2017)	File 6, no. 61	Not met	Not met	Not met	Not included
Rheingold (2008)	File 9, no. 126	Not met	Not met	Not met	Not included
Rheingold (2010)	File 9, no. 101	Not met	Not met	Not met	Not included
Rhodes et al. (2015)	File 2, no. 43	Met	Not met	Met	Not included
Riady (2014)	File 2, no. 11	Met	Met	Met	Included
Richman et al. (2017)	File 3, no. 36	Met	Not met	Not met	Not included
Riley (2015)	File 2, no. 14	Met	Not met	Not met	Not included
Riser et al. (2020)	File 2, no. 57	Met	Not met	Met	Not included
Roberson et al. (2018)	File 6, no. 52	Met	Not met	Not met	Not included
Roblyer et al. (2010)	File 6, no. 36	Met	Not met	Not met	Not included
Rodliyah (2016)	File 3, no. 93	Met	Met	Not met	Not included
Roebuck et al. (2013)	File 8, no. 22	Not met	Not met	Not met	Not included
Romero-Hall (2017)	File 6, no. 56	Met	Not met	Not met	Not included
Romero-Hall (2018)	File 6, no. 65	Not met	Not met	Not met	Not included
Root et al. (2014)	File 8, no. 73	Met	Not met	Not met	Not included
Rosatelli (2015)	File 7, no. 12	Not met	Not met	Not met	Not included
Rosenberg et al. (2016)	File 7, no. 19	Not met	Met	Not met	Not included
Rosli et al. (2016)	File 5, no. 21	Met	Not met	Not met	Not included
Rotboim et al. (2019)	File 1, no. 50	Met	Not met	Met	Not included
Rouis et al. (2011)	File 6, no. 41	Met	Not met	Not met	Not included
Rowe (2014)	File 8, no. 74	Met	Not met	Not met	Not included
Rubrico et al. (2014)	File 4, no. 41	Met	Met	Met	Included

Reference	Archival Reference	Criterion 1	Criterion 2	Criterion 3	Status
Rudick et al. (2019)	File 4, no. 100	Not met	Not met	Not met	Not included
Russell (2013)	File 6, no. 54	Not met	Not met	Not met	Not included
Russell et al. (2012)	File 5, no. 8	Not met	Not met	Not met	Not included
Rustan (2020)	File 1, no. 10	Met	Not met	Met	Not included
Sacks et al. (2012)	File 8, no. 27	Not met	Not met	Not met	Not included
Sahin (2017)	File 4, no. 80	Met	Not met	Not met	Not included
Saifudin et al. (2016)	File 2, no. 77	Met	Met	Met	Included
Saini et al. (2019)	File 4, no. 54	Met	Not met	Not met	Not included
Salameh (2017)	File 1, no. 6	Met	Met	Met	Included
Salmon et al. (2015)	File 8, no. 91	Met	Met	Not met	Not included
Samuel-Peretz et al. (2017)	File 7, no. 24	Met	Met	Not met	Not included
Samuels (2011)	File 8, no. 21	Not met	Not met	Not met	Not included
Sanderson (2018)	File 6, no. 84	Not met	Not met	Not met	Not included
Sandlin et al. (2014)	File 8, no. 8	Met	Not met	Not met	Not included
Santos (2011)	File 9, no. 112	Not met	Not met	Not met	Not included
Santoveña-Casal (2019)	File 2, no. 79	Met	Not met	Met	Not included
Sapsani (2019)	File 1, no. 33	Met	Not met	Met	Not included
Schirr (2013)	File 8, no. 12	Not met	Not met	Not met	Not included
Schmeichel et al. (2018)	File 4, no. 76	Met	Not met	Not met	Not included
Schooper et al. (2017)	File 1, no. 97	Met	Met	Not met	Not included
Schroeder et al. (2009)	File 6, no. 78	Met	Met	Met	Included
Sciutto (2015)	File 4, no. 93	Met	Not met	Not met	Not included
Seechalio (2014)	File 8, no. 39	Not met	Not met	Not met	Not included
Selwyn (2009)	File 6, no. 34	Met	Not met	Not met	Not included
Sessa (2015)	File 8, no. 78	Not met	Not met	Not met	Not included
Setiawan et al. (2020)	File 1, no. 77	Met	Not met	Met	Not included
Settle et al. (2012)	File 8, no. 63	Met	Not met	Not met	Not included
Shafie et al. (2016a)	File 2, no. 25	Met	Met	Not met	Not included
Shafie et al. (2016b)	File 2, no. 52	Met	Met	Not met	Not included
Shafie et al. (2016c)	File 2, no. 55	Met	Met	Not met	Not included
Sharma et al. (2017)	File 6, no. 64	Not met	Not met	Not met	Not included
Shaw (2016)	File 3, no. 19	Met	Not met	Not met	Not included
Sheeran et al. (2018)	File 2, no. 7	Met	Met	Met	Included
Sheffield et al. (2016)	File 2, no. 48	Not met	Met	Not met	Not included
Shelton (2017)	File 7, no. 37	Not met	Not met	Not met	Not included

Reference	Archival Reference	Criterion 1	Criterion 2	Criterion 3	Status
Shen (2019)	File 4, no. 68	Met	Not met	Not met	Not included
Shensa et al. (2018)	File 6, no. 59	Met	Not met	Not met	Not included
Shephard et al. (2019)	File 4, no. 34	Met	Not met	Not met	Not included
Sherrell et al. (2016)	File 3, no. 5	Met	Met	Not met	Not included
Shields et al. (2019)	File 4, no. 61	Met	Not met	Not met	Not included
Shih (2011)	File 6, no. 13	Met	Met	Met	Included
Shih (2013)	File 3, no. 24	Met	Met	Not met	Not included
Shorkey et al. (2014)	File 8, no. 25	Not met	Met	Not met	Not included
Simiyu et al. (2020)	File 3, no. 71	Met	Not met	Not met	Not included
Simons et al. (2016)	File 7, no. 81	Not met	Not met	Not met	Not included
Simpson (2012)	File 3, no. 29	Not met	Met	Not met	Not included
Simsek et al. (2014)	File 6, no. 81	Met	Not met	Not met	Not included
Simsek et al. (2019)	File 2, no. 2	Met	Not met	Met	Not included
Singh (2013)	File 3, no. 42	Met	Met	Not met	Not included
Singh (2017)	File 6, no. 48	Met	Not met	Not met	Not included
Sittiwong et al. (2015)	File 2, no. 75	Met	Met	Met	Included
Skerrett (2010)	File 6, no. 45	Met	Met	Not met	Not included
Sleeman et al. (2016)	File 7, no. 66	Not met	Not met	Not met	Not included
Slim et al. (2019)	File 2, no. 61	Met	Met	Met	Included
Smith (2014)	File 8, no. 37	Met	Met	Not met	Not included
Snow (2017)	File 6, no. 63	Not met	Not met	Not met	Not included
So et al. (2013)	File 8, no. 53	Met	Not met	Not met	Not included
Sobieraj et al., 2001	File 9, no. 109	Not met	Not met	Not met	Not included
Sofowora (2013)	File 6, no. 83	Met	Not met	Not met	Not included
Solmaz (2017)	File 7, no. 2	Met	Not met	Not met	Not included
Solmaz (2018)	File 1, no. 48	Met	Not met	Met	Not included
Song (2017)	File 2, no. 69	Met	Met	Not met	Not included
Sorensen-Unruh (2017)	File 7, no. 15	Met	Met	Not met	Not included
Spackman et al. (2017)	File 6, no. 26	Not met	Not met	Not met	Not included
Spikes et al. (2014)	File 8, no. 1	Met	Met	Not met	Not included
Stander et al. (2017)	File 7, no. 51	Met	Not met	Not met	Not included
Steinbrecher et al. (2012)	File 5, no. 25	Met	Not met	Not met	Not included
Stephen (2014)	File 2, no. 49	Not met	Not met	Not met	Not included
Stephenson-Abetz et al. (2012)	File 5, no. 40	Met	Not met	Not met	Not included
Stirling (2016)	File 3, no. 80	Met	Not met	Not met	Not included

Reference	Archival Reference	Criterion 1	Criterion 2	Criterion 3	Status
Stoller (2013)	File 8, no. 47	Not met	Not met	Not met	Not included
Strayhorn (2012)	File 5, no. 28	Met	Not met	Not met	Not included
Sumida Garcia et al. (2017)	File 7, no. 23	Met	Not met	Not met	Not included
Sutherland et al. (2019)	File 2, no. 12	Not met	Not met	Met	Not included
Sutherland et al. (2020)	File 1, no. 2	Not met	Not met	Met	Not included
Suwinyattichaiorn (2016)	File 7, no. 55	Not met	Met	Not met	Not included
Syme (2013)	File 8, no. 77	Not met	Not met	Not met	Not included
Szeto et al. (2016)	File 7, no. 44	Met	Not met	Not met	Not included
Tananuraksakul (2014)	File 2, no. 90	Met	Met	Not met	Not included
Tananuraksakul (2015)	File 2, no. 9	Met	Met	Not met	Not included
Tay et al. (2011)	File 8, no. 14	Not met	Met	Not met	Not included
Taylor (2019)	File 4, no. 99	Not met	Not met	Not met	Not included
Taylor et al. (2012)	File 3, no. 37	Met	Not met	Not met	Not included
Tayo et al. (2019)	File 2, no. 63	Met	Not met	Met	Not included
Teclehaimanot et al. (2011)	File 6, no. 74	Met	Not met	Not met	Not included
Teixeira et al. (2017)	File 6, no. 76	Met	Met	Met	Included
Tenorio de Azevedo (2019)	File 4, no. 1	Not met	Not met	Not met	Not included
Teo (2016)	File 3, no. 61	Met	Not met	Not met	Not included
Teo et al. (2018)	File 2, no. 84	Met	Not met	Met	Not included
Teruya et al. (2014)	File 8, no. 59	Not met	Not met	Not met	Not included
Thompson et al. (2012)	File 5, no. 37	Met	Not met	Not met	Not included
Tiggemann et al. (2010)	File 9, no. 123	Met	Not met	Not met	Not included
Torun (2020)	File 1, no. 17	Met	Not met	Met	Not included
Tran (2016)	File 2, no. 87	Met	Met	Met	Included
Tras et al. (2019)	File 1, no. 56	Met	Not met	Met	Not included
Tu (2018)	File 3, no. 28	Met	Not met	Not met	Not included
Tucker (2015)	File 7, no. 95	Met	Met	Met	Included
Tugrul (2017)	File 1, no. 49	Met	Met	Met	Included
Tull et al. (2017)	File 4, no. 59	Not met	Not met	Not met	Not included
Tur et al. (2015)	File 5, no. 24	Met	Met	Met	Included
Tuten et al. (2012)	File 8, no. 83	Not met	Not met	Not met	Not included
Tutgun-Unal (2020)	File 1, no. 58	Met	Not met	Met	Not included
Tynes et al. (2013)	File 8, no. 20	Met	Not met	Not met	Not included

Reference	Archival Reference	Criterion 1	Criterion 2	Criterion 3	Status
Unsal (2018)	File 4, no. 86	Met	Not met	Not met	Not included
Vaccaro et al. (2015)	File 7, no. 89	Met	Not met	Not met	Not included
Valiente-Neighbours (2020)	File 3, no. 43	Not met	Met	Not met	Not included
Valjataga et al. (2009)	File 9, no. 120	Met	Met	Not met	Not included
van Beynen et al. (2016)	File 3, no. 33	Not met	Not met	Met	Not included
Van Den Beemt et al. (2020)	File 3, no. 89	Not met	Not met	Not met	Not included
van Rooyen et al. (2020)	File 3, no. 75	Met	Not met	Not met	Not included
VanDoorn et al. (2013)	File 1, no. 20	Met	Met	Met	Included
Veeck et al. (2014)	File 8, no. 92	Met	Met	Not met	Not included
Veletsianos (2013)	File 8, no. 46	Not met	Not met	Not met	Not included
Veletsianos et al. (2019)	File 4, no. 8	Not met	Not met	Not met	Not included
Vie (2015)	File 7, no. 98	Not met	Not met	Not met	Not included
Vinuales et al. (2017)	File 7, no. 6	Met	Not met	Not met	Not included
Vivakaran et al. (2018)	File 3, no. 82	Not met	Not met	Not met	Not included
Vivakaran et al. (2018)	File 6, no. 98	Not met	Not met	Not met	Not included
Vivian et al. (2014)	File 4, no. 75	Met	Not met	Not met	Not included
Voivonta et al. (2018)	File 2, no. 33	Not met	Not met	Not met	Not included
Vrontis et al. (2018)	File 6, no. 80	Met	Not met	Not met	Not included
Wandera et al. (2016)	File 7, no. 86	Met	Not met	Not met	Not included
Wang et al. (2013)	File 3, no. 55	Met	Met	Met	Included
Wang et al. (2014a)	File 4, no. 72	Met	Not met	Not met	Not included
Wang et al. (2014b)	File 8, no. 85	Met	Met	Not met	Not included
Wanjohi et al. (2015)	File 2, no. 58	Met	Not met	Met	Not included
Warner et al. (2017)	File 3, no. 21	Not met	Not met	Not met	Not included
Watson (2017)	File 7, no. 46	Not met	Not met	Not met	Not included
Weiqin et al. (2016)	File 3, no. 68	Met	Not met	Not met	Not included
West (2017)	File 6, no. 19	Met	Not met	Not met	Not included
Whittaker et al. (2014)	File 4, no. 30	Met	Met	Met	Included
Whittaker et al. (2015)	File 7, no. 17	Met	Not met	Not met	Not included
Wichadee (2013)	File 3, no. 62	Met	Met	Not met	Not included
Wicks et al. (2020)	File 3, no. 9	Not met	Not met	Not met	Not included
Willems et al. (2018)	File 6, no. 23	Not met	Not met	Not met	Not included
Williams et al. (2017)	File 4, no. 90	Met	Not met	Not met	Not included
Willoughby et al. (2017)	File 4, no. 28	Met	Not met	Not met	Not included

Reference	Archival Reference	Criterion 1	Criterion 2	Criterion 3	Status
Wilner et al. (2018)	File 2, no. 27	Met	Not met	Met	Not included
Winkelmann et al. (2018)	File 6, no. 11	Met	Not met	Not met	Not included
Witte (2014)	File 4, no. 17	Not met	Not met	Not met	Not included
Woodley et al. (2012)	File 8, no. 76	Met	Not met	Not met	Not included
Woodley et al. (2014)	File 8, no. 88	Not met	Not met	Not met	Not included
Woods et al. (2019)	File 4, no. 94	Met	Met	Not met	Not included
Wright et al. (2017)	File 7, no. 36	Met	Not met	Not met	Not included
Wu et al. (2012)	File 3, no. 85	Met	Not met	Not met	Not included
Wu et al. (2020)	File 1, no. 98	Met	Met	Met	Included
Xu et al. (2014)	File 6, no. 1	Met	Not met	Not met	Not included
Yagci (2015)	File 5, no. 11	Met	Met	Met	Included
Yaman (2016)	File 2, no. 26	Met	Not met	Met	Not included
Yaman (2020)	File 1, no. 44	Met	Not met	Met	Not included
Yang (2018)	File 1, no. 60	Met	Not met	Not met	Not included
Yang et al. (2020)	File 1, no. 42	Met	Not met	Met	Not included
Yavich et al. (2019)	File 2, no. 59	Met	Not met	Met	Not included
Yeh et al. (2019)	File 2, no. 60	Met	Not met	Met	Not included
Yeh et al. (2020)	File 3, no. 16	Met	Met	Not met	Not included
Yen et al. (2015)	File 4, no. 13	Met	Met	Not met	Not included
Yen et al. (2015b)	File 8, no. 7	Met	Not met	Not met	Not included
Yeo (2014)	File 6, no. 62	Met	Not met	Not met	Not included
Yesilyurt et al. (2020)	File 1, no. 51	Met	Not met	Met	Not included
Yildirim et al. (2018)	File 4, no. 50	Met	Not met	Not met	Not included
Young (2009)	File 9, no. 111	Not met	Not met	Not met	Not included
Young et al. (2014)	File 8, no. 67	Not met	Not met	Not met	Not included
Yu (2014)	File 2, no. 1	Met	Met	Met	Included
Zachos et al. (2018)	File 4, no. 65	Not met	Not met	Not met	Not included
Zgheib et al. (2020)	File 1, no. 65	Not met	Met	Met	Not included
Zhang et al. (2014)	File 4, no. 88	Met	Met	Met	Included
Zhang et al. (2016)	File 4, no. 18	Met	Met	Not met	Not included
Zhu (2019)	File 4, no. 62	Met	Not met	Not met	Not included
Zhu et al. (2020)	File 3, no. 90	Met	Not met	Not met	Not included
Zhuravleva et al. (2016)	File 7, no. 96	Not met	Not met	Not met	Not included
Zickar et al. (2018)	File 2, no. 95	Not met	Not met	Met	Not included
Zimmerman et al. (2020)	File 1, no. 28	Met	Not met	Not met	Not included

Notes. Criterion 1, Undergraduate or graduate level students; Criterion 2, Use of social media for educational purposes; Criterion 3, Quantitative variable of engagement in social media. Archival reference files can be downloaded from <https://www.dropbox.com/s/tus5zkwf8wom7r1/Archival%20Reference%20Files.zip?dl=0>

Anexo D. Effect of the Behavioral Education and Social Media (BE-Social) Program on Postgraduate Academic Performance: A Randomized Controlled Trial

SUPPLEMENTARY ONLINE INFORMATION

Appendix 1 Examples of Self-Management Posts

#	Procedure	Content Type	Target Skill	Selection of Textual Content
1	Self-recording	Textual and visual	Independent study habits	"Hey, here is a simple example of a self-recording form: "Total number of hours I spent studying last week, 2 + 3 + 2 + 1 + 2 + 2 + 1 + 2 + 1 + 2 + 1 + 1 + 1 = 21 h"
2	Goal setting	Textual and visual	Independent study habits	"Check out this list of topics you have to study and next to it write down the number of pages in the study material for each of them. Please, do the same for other sections and share away!"
3	Goal setting	Textual and visual	Independent study habits	"Hey, let's write down on a sheet of paper the activities you have to do before you start studying: (1) collect the notes taken from videos and literature on the subject, (2) organize your study space, (3) warn those around you that you are not to be disturbed, (4) switch off your phone."
4	Goal setting, self-recording, self-reinforcement	Textual and visual	Independent study habits	"Let's create a schedule with the day of the week and your available study hours. Then divide the time as follows: 30 minutes study, 10 minutes rest, 30 minutes study, 10 minutes rest, etc."
5	Goal setting, self-recording	Textual and visual	Independent study habits	"When studying a new lesson, create a list with all the new concepts. Then try to memorize them. Write up the concepts in a blank piece of paper keeping the reference material away from you. When you are done, mark the concepts that you did recall correctly, and note the time it took you. Keep track of your memorization timing!"

Appendix 2 Example Timeline of a Video Modeling Post

Event #	Start and End h:mm:ss, h:mm:ss	Event Description
1	N/A	Post text: "I am going to demonstrate the studying strategy discussed in an earlier post by memorizing concepts written on flashcards. Please, use your own flashcards to complete this activity. You practice at the same time and repeat the video to help you stay on track as many times as you like. Let us know about how it goes in the comments."
2	0:00:01, 0:00:05	The instructor sets up a 20-min timer alarm on an iPad.
3	0:00:05, 0:05:00	The instructor works on a large whiteboard. There are 50 small color flashcards attached to the wall to the left of the whiteboard. Each card contains the definition of a key course concept. The instructor writes in the whiteboard a list with the first 20 concepts that are defined in the cards. The text in the flashcards and the whiteboard are not visible in the video. [Students need to use their own flashcards to participate in the activity].
4	0:05:00, 0:20:00	The instructor reads out loud the definition of the first concept, steps away from the flashcard, and writes the definition on the whiteboard in his/her own words. The process is repeated for every concept. Each reading and re-writing cycle takes approximately one minute. The instructor paces herself as to complete the process for the 20 first concept before the timer alarm goes off.
5	0:20:00, 0:30:00	The alarm goes off and the instructor gives a 10-min break. The instructor sets a new 10-min timer alarm. The video streaming continues during the break: the instructor is shown quietly in her desk resting and looking at her phone. The alarm goes off signaling the end of the break period.
6	0:30:01, 0:30:05	The instructor sets a new 30-min timer alarm.
7	0:30:06, 0:30:10	The instructor writes a list with the remaining 30 concepts on the whiteboard.
8	0:30:10, 1:00:00	The process described in event #3 is repeated for the 30 remaining flashcards.
9	1:00:00, 1:00:30	The alarm goes off and the session ends. The instructor thanks those who attended and encourage students to use the video to practice the strategy as often as they like.

Note. N/A = Non-applicable.

Anexo E. Enhancing Academic Performance in Tertiary Education through Social Media: A Multiple-Arm RCT

SUPPLEMENTARY ONLINE MATERIAL

Table A Sociodemographic Characteristics of Individuals
Initially Invited to Participate (n = 170)

	Total (n = 170)	G1 (n = 36)	G2 (n = 34)	G3 (n = 31)	G4 (n = 34)	G5 (n = 35)
Gender, % (n)						
Female	96.87 (155)	91.7 (33)	91.18 (31)	98.8 (30)	97.1 (33)	80 (28)
Male	9.37 (15)	8.4 (3)	8.82 (3)	3.22 (1)	2.94 (1)	20 (7)
Mean age, years (SD)	41.63 (8.59)	40.61 (8.72)	41.48 (8.34)	42.74 (7.16)	43.32 (9.16)	40.17 (9.30)
Years of education (SD)	16.50 (5.44)	17.53 (5.14)	16.65 (5.53)	16.06 (4.94)	17.56 (5.36)	14.51 (5.84)
Mean BL performance (SD)	12.26 (2,18)	12.19 (1.64)	12.42 (2.04)	12.62 (2.11)	12.35 (2.42)	11.79 (2.62)
Country, % (n)						
U. S	95.62 (143)	91.7 (33)	85.29 (29)	93.55 (29)	94.12 (32)	82.85 (29)
Others	10.62 (17)	8.4 (3)	14.28 (5)	6.45 (2)	5.88 (2)	17.14 (6)
Socioeconomic Status, % (n)						
Lower-middle	62.35 (106)	50.00 (18)	67.65 (24)	70.97 (22)	50.00 (17)	71.43 (26)
Middle	4.70 (8)	2.78 (5)	2.94 (1)	3.23 (1)	2.94 (2)	5.71 (2)
Upper-middle	31.17 (53)	38.89 (14)	23.53 (9)	16.13 (7)	41.18 (15)	20.00 (7)
Upper class	1.76 (3)	8.33 (3)	0.00 (0)	3.23 (1)	0.00 (0)	0.00 (0)

Table B Summary of Procedural Integrity Analysis

	G1	G2	G3	G4	G5
<i>Baseline</i>					
No. of daily posts	3	3	3	3	3
No. weekly videos	2	2	2	2	2
Instructor feedback latency ¹	74.4±37.8	99.7±46.1	82.4±39.7	81.7±35.7	94.2±47.7
Relevant video content ² , %	0.0	0.0	0.0	0.0	0.0
No. self-management posts	0	0	0	0	0
Self-management posts format ³	0:0	0:0	0:0	0:0	0:0
No. cooperative learning posts	0	0	0	0	0
<i>Intervention</i>					
No. of daily posts	3	3	3	3	3
No. weekly videos	2	2	2	2	2
Instructor feedback latency ¹	0.82±1.9	83.10±36.4	3.9±6.4	5.8±8.1	107.7±24.3
Relevant video content ² , %	95.4	91.7	0.0	95.6	0.0
No. self-management posts	1	1	0	3	0
Self-management posts format ³	4:8	4:8	0:8	8:8	0:0
No. cooperative learning posts	1	1	3	0	0
<i>Post-Intervention</i>					
No. of daily posts	3	3	3	3	3
No. weekly videos	2	2	2	2	2
Instructor feedback latency ¹	120.4±75.8	84.8±39.3	72.6±41.0	85.3±46.9	110.4±38.3
Relevant video content ² , %	0.0	0.0	0.0	0.0	0.0
No. self-management posts	0	0	0	0	0
Self-management posts format ³	0:0	0:0	0:0	0:0	0:0
No. cooperative learning posts	0	0	0	0	0

Notes. G1 = BE-Social; G2 = BE-Social without semi-immediate feedback; G3 = Cooperative learning; G4 = Self-management; G5 = Control group. Instructor feedback latency and relevant video content as coded on 20% of randomly selected instructor video posts within each group and study phase. 1. Hours (Mean ± SD); 2. Percentage of partial intervals with the instructor presenting self-management skills during video broadcasts; 3. Video posts vs. textual posts (V:T).

Table C Unsolicited Positive Comments during Farewell Video Broadcast

Group	n (%)
Control	9 (33.3)
G1: BE-Social	12 (44.4)
G2: BE-Social without semi-immediate feedback	14 (51.9)
G3: Cooperative learning	6 (18.2)
G4: Self-Management	5 (18.5)
Total	44 (32.6)

Appendix 1 Examples of Self-Management Posts

#	Procedure	Content Type	Target Skill	Selection of Textual Content
1	Self-recording	Textual and visual	Independent study habits	"Hey, here is a simple example of a self-recording form: "Total number of hours I spent studying last week, 2 + 3 + 2 + 1 + 2 + 2 + 1 + 2 + 1 + 2 + 1 + 1 + 1 + 1 = 21 h"
2	Goal setting	Textual and visual	Independent study habits	"Check out this list of topics you have to study and next to it write down the number of pages in the study material for each of them. Please, do the same for other sections and share away!"
3	Goal setting	Textual and visual	Independent study habits	"Hey, let's write down on a sheet of paper the activities you have to do before you start studying: (1) collect the notes taken from videos and literature on the subject, (2) organize your study space, (3) warn those around you that you are not to be disturbed, (4) switch off your phone."
4	Goal setting, self-recording, self-reinforcement	Textual and visual	Independent study habits	"Let's create a schedule with the day of the week and your available study hours. Then divide the time as follows: 30 minutes study, 10 minutes rest, 30 minutes study, 10 minutes rest, etc."
5	Goal setting, self-recording	Textual and visual	Independent study habits	"When studying a new lesson, create a list with all the new concepts. Then try to memorize them. Write up the concepts in a blank piece of paper keeping the reference material away from you. When you are done, mark the concepts that you did recall correctly, and note the time it took you. Keep track of your memorization timing!"

Appendix 2 Example Timeline of a Video Modeling Post

Event #	Start and End h:mm:ss, h:mm:ss	Event Description
1	N/A	Post text: "I am going to demonstrate the studying strategy discussed in an earlier post by memorizing concepts written on flashcards. Please, use your own flashcards to complete this activity. You practice at the same time and repeat the video to help you stay on track as many times as you like. Let us know about how it goes in the comments."
2	0:00:01, 0:00:05	The instructor sets up a 20-min timer alarm on an iPad.
3	0:00:05, 0:05:00	The instructor works on a large whiteboard. There are 50 small color flashcards attached to the wall to the left of the whiteboard. Each card contains the definition of a key course concept. The instructor writes in the whiteboard a list with the first 20 concepts that are defined in the cards. The text in the flashcards and the whiteboard are not visible in the video. [Students need to use their own flashcards to participate in the activity].
4	0:05:00, 0:20:00	The instructor reads out loud the definition of the first concept, steps away from the flashcard, and writes the definition on the whiteboard in his/her own words. The process is repeated for every concept. Each reading and re-writing cycle takes approximately one minute. The instructor paces herself as to complete the process for the 20 first concept before the timer alarm goes off.
5	0:20:00, 0:30:00	The alarm goes off and the instructor gives a 10-min break. The instructor sets a new 10-min timer alarm. The video streaming continues during the break: the instructor is shown quietly in her desk resting and looking at her phone. The alarm goes off signaling the end of the break period.
6	0:30:01, 0:30:05	The instructor sets a new 30-min timer alarm.
7	0:30:06, 0:30:10	The instructor writes a list with the remaining 30 concepts on the whiteboard.
8	0:30:10, 1:00:00	The process described in event #3 is repeated for the 30 remaining flashcards.
9	1:00:00, 1:00:30	The alarm goes off and the session ends. The instructor thanks those who attended and encourage students to use the video to practice the strategy as often as they like.

Note. N/A = Non-applicable.

Anexo F. Publicaciones, participación en proyectos, ponencias internacionales

1. PUBLICACIONES REALIZADAS DURANTE EL DOCTORADO, PERO NO INCLUIDAS EN LA TESIS (2019-2022)

- Fuentes, B., de la Fuente-Gómez, L., Sempere-Iborra, C., Delgado-Fernández, C., Tarifa-Rodríguez, A., Alonso de Leciñana, M., de Celis-Ruiz, E., Gutiérrez-Zúñiga, R., López-Tàpper, J., Martín Alonso, M., Pastor-Yborra, S., Rigual, R., Ruiz-Ares, G., Rodríguez-Pardo, J., Virués-Ortega, J., Borobia, A. M., Blanco, P., & Bueno-Guerra, N. (2022). DUBbing Language-therapy CINema-based in Aphasia post-Stroke (DULCINEA): study protocol for a randomized crossover pilot trial. *Trials*, 23(1), 21. <https://doi.org/10.1186/s13063-021-05956-5>
- Virues-Ortega, J., Perez-Bustamante, A., & Tarifa Rodríguez, A. (2022). Evidence-Based Applied Behavior Analysis (ABA) Autism Treatments: An Overview of Comprehensive and Focused Meta-Analyses. In J. L. Matson & P. Sturmey (Eds.), *Handbook of Autism and Pervasive Developmental Disorders*, Springer. <https://doi.org/10.1007/978-3-030-88538-0>
- Barbera, M. L. (2022). *El enfoque de la conducta verbal* (A. Tarifa-Rodríguez & J. Virues-Ortega, eds. & trans.). ABA España. (Original published 2007).
- Esch, B. A., Tarifa-Rodríguez, A., y Virués-Ortega, J. (2021). Adaptación al español de la Evaluación de Habilidades Ecoicas Tempranas (Early Echoic Skills Assessment, EESA): Instrucciones de administración. En M. L. Sundberg (Ed.), *VB-MAPP, Evaluación y programa de ubicación curricular de los hitos de la conducta verbal: Guía* (págs. 43-47, 162-163, 182-183). ABA España. https://doi.org/10.26741/978-84-09-33123-9_0
- Sundberg, M. L. (2021). *VB-MAPP, Evaluación y programa de ubicación curricular de los hitos de la conducta verbal: Guía* (A. Tarifa Rodríguez y J. Virués Ortega, Eds. y

- trads.). ABA España. <https://doi.org/10.26741/978-84-09-33123-9> (Última edición del original publicada en 2014)
- Sundberg, M. L. (2021). *VB-MAPP, Evaluación y programa de ubicación curricular de los hitos de la conducta verbal: Protocolo* (A. Tarifa Rodríguez y J. Virués Ortega, Eds. y trads.). ABA España. <https://doi.org/10.26741/978-84-09-33124-6> (Última edición del original publicada en 2017)
 - Bailey, J. S., & Burch, M. R. (2021). *25 habilidades esenciales y estrategias para analistas de conducta: Consejos de expertos para ser profesionales más eficaces* (J. Virués-Ortega, J. I. Navarro-Guzman, & Aida Tarifa-Rodriguez, Eds. & Trans.). ABA España. <https://doi.org/10.26741/978-84-09-31731-8>
 - Cooper, J. O., Heron, T. E., y Heward, W. L. (2020). Reforzamiento diferencial (A. Tarifa Rodríguez, J. Virues-Ortega, trad.). En *Análisis Aplicado de Conducta*, tercera edición en español (J. Virues Ortega, ed.) (págs. 595-612). ABA España <https://doi.org/10.26741/abaespana/2020/Cooper25>
 - Cooper, J. O., Heron, T. E., y Heward, W. L. (2020). Imitación, modelado y aprendizaje observacional (A. Tarifa Rodríguez, J. Virues-Ortega, trad.). En *Análisis Aplicado de Conducta*, tercera edición en español (J. Virues Ortega, ed.) (págs. 529-540). ABA España <https://doi.org/10.26741/abaespana/2020/Cooper21>
 - Cooper, J. O., Heron, T. E., y Heward, W. L. (2020). Generalización y mantenimiento del cambio de conducta (A. Tarifa Rodríguez, J. Virues-Ortega, trad.). En *Análisis Aplicado de Conducta*, tercera edición en español (J. Virues Ortega, ed.) (págs. 716-758). ABA España <https://doi.org/10.26741/abaespana/2020/Cooper30>
 - ABA España (2020). *Medidas preventivas para terapeutas de servicios de análisis aplicado de conducta (ABA) frente al COVID-19* (J. Virues-Ortega, editor/coordinador). <https://doi.org/10.26741/covid19>
 - Bailey, J. S., & Burch, M. R. (2019). Evaluación de la conducta. En *Ética para analistas de conducta*, 3ª ed. (págs. 121-136) J. Virues-Ortega ed., A. Tarifa-Rodriguez, trad. ABA España Publicaciones. 10.26741/abaspain/2019/Bailey08 ISBN 978-84-09-07803-5. <https://doi:10.26741/abaspain/2019/Bailey8>
 - Bailey, J. S., & Burch, M. R. (2019). El analista de conducta y el programa de cambio de conducta. En *Ética para analistas de conducta*, 3ª ed. (págs. 137-157) J. Virues-Ortega

ed., A. Tarifa-Rodríguez, trad.). ABA España Publicaciones. 10.26741/abaspain/2019/Bailey09 ISBN 978-84-09-07803-5. [https:// doi:10.26741/abaspain/2019/Bailey9](https://doi.org/10.26741/abaspain/2019/Bailey9)

2. PUBLICACIONES REALIZADAS DURANTE EL DOCTORADO, PERO NO INCLUIDAS EN LA TESIS (EN REVISIÓN)

- Virues-Ortega, J., Pérez-Bustamante Pereira, A., Martín, N., Moeyaert, M., Krause, P. A., Tarifa-Rodríguez, A., Trujillo, C., & Sivaraman, M. Reducing face-touching through motion sensing and vibrotactile cueing during the COVID-19 pandemic. Manuscript under review at *Journal of Applied Behavior Analysis*.

3. PUBLICACIONES (PROYECTO PERSONAL)

- Tarifa Rodríguez, A (2022). Nala y su analista de conducta: Enséñame mi voz. ABA España ISBN: 978-84-09-34755-1
- Tarifa Rodríguez, A & Keenan, M (2022). Nala and her behavior analyst: Finding my voice. ABA España ISBN: 978-8409374014

4. PRESENTACIONES RELACIONADAS CON LA TESIS DOCTORAL EN CONGRESOS INTERNACIONALES (2019-2022)

- Tarifa- Rodríguez, A. (May, 2021). An experimental evaluation of a Facebook group's contribution to academic engagement and performance among postgraduate students. *XLVIII Conference of the Association for Behavior Analysis International*. Washington, EEUU.
- Tarifa- Rodríguez, A. (May, 2020). Análisis experimental de la interacción mediada por redes sociales y su impacto en el rendimiento académico. Universidad Autónoma de Madrid, Madrid.

5. PRESENTACIONES NO RELACIONADAS CON LA TESIS DOCTORAL EN CONGRESOS INTERNACIONALES

- Tarifa- Rodríguez, A. (May, 2022). Automated detection of face touching as an aid to face touching reduction studies. *XLVIII Conference of the Association for Behavior Analysis International*. Boston, EEUU.

- Tarifa- Rodríguez, A. (May, 2020). Current trends in the assessment of treatment outcomes of behavioral services for autism and other developmental disabilities. *XLVI Conference of the Association for Behavior Analysis International*. Washington, EEUU.

6. ESTUDIOS PRESENTADOS EN CONGRESOS INTERNACIONALES EN LOS QUE HE PARTICIPADO

- Tarifa- Rodríguez, A. (May, 2022). Reducing face-touching through motion sensing and vibrotactile cueing during the covid-19 pandemic: treatment effects and disruptor descriptive analysis. *XLVIII Conference of the Association for Behavior Analysis International*. Boston, EEUU.
- Tarifa- Rodríguez, A. (May, 2021). Suppressive effect of contingent vibrotactile stimulation on face touching during the COVID-19 Pandemic: An experimental treatment evaluation. *XLVII Conference of the Association for Behavior Analysis International*. Washington, EEUU.
- Tarifa-Rodríguez, A. (May, 2021). Current trend in the assessment of treatment outcomes of behavioral services for autism and other developmental disabilities. *XLVII Conference of the Association for Behavior Analysis International*. Washington, EEUU.

7. ESTANCIAS DE INVESTIGACIÓN

- Durante los meses de julio, agosto, setiembre y octubre del 2021 realicé una estancia de investigación en la Universidad de Auckland (Nueva Zelanda), en concreto, en el departamento de análisis experimental de la conducta. Durante este periodo impartí clases teóricas y prácticas a estudiantes de la universidad sobre metodología y recogida de datos, como parte de una investigación en curso sobre enriquecimiento ambiental. Ayudé a tomar datos y procesarlos en diferentes proyectos en curso. Asistí a cursos de ética con el fin de aplicar correctamente procedimientos rutinarios en el laboratorio; los cuales llevé a cabo durante todo el periodo. Durante mi estancia aprendí y profundicé en diseños y análisis experimentales, lo cuáles he intentado plasmar en los estudios presentados en la presente tesis.

8. CONTRATO CON CARGO A PROYECTO

- Contrato predoctoral con cargo a proyecto (enero-diciembre 2020). Financiado mediante la dotación adicional de una ayuda Ramón y Cajal del Ministerio de Ciencia e Innovación (RYC-2016-20706).

9. PROYECTO DE INVESTIGACIÓN (2020-ACTUALIDAD)

Proyecto: *DUBbing Language-therapy CINema-based in Aphasia post-Stroke (DULCINEA): A feasibility cross-over pilot-trial* (HR18-00026)

IP: Dra. Blanca Fuentes, Hospital Universitario La Paz

Financiación: Fundación La Caixa (308.066€)

Periodo: 2019-2023

Rol: Asistente de investigación

El proyecto de investigación Dulcinea está liderado por la Dr. Fuentes en el Hospital Universitario la Paz (Madrid). Esta investigación innovadora está dirigida a la mejora de la comunicación de pacientes que han sufrido un ictus. La intervención consta de un software basado en técnicas de doblaje de series de televisión. Actualmente se está validando y evaluando la eficacia de la intervención. Mi papel en este proyecto es evaluar la comunicación y el lenguaje mediante pruebas estandarizadas, tanto para la selección de los participantes en el proyecto como trimestralmente.