UNIVERSIDAD AUTÓNOMA DE MADRID

FACULTAD DE PSICOLOGÍA

Doctorado en Metodología de las Ciencias del Comportamiento y de la Salud

Resolución de problemas relacionados con el comportamiento y la salud mediante la aplicación de técnicas avanzadas de reconocimiento de patrones

Autor: David Delgado-Gómez

Directores:

Dr. David Aguado

Dr. Hilario Blasco-Fontecilla

Dr. Vicente Ponsoda

Madrid, Noviembre 2015

Agradecimientos:

Esta tesis concluye la primera etapa del camino iniciado en los últimos años que me ha permitido abordar interesantes problemas en el campo de las ciencias de la salud y del comportamiento. Antes de proceder a exponer los resultados obtenidos quiero primeramente agradecer a mis supervisores David, Hilario y Vicente todo el apoyo recibido sin el que no hubiera sido posible la realización de esta tesis. También quiero a gradecer a mis familiares, y en especial a Elena, la paciencia que han tenido conmigo en los momentos difíciles. Finalmente, esta tesis también contiene pedacitos de conocimiento obtenidos en las numerosas charlas científicas con los compañeros de los distintos centros a los que he tenido la suerte de pertenecer. Estos centros incluyen el Instituto de Ingeniería del Conocimiento de la Universidad Autónoma de Madrid, el Instituto de Modelado Matemático de la Universidad Técnica de Dinamarca, el Departamento de Biometría de la Universidad Pompeu Fabra, el Departamento de Psiquiatría de la Fundación Jiménez Díaz y el Institut National de la Santé et de la Resecher Medicale de MontPellier.

A todos vosotros, muchas gracias.

Tabla de contenido

“There are only patterns, patterns on top of patterns, patterns that affect other patterns.
Patterns hidden by patterns. Patterns within patterns.
If you watch close, history does nothing but repeat itself.
What we call chaos is just patterns we haven't recognized. What we call random is just patterns we can't decipher. What we can't understand we call nonsense. What we can't read we call gibberish.
There is no free will.
There are no variables.”

Survivor

by Chuck Palahniuk

Resumen

En esta tesis, se investigan los beneficios que se obtendrían al aplicar técnicas recientes de reconocimiento de patrones en la resolución de diversos problemas existentes en el área de la Psicología del comportamiento y de la salud. En las ciencias de la salud, se observa que estas técnicas son capaces de mejorar, en más de un 10%, la precisión o capacidad predictiva de las técnicas actuales de identificación de personas en riesgo de cometer un intento de suicidio. Además, se muestra que estas técnicas pueden ser utilizadas para construir escalas de propósito específico mediante la selección de los ítems más adecuados de escalas de propósito general. En el área de las ciencias del comportamiento, estas técnicas son capaces de identificar a los candidatos más adecuados en los procesos de selección de personal lo que permite reducir considerablemente los gastos de las empresas y facilitar su crecimiento. También se muestra que estas técnicas pueden ser utilizadas para verificar la identidad del participante cuando parte del proceso de selección se realiza online. Se presentan 6 artículos en los que se comparan diversas técnicas (regresión lineal, regresión lineal con selección de variables, regresión logística, análisis discriminante lineal y cuadrático, análisis discriminante de Fisher, boosting, árboles de decisión, Máquinas de Vectores Soporte, y el algoritmo Lars-en) con distintos objetivos, resultando especialmente eficientes las dos últimas.

Esta tesis también investiga la creación de predictores más discriminativos analizando el comportamiento facial y corporal del participante mientras realiza las pruebas. Para ello, en el séptimo artículo se estudia la posibilidad de combinar test psicométricos informatizados conductuales con técnicas de reconocimiento de patrones y visión por ordenador. Se observa que es posible encontrar determinados patrones de movimientos que mejoran las valoraciones de impulsividad. Estos hallazgos abren nuevas líneas de investigación que serán exploradas en los próximos años.

Abstract

This thesis analyses the benefits that can be obtained when pattern recognition techniques are utilized for solving several existing problems in the research areas of health and behavioral psychology. In the field of health, it is shown that these techniques are capable of improving, in more than 10%, the accuracy obtained by the current techniques in identifying suicidal behavior. Moreover, it is also exposed that these techniques can build accurate scales for specific purposes by selecting the most suitable items from general purpose scales. On the other hand, in the research area of behavioral psychology, it is observed that recent pattern recognition techniques are able to identify the most suitable candidates in the recruitment processes, which allows to reduce considerably the companies' costs and, therefore, making easier their growth. In this area, it is also shown that these techniques can be used to verify the identity of the participant when some assessments during the recruitment process are realized online. This PhD thesis contains six articles in which several statistical and pattern recognition techniques are applied. The techniques are linear regression, stepwise linear regression, linear and quadratic discriminant analysis, Fisher discriminant analysis, boosting, decision trees, support vector machines and the Lars-en algorithm. These techniques were used with different purposes. Support vector machine and the Lars-en algorithm showed to obtain accurate results.

This thesis also investigates the development of new discriminative predictors by analyzing the facial and corporal behavior of the examinee while performing some behavioral computerized tasks. This goal is achieved by combining pattern recognition and computerized psychometrical tests with computer vision techniques. Obtained results show that it is possible to find patterns related to corporal movement that can be used to improve the assessment of impulsivity. These findings open new future research lines that will explored in the following years.

1. Introducción

Un *patrón* puede definirse como una relación, estructura o regularidad inherente en una determinada fuente de datos (Shawe-Taylor & Cristianini, 2004). Detectar patrones permite realizar predicciones sobre otros datos provenientes de la misma fuente. En los últimos años, se han desarrollado varias técnicas basadas en el reconocimiento de patrones. Estas técnicas han proporcionado soluciones a distintos problemas existentes en diferentes áreas de investigación. A modo de ejemplo, el algoritmo *Adaboost,* que permite combinar clasificadores sencillos para crear uno con mucha mayor precisión, es actualmente la técnica más utilizada en identificar regiones que contienen caras en imágenes digitales (Pavani, Delgado, & Frangi, 2010). El número de aplicaciones donde se han aplicado las Máquinas de Vectores Soporte (*support vector machines*) es innumerable: visión por ordenador (Cheng, Zheng, & Qin, 2005), finanzas (Tay & Cao, 2001), genética (Guyon, Weston, Barnhill, & Vapnik, 2002) o clasificación de textos (Tong & Koller, 2002). Una técnica desarrollada recientemente, que está capturando una considerable atención, es el denominado aprendizaje profundo (Deep learning) (Hilton, Osindero, & Teh, 2006)

A pesar de la amplia utilización que han tenido estas técnicas en numerosas áreas de investigación, su adopción en las ciencias del comportamiento y de la salud ha sido marginal hasta hace relativamente poco. Probablemente, uno de los principales motivos que han limitado su uso es que, aunque estas técnicas mejoran considerablemente la precisión de las clasificaciones, pueden reducir (o imposibilitar) la interpretación de los resultados ya que algunas de ellas son consideradas cajas negras (Suykens, 2001). Además, por otro lado, muchos profesionales de la salud están acostumbrados a resumir las respuestas proporcionadas por sus pacientes a diferentes cuestionarios a través de la suma de las puntuaciones de sus ítems y en muchos casos desconocen las posibilidades de estas técnicas.

En relación con lo expuesto, esta tesis persigue analizar y mostrar las ventajas de introducir estas técnicas recientes de reconocimiento de patrones en las ciencias de la salud y del comportamiento. En estos momentos, es un tema muy actual y de gran interés como se puede apreciar en el hecho que, en el año 2012, la "National Science Foundation" financió, con 10 millones de dólares, un proyecto destinado a desarrollar técnicas que midan y analicen comportamientos infantiles mediante el desarrollo de un sistema multimodal basado en el reconocimiento de patrones[1]. En este proyecto están participando universidades de reconocido prestigio como son el Instituto Tecnológico de Georgia, Carnegie Mellon o el Instituto de Tecnología de Massachusetts. Los investigadores de estos centros han

[1] http://www.cbs.gatech.edu/

denominado a esta disciplina “ciencia computacional del comportamiento”, combinación de Ciencias de la Computación y Psicología.

En particular, en esta tesis, se persigue identificar patrones que permitan proporcionar soluciones a dos problemas actuales. Por un lado, en el área de la salud, se busca encontrar patrones que caractericen la conducta suicida. Como se explicará más adelante en detalle, este problema es de gran relevancia debido a las consecuencias sociales y económicas que produce. Por otro lado, en el área de las ciencias del comportamiento, se pretende identificar patrones que caractericen a los mejores candidatos en un proceso de selección. Además, en esta segunda área, también se persigue desarrollar un método que permita reducir el tiempo de los procesos de selección, mediante la realización online de parte de la evaluación, garantizando que la persona que realiza las pruebas a distancia es el candidato.

Estos objetivos se tratarán como problemas de clasificación. Por tanto, intervendrán principalmente dos elementos: las variables predictivas y las técnicas de clasificación o clasificadores. En relación a las variables predictivas, en la primera parte de esta tesis, se utilizaran los predictores más comúnmente usados en el campo de la Psicología. Es decir, las respuestas proporcionadas por los participantes a los ítems de diversas escalas. Estas respuestas estarán acompañadas, en ocasiones, de una etiqueta que indicará si el examinado que ha proporcionado las respuestas posee determinada característica (por ejemplo, si ha realizado un intento de suicidio recientemente o si ha mostrado un buen rendimiento en determinada tarea). Estas variables predictivas serán utilizadas por varias técnicas de clasificación para encontrar patrones que posibiliten una posible solución a los problemas estudiados. Debido a que cada técnica posee sus ventajas y sus inconvenientes, se emplearan varias de ellas y se combinaran sus resultados para mejorar su precisión e interpretabilidad. Cuando las variables predictivas posean una etiqueta, se utilizaran técnicas supervisadas como las Máquinas de Vectores Soporte, el boosting o técnicas estadísticas multivariantes clásicas. Cuando no se disponga de estas etiquetas, se utilizaran técnicas no supervisadas como el análisis factorial o técnicas desarrolladas en el área de la psicometría. Esta distinción, entre técnicas supervisadas y no supervisadas, marcará la estructura de esta tesis. En particular, el resto de esta tesis se estructura de la siguiente forma.

El *capítulo 2* se centra en el uso de *técnicas supervisadas* de reconocimiento de patrones para identificar, por un lado, individuos con riesgo de cometer suicidio y, por otro lado, para caracterizar a los mejores candidatos en un proceso de selección de una conocida empresa de seguros española.

El *capítulo 3* aborda problemas relacionados a las dos áreas de estudio desde un punto de vista *no supervisado*. En el lado de las ciencias de la salud, se buscará encontrar patrones que caractericen las consecuencias de atentados terroristas tanto en víctimas directas como indirectas. En el área de las ciencias del comportamiento, se buscará desarrollar un método que permita verificar la identidad del participante cuando este realiza una prueba a distancia.

El *capítulo 4*, motivado por los resultados obtenidos en los capítulos anteriores, se centra en el desarrollo de nuevos predictores que solucionen algunas de las de las debilidades que poseen las escalas. Para ello, se fusionaran test psicométricos existentes con técnicas de visión por ordenador y reconocimiento de patrones.

Esta tesis concluye en *capítulo 5* donde inicialmente se realiza un resumen de los principales hallazgos junto con una exposición de sus limitaciones. Tras ello, ya que esta tesis no es un punto final sino un punto seguido, se muestran cuáles van a ser las líneas de investigación futuras; algunas de las cuales ya se han iniciado.

2. Utilización de técnicas supervisadas de reconocimiento de patrones, desarrolladas recientemente, en valoraciones de personalidad

Hasta hace relativamente poco, la mayoría de los trabajos aparecidos en el área de Psicología clínica que tenían como objetivo descubrir patrones que mejoraran la caracterización de las enfermedades mentales presentaban varias debilidades. La primera de ellas consistía en que los trabajos aparecidos se centraban principalmente en demostrar, mediante test estadísticos clásicos, la relación entre el predictor propuesto y determinada enfermedad mental (Turvey et. al., 2002; Callicot, Bertolino, Egan, Mattay, Langheim, & Weinberger, 2000). Aunque estos test permitían identificar ciertos predictores potenciales, no era posible valorar su utilidad clínica ya que se centraban únicamente en el establecimiento de relaciones entre las variables dejando de lado la puesta a prueba de modelo de predicción y clasificación. Tampoco permitían identificar patrones complejos. La segunda debilidad, de mayor importancia, es que diversas publicaciones mostraban la precisión con la que identificaban determinadas enfermedades utilizando únicamente un conjunto de datos (Keilp et. al., 2006). Es decir, utilizaban el mismo conjunto de datos para construir su clasificador y para evaluarlo. Esto puede producir el fenómeno conocido como sobreajuste o lo que es lo mismo obtener una clasificación muy buena en los datos analizados pero una pobre precisión en datos futuros.

La primera parte de este capítulo aplica y adapta técnicas actuales de reconocimiento de patrones, simulando escenarios reales, con el objetivo de encontrar patrones que puedan ser utilizados en la labor clínica diaria. En particular, se centrará en la identificación de riesgo suicida, aunque la metodología utilizada puede ser aplicada a otros diagnósticos. La elección de riesgo suicida fue debida a diversos motivos. En primer lugar, la problemática relacionada con el suicidio le convierte en uno de los principales problemas en el área de la salud. Esto puede apreciarse cuando se observan las distintas estadísticas asociadas con el suicidio. Por ejemplo, se ha reportado que cada 40 segundos se produce un suicidio en el mundo (World report on violence and health, 2002). Además, el suicidio es la tercera causa de mortalidad entre personas con edades comprendidas entre los 15 y los 44 años (Holmes, Crane, Fennell, & Williams, 2007). Aparte del coste humano, el suicidio también tiene un coste económico que ha sido estimado en 33.000 millones de dólares únicamente en los Estados Unidos (Coreil, Bryant, & Henderson, 2001). El segundo motivo, es que pudimos disponer de una base de datos excepcional que recogía las respuestas a distintos cuestionarios psicométricos de más de 1000 individuos que incluía tanto controles (donantes de sangre y pacientes psiquiátricos sin historia de suicidio) como individuos que acababan de realizar un intento de suicidio. Esta base de datos nos posibilitaría descubrir si las técnicas recientes de reconocimiento de patrones pueden identificar de forma precisa la tendencia suicida. En caso afirmativo, esto permitiría reducir las tasas de suicidio ya que se ha mostrado que es posible reducir estas tasas entre un 25% y un 75% con los tratamientos adecuados a personas identificadas en riesgo (Isaacson, 2000; Hampton, 2010).

La segunda parte de este capítulo se centra en la aplicación de estas técnicas en el área de selección de personal. En gran medida, de los procesos de selección de personal depende la expansión y supervivencia de las empresas (Kangis & Lago, 1997). A modo de ejemplo, según un reciente estudio desarrollado por Careerbuilder[2], cerca del 40% de las empresas tuvieron un coste de 25000 dólares como consecuencia de un mal proceso de selección. Además, este estudio también indica que un 25% de las empresas sufrió, por la misma razón, un coste superior a 50000 dólares. En el sector de ventas, se ha observado que únicamente un 20% de los vendedores son responsables del 80% de las ventas de la compañía (Greenberg & Greenberg, 1980). El estudio presentado se desarrolló junto con una conocida compañía española de seguros.

2.1 Caracterización de la conducta suicida

Como se ha comentado en la introducción de este capítulo, esta primera parte pretende descubrir patrones que permitan identificar la conducta suicida. Para ello, la investigación desarrollada se ha realizado de forma secuencial en tres etapas donde los resultados obtenidos en las dos primeras guiaron la investigación de la segunda y tercera etapa respectivamente. Como se mostrará a continuación, el resultado final de las tres investigaciones ha sido la creación de una herramienta para identificar riesgo suicida con alta precisión, especificidad, y sensibilidad.

2.1.1 Identificación de riesgo suicida a través de escalas de impulsividad y desordenes de personalidad

La investigación desarrollada en esta primera parte de esta tesis comenzó buscando la respuesta a la pregunta más directa que podíamos formular: ¿pueden mejorar las técnicas de reconocimiento de patrones la precisión obtenida por los procedimientos actuales que resumen las respuestas al cuestionario por la suma (en ocasiones, ponderada) de la puntuación obtenida en cada uno de los ítems? Para responder esta primera pregunta, analizamos la respuesta de una muestra de 879 sujetos (345 individuos que habían cometido

---

2

http://www.careerbuilder.com/share/aboutus/pressreleasesdetail.aspx?sd=12/13/2012&id=pr730&ed=12/31/2012.

recientemente un intento de suicidio, 384 donantes de sangre y 150 pacientes psiquiátricos sin historia de suicidio) a dos cuestionarios de evaluación de tendencia suicida: la versión 11 de la escala de impulsividad de Barratt (BIS-11) en su adaptación española y el Internacional Personality Disorder Evaluation Screening Questionnaire (IPDE-SQ). El cuestionario BIS-11 es uno de los cuestionarios más utilizados en la evaluación de la personalidad. Consta de 30 ítems que evalúan tres factores de impulsividad (atencional, motora y planificación). Por otro lado, el IPDE-SQ consta de 75 ítems que evalúan los diferentes desordenes de personalidad del ICD-10 y DSM-IV entre los que se puede encontrar inestabilidad emocional, personalidad limite o paranoide. Estos dos cuestionarios se eligieron porque tanto la impulsividad como los desórdenes de personalidad han sido ampliamente relacionados con el suicidio (Gvion & Apter, 2011; Soloff, Lynch, Kelly, Malone, & Mann, 2014; Kohut, 2013).

Con el objetivo de responder esta pregunta inicial, se aplicaron diversos clasificadores a estas dos escalas. Antes de realizar los análisis, se dividieron aleatoriamente los datos en tres conjuntos: entrenamiento, evaluación y test. En esta configuración, el conjunto de entrenamiento es utilizado para estimar los distintos parámetros de los clasificadores, el conjunto de evaluación para determinar el umbral que mejor discrimina al grupo control de las personas que habían realizado un intento de suicidio y, finalmente, el conjunto de test es utilizado para evaluar el rendimiento del clasificador. De esta forma, se espera que los resultados que se obtengan posteriormente en la actividad clínica diaria sean similares a los obtenidos en el conjunto de test. Además del método tradicional, se analizó el rendimiento de cuatro tipos de clasificadores: las Máquinas de Vectores Soporte (SVM), boosting, el análisis discriminante de Fisher (FLDA) y el análisis discriminante lineal. El primero de estos cuatro clasificadores, las Máquinas de Vectores Soporte, fue seleccionado porque actualmente es considerado una de las técnicas más precisas (Van Gestel, Suykens, Baesens, Viaene, Vanthienen, Dedene, & Vandewalle (2004). Boosting, consistente en combinar clasificadores sencillos para obtener un clasificador no lineal, fue seleccionado porque proporcionó buenos resultados en previos estudios (Pavani, Delgado, & Frangi, 2010). Finalmente, los últimos dos clasificadores fueron seleccionados del área de Estadística: el primero fue seleccionado porque, a diferencia de los dos primeros, permite estimar la importancia de cada uno de los ítems en la clasificación; el segundo fue seleccionado por su simplicidad y eficiencia computacional. Los resultados obtenidos mostraron que las técnicas de reconocimiento de patrones eran capaces de mejorar en más de un 10% la tasa de clasificación correcta obtenida con las técnicas actuales. Además, también se observó que las Máquinas de Vectores Soporte eran la técnica que mejores resultados obtenía. Obtuvo un 63.68% de clasificación correcta cuando era aplicada al BIS-11 y un 73.43% cuando era aplicada al IPDE-SQ. Esto, indicó que el cuestionario IPDE-SQ parece tener más información que el BIS-11 para diferenciar controles de personas que han realizado un intento de suicidio.

A continuación se muestra en su totalidad este trabajo que fue publicado en la revista *Artificial Intelligence In Medicine* (Delgado-Gomez, Blasco-Fontecilla, Alegria, Legido-Gil, Artes-Rodriguez, & Baca-Garcia, 2011).

Artificial Intelligence in Medicine 52 (2011) 165–168

Contents lists available at ScienceDirect

Artificial Intelligence in Medicine

journal homepage: www.elsevier.com/locate/aiim

Improving the accuracy of suicide attempter classification

David Delgado-Gomez [a,*], Hilario Blasco-Fontecilla [b], AnaLucia A. Alegria [c], Teresa Legido-Gil [b], Antonio Artes-Rodriguez [d], Enrique Baca-Garcia [b]

[a] Department of Statistics, Universidad Carlos III, Calle Madrid 126, Getafe 28903, Spain
[b] Department of Psychiatry, Jimenez Diaz Foundation, IIS, CIBERSAM, Madrid, Spain
[c] Department of Psychiatry, Ramon y Cajal Hospital, IRYCIS, CIBERSAM, Madrid, Spain
[d] Department of Signal Theory and Communications, Universidad Carlos III, Leganes, Spain

ARTICLE INFO

Article history:
Received 29 November 2009
Received in revised form 9 May 2011
Accepted 9 May 2011

Keywords:
Support vector machines
Boosting
Suicide prediction
Barratt's impulsiveness scale
International personality disorder evaluation screening questionnaire

ABSTRACT

Objective: Psychometrical questionnaires such as the Barrat's impulsiveness scale version 11 (BIS-11) have been used in the assessment of suicidal behavior. Traditionally, BIS-11 items have been considered as equally valuable but this might not be true. The main objective of this article is to test the discriminative ability of the BIS-11 and the international personality disorder evaluation screening questionnaire (IPDE-SQ) to predict suicide attempter (SA) status using different classification techniques. In addition, we examine the discriminative capacity of individual items from both scales.
Materials and methods: Two experiments aimed at evaluating the accuracy of different classification techniques were conducted. The answers of 879 individuals (345 SA, 384 healthy blood donors, and 150 psychiatric inpatients) to the BIS-11 and IPDE-SQ were used to compare the classification performance of two techniques that have successfully been applied in pattern recognition issues, Boosting and support vector machines (SVM) with respect to linear discriminant analysis, Fisher linear discriminant analysis, and the traditional psychometrical approach.
Results: The most discriminative BIS-11 and IPDE-SQ items are "I am self controlled" (Item 6) and "I often feel empty inside" (item 40), respectively. The SVM classification accuracy was 76.71% for the BIS-11 and 80.26% for the IPDE-SQ.
Conclusions: The IPDE-SQ items have better discriminative abilities than the BIS-11 items for classifying SA. Moreover, IPDE-SQ is able to obtain better SA and non-SA classification results than the BIS-11. In addition, SVM outperformed the other classification techniques in both questionnaires.

1. Introduction

Suicide is a considerable public health issue. In 2002, suicide accounted for nearly one million people worldwide [1]. In spite of this negative figure, it is encouraging to know that it is possible to prevent suicide [2]. The use of the correct preventive measures to individuals at increased risk can reduce suicide rate by 25% [3]. Therefore, the development of a system capable of detecting suicide attempters (SA) may aid clinicians in the better distribution of resources (i.e. dedicating most of the available resources to monitor these at-risk individuals) and ultimately decrease suicide rates.

In order to effectively predict suicide, it is vital to identify the clinical characteristics associated with increased risk for suicide. One of the best predictors of suicide is the presence of a previous suicide attempt [4]. Several other characteristics shown to discriminate SA from non-SA include socio-demographic factors (i.e. gender, age) [5,6], family history of suicidal behavior [7], high impulsiveness [8] or aggressiveness [9], and presence of psychiatric disorders such as major depressive disorders (MDD) [10], personality disorders (PD) [11] and drug addiction [12]. Although several studies tested the discriminative capability of some scales, only few of them reported their classification accuracy. Because accuracy is not reported in most studies, it is difficult to state if these factors (i.e. previous suicide attempt, presence of MDD) are useful to predict suicidal behaviors. Moreover, the few studies addressing suicide prevention from a classification point of view offered discouraging results [13]. For instance, Pokorny [14] conducted a study aimed at identifying people who committed suicide using linear discriminant analysis. Although some features were highly correlated with suicide, suicide prediction was not feasible due to the low sensitivity and specificity of classification procedures. Later, Motto et al. [15] concluded that even if the idea of a questionnaire is able to predict suicide in subjects is difficult to attain, the development of clinically derived scales might pay off in the long run.

Impulsiveness and PD are well-known risk factors of suicidal behavior [9,11]. Therefore, the use of scales evaluating impulsive-

* Corresponding author. Tel.: +34 916249848; fax: +34 916249849.
E-mail address: ddelgado@est-econ.uc3m.es (D. Delgado-Gomez).

0933-3657/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.artmed.2011.05.004

ness and PD may help in detecting patients at risk of suicide. The main objective of this article is to test the discriminative ability of the Barratt's impulsiveness scale version 11 (BIS-11) and the international personality disorder evaluation screening questionnaire (IPDE-SQ) to predict SA status using different classification techniques. We compare the classification performance of two relatively novel techniques that have successfully been applied in pattern recognition issues, boosting and support vector machines (SVM) [16], with respect to linear discriminant analysis (LDA), Fisher linear discriminant analysis (FLDA) and the traditional psychometrical approach consisting in the sum of all items scores. In addition, we examine the discriminative capacity to detect SA using both scales.

2. Methods

2.1. Sample

In this article, a set of 879 consenting subjects (345 SA: 534 non-SA) was used. The 345 SA were admitted to the emergency departments at two General Hospitals in Madrid, Spain, between January 1999 and January 2003. According to the National Institute of Mental Health for research on suicidal behavior, a suicide attempt was defined as "a self-destructive behavior with intent to end one's life independent of resulting damage" [17,18]. The non-SA group was composed by healthy blood donors ($n = 384$) and psychiatric inpatients ($n = 150$). The present study was approved by the appropriate Ethical Committees.

2.2. Questionnaires

Impulsiveness is a well-known risk factor of suicidal behavior [19,20]. The Spanish version of the BIS-11 was used to measure lifetime impulsiveness. The BIS-11 is a 30-items self-report scale, with each item allowing 4-point ratings. This scale is widely used in the study of impulsiveness in SA [21].

PD were ascertained with the diagnostic and statistical manual of mental disorders (DSM-IV, 4th ed.) version of the IPDE-SQ. The DSM-IV is a 77 True/False self-report screening questionnaire that yields ten PD. As described in DSM-IV, PD are further grouped into three clusters. Cluster A PD are characterized by odd or eccentric behaviors and includes paranoid, schizoid and schizotypal PD. Cluster B PD, the most dramatic, emotional or erratic disorders includes antisocial, borderline, histrionic, and narcissistic PD. Finally, Cluster C related to anxious mood and fear includes avoidant, dependent, and obsessive–compulsive PD. The DSM-IV is a well-validated, reliable, and easy to handle instrument to diagnose PD [22]. This instrument has been previously used in the study of PD and suicidal behaviors with consistent results [23,24].

2.3. Techniques

Here we briefly describe the different techniques used to identify SA.

- Traditional psychometrical approach
  The traditional approach consists of adding up the scores of the individual items in order to obtain a global score. The higher the BIS-11 global score, the more likely the individual exhibits impulsive suicidal behavior [25,26].
- Linear discriminant analysis (LDA)
  This technique emerged as a possible solution to the problem of classifying an individual. This technique was selected because of its simplicity and computational efficiency [27].
- Fisher linear discriminant analysis (FLDA)
  For a two class problem, a standard FLDA maximizes the ratio of the between-class scatter to the within-class scatter matrices. This technique considers the different discriminative abilities of each item [28].
- Boosting
  Boosting [29] is a meta-algorithm that refers to a method of producing a strong classifier by additively combining a set of weak classifiers. Different types of weak classifiers such as linear regressors [30], decision trees [31] or thresholding of an individual feature [32] have been used in the literature. In the present study, the weak classifiers were generated by applying a threshold to an individual feature. Boosting is one of the most important recent developments in prediction [16]. This method has successfully been applied in several classification problems [30,32].
- Support vector machines (SVM)
  SVM are non-linear classifiers that have captured the attention of the scientific community in the last years because of their excellent performance in many classification problems [33,34]. In this study, the Gaussian kernel is used.

2.4. Procedure

Two experiments aimed at discriminating between SA and non-SA were conducted. The first experiment tested the discriminative capacity of the BIS-11, whereas the second experiment focused on the IPDE-SQ. In both experiments, data were randomly divided into three sets: training set, evaluation set and test set. The training set was used to build the different classifiers. Because the accuracy of classifiers depends on different parameters, the evaluation set was used to tune them. These parameters were the σ and C parameters for the SVM, the number of iterations T for boosting, the prior for the LDA, and the thresholds for the FLDA and the traditional psychometrical approach. These parameters were set to the values that maximize the accuracy in the evaluation set. Once these parameters were established, the test set was used to obtain the accuracy of the different classifiers. For each of the following analyses, 100 repetitions of this set-up were conducted to obtain statistically meaningful results.

3. Results

3.1. Experiment 1 (BIS-11)

The mean absolute difference of the scores obtained for each single item in the SA and non-SA groups are shown in Table 1. The values obtained indicate significant differences in the discriminative capacity of the BIS-11 items. For instance, the three most discriminative items were item 6 ("I am self controlled"), item 15 ("I act on impulse") and item 27 ("I often have extraneous thoughts when thinking"), whereas the less discriminative was item 28 ("I am more interested in the present than the future").

The classification accuracy, specificity and sensitivity mean values obtained with the BIS-11 are displayed in Table 2. All the

Table 1
Mean absolute differences (range 0–3) of the different items of the BIS-11 between SA and non-SA groups.

Item 6	**Item 15**	**Item 27**	**Item 7**	**Item 22**	**Item 11**
0.98	0.80	0.77	0.74	0.72	0.65
Item 16	**Item 8**	**Item 29**	**Item 25**	**Item 5**	**Item 12**
0.64	0.58	0.56	0.53	0.53	0.50
Item 14	**Item 2**	**Item 10**	**Item 21**	**Item 3**	**Item 20**
0.46	0.45	0.43	0.39	0.37	0.34
Item 9	**Item 4**	**Item 1**	**Item 23**	**Item 18**	**Item 17**
0.33	0.33	0.29	0.27	0.26	0.26
Item 30	**Item 19**	**Item 24**	**Item 13**	**Item 26**	**Item 28**
0.26	0.25	0.23	0.05	0.05	0

Table 2
Average classification accuracy (*acc*) together with the average specificity (*spc*) and the average sensitivity (*sens*) using the BIS-11.

Technique		SVM (%)	Boosting (%)	FLDA (%)	LDA (%)	Traditional (%)
Eval.	*acc*	78.81	76.83	77.92	77.50	73.72
	spc	86.95	85.42	86.50	85.26	84.03
	sens	66.20	63.55	64.64	65.48	57.77
Test	*acc*	76.71	74.41	76.13	75.94	71.46
	spc	85.13	82.95	84.41	83.66	82.38
	sens	63.68	61.21	63.31	63.98	54.56

classification techniques achieved better classification accuracy than the traditional psychometrical approach.

3.2. Experiment 2 (IPDE-SQ)

This experiment was aimed at examining the differential accuracy attained between the IPDE-SQ and the BIS-11. Before conducting the experiment, two items of the IPDE-SQ were removed. The item "I've never threatened suicide or injured myself on purpose" (item 25) was excluded to avoid a tautological problem. The item 49 of the IPDE-SQ "I often seek advice or reassurance about everyday decisions" was also removed due to a high rate of non-responding subjects, hence diminishing considerably our statistical power. The three most discriminative items for the IPDE were item 40 ("I feel often empty inside"), item 43 ("I have tantrums or angry outburst") and item 1 ("I usually get fun and enjoyment out of life") whereas the less discriminative was item 7 ("I get upset when I heard bad news about someone I know"). All results obtained are displayed in Table 3.

Table 3
Mean absolute differences (range 0–1) of the different items of the IPDE-SQ between SA and non-SA groups.

Item 40	**Item 43**	**Item 1**	**Item 66**	**Item 33**	**Item 42**	**Item 26**
0.56	0.44	0.42	0.39	0.39	0.38	0.36
Item 34	**Item 4**	**Item 27**	**Item 16**	**Item 51**	**Item 53**	**Item 60**
0.35	0.34	0.33	0.32	0.32	0.32	0.32
Item 75	**Item 8**	**Item 36**	**Item 39**	**Item17**	**Item 14**	**Item 24**
0.32	0.31	0.30	0.28	0.28	0.27	0.27
Item 70	**Item 6**	**Item 64**	**Item 73**	**Item 31**	**Item 2**	**Item 15**
0.27	0.26	0.24	0.24	0.24	0.22	0.22
Item 58	**Item 56**	**Item 69**	**Item 13**	**Item 61**	**Item 63**	**Item 54**
0.22	0.22	0.22	0.21	0.21	0.20	0.20
Item 76	**Item 28**	**Item 49**	**Item 71**	**Item 52**	**Item 68**	**Item 72**
0.20	0.19	0.19	0.18	0.18	0.18	0.18
Item 77	**Item 12**	**Item 45**	**Item 74**	**Item 32**	**Item 30**	**Item 29**
0.17	0.17	0.17	0.15	0.15	0.15	0.14
Item 11	**Item 23**	**Item 62**	**Item 20**	**Item 44**	**Item 65**	**Item 9**
0.13	0.13	0.13	0.12	0.11	0.10	0.10
Item 59	**Item 38**	**Item 55**	**Item 41**	**Item 22**	**Item 67**	**Item 46**
0.09	0.08	0.08	0.08	0.08	0.07	0.07
Item 19	**Item 35**	**Item 48**	**Item 57**	**Item 33**	**Item 18**	**Item 5**
0.07	0.06	0.05	0.05	0.04	0.03	0.03
Item 10	**Item 37**	**Item 21**	**Item 7**			
0.03	0.03	0	0			

Table 4
Average classification accuracy (*acc*) together with the average specificity (*spc*) and the average sensitivity (*sens*) using the IPDE-SQ.

Technique		SVM (%)	Boosting (%)	FLDA (%)	LDA (%)
Eval.	*acc*	82.29	81.29	78.78	78.51
	spc	86.67	86.32	85.34	83.82
	sens	75.52	73.51	68.63	70.28
Test	*acc*	80.26	79.12	77.00	77.24
	spc	84.68	84.18	83.18	82.37
	sens	73.43	71.28	67.43	69.29

The average classification results of the IPDE-SQ are displayed in Table 4. The best classification techniques for the IPDE-SQ were SVM and boosting.

Finally, we tested if the combination of all the items of the BIS-11 and IPDE-SQ offered better classification accuracy results than using these instruments individually. SVM were again the best classification technique. However, the accuracy improvement was minimal (80.74 vs. 80.26).

4. Discussion

Our results suggest that the traditional psychometrical approach of considering all items as equally important to calculate a global score might not be the best option. Our findings suggest that some items of these questionnaires are more valuable than others in the identification of SA. Additionally, our results showed than the mean absolute different values between the scores obtained for the SA and no-SA groups of the IPDE-SQ (range 0–1) were higher than the values observed in the BIS-11 (range 0–3). This is interesting, because the BIS-11 is more frequently used than the IPDE-SQ for the study of suicidal behaviors [35,36]. Our results highlight the importance of evaluating other personality dimensions apart from impulsiveness in the assessment of suicide risk.

The most discriminative BIS item was lack of self-control, which is in accordance with previous reports [37]. Regarding to the IPDE-SQ, both empty feelings [38] and temper tantrums [39] were the items that better characterized SA. These items are typically displayed by subjects with borderline personality disorder. This disorder has traditionally been associated to suicidal behavior [40].

Concerning the accuracy, the IPDE-SQ provided better classification results than the BIS-11. Our results suggest that the IPDE-SQ items are more valuable than the BIS-11 items to discriminate between SA and non-SA. With respect to the classifiers, all the techniques offered better classification results than the traditional psychometric classification approach, but SVM attained the best results.

5. Conclusions

This article focused on the capacity of some classification techniques to increase the power of suicidal behavior assessment of two general psychometric questionnaires, namely the BIS-11 and the IPDE-SQ. Traditional multivariate techniques such as LDA and FLDA, and non-linear approaches such as boosting and SVM were used. Although all these techniques improved the classification of SA compared with the traditional approach, SVM technique outperformed the others. In addition, the obtained results suggest that the IPDE-SQ is a better instrument to classify SA than BIS-11.

References

[1] Baca-Garcia E, Perez-Rodriguez MM, Keyes KM, Oquendo MA, Hassin DS, Grand BF, et al. Suicidal ideation and suicide attempts in the United States: 1991–1992 and 2001–2002. Mol Psychiatry 2010;15(3):250–9.

[2] Jamison KR. Suicide and bipolar disorder. J Clin Psychiatry 2000;61(9):47–51.

[3] Isaacson G. Suicide prevention—a medical breakthrough? Acta Psychiatr Scand 2000;102(2):103–17.
[4] Leon AC, Friedman RA, Sweeney JA, Brown RP, Mann JJ. Statistical issues in the identification of risk factors for suicidal behavior: the application of survival analysis. Psychiatry Res 1990;31(1):99–108.
[5] Smith JC, Mercy JA, Conn JM. Marital status and the risk of suicide. Am J Public Health 1988;78(1):78–80.
[6] Spicer RS, Miller TR. Suicide acts in 8 states: incidence and case fatality rates by demographics and method. Am J Public Health 2000;90(12):1885–91.
[7] Sarchiapone M, Carli V, Janiri L, Marchetti M, Cesaro C, Roy A. Family history of suicide and personality. Arch Suicide Res 2009;13(2):178–84.
[8] Patton JH, Stanford MS, Barratt ES. Factor structure of the Barratt impulsiveness scale. J Clin Psychol 1995;51(6):768–74.
[9] Mann JJ, Waternaux C, Haas GL, Malone KM. Toward a clinical model of suicidal behavior in psychiatric patients. Am J Psychiatry 1999;156(2):181–9.
[10] Beck AT, Steer RA, Carbin MG. Psychometric properties of the Beck Depression Inventory: twenty-five years of evaluation. Clin Psychol Rev 1998;8(1):77–100.
[11] Mann AH, Raven P, Pilgrim J, Khanna S, Velayudham A, Suresh KP, et al. An assessment of the standardized assessment of personality as a screening instrument for the international personality disorder examination: a comparison of informant and patient assessment for personality disorder. Psychol Med 1999;29(4):985–9.
[12] Miller M, Hemenway D, Bell NS, Yore MM, Amoroso PJ. Cigarette smoking and suicide: a prospective study of 300,000 male active-duty Army soldiers. Am J Epidemiol 2000;151(11):1060–3.
[13] Goldney RD, Spence ND. Is suicide predictable? Aust N Z J Psychiatry 1987;21(1):3–4.
[14] Pokorny AD. Prediction of suicide in psychiatric patients. Arch Gen Psychiatry 1983;40(3):249–57.
[15] Motto JA, Heilbron DC, Juster RP. Development of a clinical instrument to estimate suicide risk. Am J Psychiatry 1985;142(6):680–6.
[16] Friedman JH. Recent advances in predictive (machine) learning. J Classification 2006;23(2):175–97.
[17] O'Carrol PW, Berman AL, Maris RW, Moscicki EK, Tanney BL, Silverman MM. Beyond the tower of Babel: a nomenclature for suicidology. Suicide Life Threat Behav 1996;26(3):237–52.
[18] Silverman MM, Berman AL, Sanddal ND, O'Carrol PW, Joiner TE. Rebuilding the tower of Babel: a revised nomenclature for the study of suicide and suicidal behaviors. Part 1: Background, rationale, and methodology. Suicide Life Threat Behav 2007;37(3):248–63.
[19] Mann JJ. The biology of suicide. Nat Med 1998;4(1):25–30.
[20] Mann JJ. The biology of impulsivity and suicidality. Psychiatr Clin North Am 2000;23(1):11–25.
[21] Oquendo MA, Baca-Garcia E, Graver R, Morales M, Montalvan M, Mann JJ. Spanish adaptation of the Barratt Impulsiveness Scale (BIS-11). Eur J Psychiatry 2001;15(3):147–55.
[22] Egan V, Austin E, Elliot D, Patel D, Charlesworth P. Personality traits, personality disorders and sensational interests in mentally disordered offenders. Crim Psychol 2003;8(1):51–62.
[23] Blasco-Fontecilla H, Oquendo MA, Baca-Garcia E. Recurrence of self-harm and severity of personality disorder. Acta Psychiatr Scand 2009;120(1):82–3.
[24] Tyrer P. Recurrence of self-harm and severity of personality disorder. Acta Psychiatr Scand 2009;120(1):82.
[25] Zouk H, Tousignant M, Seguin M, Lesage A, Turecki G. Characterization of impulsivity completers: clinical, behavioral and psychosocial dimensions. J Affect Disord 2006;92(2–3):195–204.
[26] Gut-Fayand A, Dervaux A, Olie JP, Loo H, Poirier MF, Krebs MO. Substance abuse and suicidality in schizophrenia: a common risk factor linked to impulsivity. Psychiatry Res 2001;102:65–72.
[27] Duda RO, Hart P, Stork DG. Pattern classification. 2nd ed. Wiley-Interscience; 2000.
[28] Delgado-Gomez D, Fagertun J, Ersboll B, Sukno FM, Frangi AF. Similarity-based Fisherfaces. Pattern Recog Lett 2009;30(12):1110–6.
[29] Schapire RE. Theoretical views of boosting and applications. Proc Int Conf Algorithm Learn 1999;1720:13–25.
[30] Skurichina M, Duin F. Bagging, boosting and the random subspace method for linear classifiers. Pattern Anal Appl 2002;5(2):121–35.
[31] Dietterich TG. An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting and randomization. Mach Learn 2000;40(2):139–57.
[32] Pavani SK, Delgado-Gomez D, Frangi AF. Haar-like features with optimally weighted rectangles for rapid object detection. Pattern Recog 2010;43(1):160–72.
[33] Shawe-Taylor J, Cristianini N. Support vector machines and other kernel-based learning methods. Cambridge: Cambridge University Press; 2000.
[34] Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. Mach Learn 2002;46(1):389–422.
[35] Baca-Garcia E, Diaz-Sastre C, García-Resa E, Blasco-Fontecilla H, Braquehais-Conesa D, Oquendo MA, et al. Suicide attempts and impulsivity. Eur Arch Psychiatry Clin Neurosci 2005;255(2):152–6.
[36] Barratt ES, Stanford MS, Dowdy L, Liebman MJ, Kent TA. Impulsive and premeditated aggression: a factor analysis of self-reported acts. Psychiatry Res 1999;86(2):163–73.
[37] Wu CS, Liao SC, Lin KM, Tseng MM, Wu EC, Liu SK. Multidimensional assessments of impulsivity in subjects with history of suicidal attempts. Compr Psychiatry 2009;50(4):315–21.
[38] Klonsky ED. What is emptiness? Clarifying the 7th criterion for borderline personality disorder. J Pers Disord 2008;22(Aug (4)):418–26.
[39] Hendin H, Maltsberger JT, Haas AP, Szanto K, Rabinowicz H. Desperation and other affective states in suicidal patients. Suicide Life Threat Behav 2004;34(4):386–94 [Winter].
[40] Berk MS, Jeglic E, Brown GK, Henriques GR, Beck AT. Characteristics of recent suicide attempters with and without Borderline Personality Disorder. Arch Suicide Res 2007;11(1):91–104.

2.1.2 Caracterizando el comportamiento suicida a través de escalas de eventos vitales y desordenes de personalidad

Los resultados obtenidos en la investigación anterior indicaron que las técnicas de reconocimiento de patrones pueden aportar una contribución significativa en la valoración psicológica de los pacientes. También planteaban la pregunta de si sería posible mejorar la precisión obtenida utilizando predictores más adecuados. Para resolver esta pregunta, este segundo trabajo está basado en el concepto ampliamente aceptado que la conducta de un individuo es el resultado de la interacción de factores estables y consistentes de su personalidad (rasgos) con el contexto (Mann, Waternaux, Haas, & Malone, 1999). En concreto, la pregunta principal que pretendía responderse en este segundo trabajo es: ¿Sería posible mejorar los resultados obtenidos anteriormente combinando la escala IPDE-SQ (aproximación simple a la caracterización del rasgo) con la escala de acontecimientos vitales de Holmes & Rae (influencia del contexto)? Por otro lado, desde el punto de vista metodológico, también se consideró incluir técnicas de selección de variables. Utilizar un número reducido de variables mejora la capacidad de generalización del clasificador. Además, pasar un cuestionario con un número menor de ítems es más beneficioso para departamentos médicos que disponen de un tiempo limitado para realizar las valoraciones, como pueden ser los servicios de urgencias.

La escala de valoración de ajuste social de Holmes-Rahe (Holmes & Rae, 1967), usada en este trabajo, consta de 43 ítems que hacen referencia a diversos acontecimientos que puede haber sufrido el entrevistado en los últimos años. Estos acontecimientos están relacionados con su vida personal (cambio importante en hábitos alimenticios, cambio importante de hábitos de sueño), con sus relaciones de pareja/familia (divorcio, casarse, problemas con la familia política) y con su vida laboral (retiro, despido). Aunque esta escala fue inicialmente desarrollada para medir el estrés que sufre el examinado, en un estudio previo se mostró que es posible utilizarla para evaluar tendencia suicida modificando adecuadamente la importancia dada a cada ítem. (Blasco-Fontecilla, Delgado-Gómez, Legido-Gil, De Leon, Perez-Rodriguez, & Baca-García, 2012).

Igual que en el trabajo anterior, se pasaron las respuestas proporcionadas a estos dos cuestionarios por una muestra similar de entrevistados (526 controles, 347 individuos que acababan de realizar un intento de suicidio) a cinco clasificadores diferentes. Estos cinco clasificadores fueron seleccionados por diferentes razones. Al igual que antes, el discriminante lineal fue escogido por su sencillez y eficacia computacional. Además, puede ser considerado como una referencia base en relación a las precisiones de clasificación. Como se comentó anteriormente, las Máquinas de Vectores Soporte es considerada una de las técnicas de clasificación más precisas en la actualidad. Los arboles de decisión permiten realizar clasificadores no lineales a partir de reglas sencillas. Finalmente se añadieron dos clasificadores que permitían seleccionar variables. La regresión lineal con selección de variables y el método Lars-en. Este último consiste en una regresión lineal con penalizadores que es un tema que actualmente está siendo investigado en gran detalle.

El resultado principal obtenido es que combinando estas dos escalas era posible obtener una clasificación de 86.4%; casi un 10% de mejora respecto a nuestros resultados anteriores. Este resultado indica la necesidad de combinar personalidad y contexto a la hora de identificar la conducta suicida. Además, otro aspecto a tener en cuenta es que este número fue conseguido con la técnica Lars-en que permitía reducir el número de variables utilizadas. Esta técnica seleccionó un promedio de 34 ítems de los 90 disponibles.

Los resultados de este trabajo fueron publicados en la revista Neurocomputing (Delgado-Gomez, Blasco-Fontecilla, Sukno, Ramos-Plasencia, & Baca-García, 2012). El trabajo en su totalidad se expone a continuación.

Neurocomputing 92 (2012) 3–8

Contents lists available at SciVerse ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Suicide attempters classification: Toward predictive models of suicidal behavior

David Delgado-Gomez [a,*], Hilario Blasco-Fontecilla [b], Federico Sukno [c], Maria Socorro Ramos-Plasencia [d], Enrique Baca-Garcia [e]

[a] Department of Statistics, Carlos III University, Calle Madrid 126, Getafe 28903, Spain
[b] Department of Psychiatry, Jimenez Diaz Foundation, IIS, CIBERSAM, Madrid, Spain
[c] Centre for Image Processing and Analysis, Dublin City University, Dublin 9, Ireland
[d] Department of Linguistics, I.E.S Gran Tarajal, Fuerteventura, Spain
[e] Department of Psychiatry, Columbia University Medical Center, New York, USA

ARTICLE INFO

Available online 24 February 2012

Keywords:
Support vector machines
Lars-en
Suicide prediction
Holmes–Rahe social readjustment rating scale
International personality disorder examination screening questionnaire

ABSTRACT

Suicide is a major public health issue with considerable human and economic cost. Previous attempts to delineate techniques capable of accurately predicting suicidal behavior proved unsuccessful. This paper aims at classifying suicide attempters (SA) as a first step toward the development of predictive models of suicidal behavior. A sample of 883 adults (347 SA and 536 non-SA) admitted to two university hospitals in Madrid, Spain, between 1999 and 2003 was used. Five multivariate techniques (linear regression, stepwise linear regression, decision trees, Lars-en and support vector machines) were compared with regard to their capacity to accurately classify SA. These techniques were applied to the Holmes–Rahe social readjustment rating scale and the international personal disorder examination screening questionnaire. Combining both scales, the Lars-en and stepwise linear regression techniques achieved 83.6% and 82.3% classification accuracy, respectively. In addition, these classification results were obtained using less than half of the available items. Multivariate techniques demonstrated to be useful in classifying SA using a combination of life events and personality criteria with reasonable accuracy, sensitivity and specificity.

1. Introduction

Suicide prevention is a worldwide health service priority [1]. In 2002, it was estimated that a person committed suicide every 40 s [2]. In recent years, the number of suicides has increased becoming the third leading cause of death worldwide among people aged 15–44 [3]. Besides the human cost, suicidal behavior (suicide attempts and suicide completion) conveys considerable economical burden. The annual cost of suicidal behavior has been estimated in $33 billions only in the United States [4]. Despite these negative figures, it has been shown that it is possible to prevent suicide [5]. Treating subjects at risk with the appropriate preventive measures, such as cognitive behavior therapies [6], can reduce suicide rate up to 25% [7]. Consequently, the development of a system able to identify suicide attempters (SA) might help to decrease suicide rates.

In order to build such a system, it is fundamental to identify the factors that are most closely related to suicidal behavior. The presence of a suicide attempt is the most compelling predictor of suicide [8]. In addition, family history of suicidal behavior [9], socio-demographic factors (i.e. gender, age) [10,11], high impulsiveness [12], aggressiveness [13], depression [14], personality disorders (PD) [15] and drug use [16] are associated with an increased risk of suicide. Unfortunately, all these potentially predictive factors have shown rather low specificity.

An explanatory variable, which has not received enough attention, is the presence of life events (LE). This is surprising, as most suicides are preceded by LE [17] and its predictive capacity has outperformed that of biological factors [18]. Those LE that are most frequently associated with suicidal behavior include interpersonal conflicts (i.e. marital separation, divorce), personal losses, physical illness and financial problems [19]. It has been shown that inability to cope with LE may lead to attempted or completed suicides [20].

The present study has three objectives: (i) to test whether we can accurately classify SA using a scale measuring LE; (ii) to find a subset of LE that discriminates SA from non-SA, and can easily be adopted by clinicians; and (iii) to test if combining LE with PD scales outperforms the classification results obtained by each scale separately.

2. Related work

As stated above, almost all previous research on suicidal behavior have been focused in finding explanatory variables. However, these studies did not quantify the importance of these

* Corresponding author.
E-mail address: ddelgado@est-econ.uc3m.es (D. Delgado-Gomez).

0925-2312/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2011.08.033

variables. Pokorny [21] provided the first classification results on suicide prediction. His results were disappointing as he reported that accuracy, specificity and sensitivity did not reach 70%. Therefore, he concluded that there was not a combination of items that could provide an adequate identification of suicidal populations. Recently, Hendin et al. [22] slightly improved Pokorny's results. They obtained an accuracy of 71.67% with a specificity of 74% and sensitivity of 60% in predicting risk for suicidal behavior. This improvement might be explained by the use of a different questionnaire (Affective States Questionnaire) as they used a basic classification technique (they simply added up the scores of each item). Therefore, this raises the question of whether these results might be improved by using multivariate classifiers.

3. Scales

LE within two years preceding a suicide attempt were coded according to the standardized and adapted Spanish version [23] of the Holmes–Rahe social readjustment rating scale (SRRS) [24]. The SRRS includes 43 LE ranked according to the degree of severity. Death of spouse is considered the most severe item while minor violations of the law is the mildest.

PD were ascertained with the diagnostic and statistical manual of mental disorders (DSM-IV, fourth edition) version of the international personality disorder examination screening questionnaire (IPDE-SQ), which is a well-validated, reliable, and easy to handle instrument to diagnose PD [25]. The DSM-IV IPDE-SQ is a 77 true/false self-report screening questionnaire that yields 10 PD. PD are grouped into three clusters. Cluster A PD are characterized by odd or eccentric behavior and includes paranoid, schizoid and schizotypical PD. Cluster B PD includes antisocial, borderline, histrionic, and narcissistic PD. Finally, cluster C is related to anxious mood and fear and includes avoidant, dependent, and obsessive-compulsive PD.

4. Techniques

The accuracy of the following multivariate classification techniques in classifying SA and non-SA was investigated:

- Linear regression (LR) was selected to set a classification baseline. LR is a simple multivariate technique that assumes a linear relationship between the dependent variable and the regressors. Namely, $y=\beta\cdot\vec{x}$, where $\vec{x}$ is the vector of regressors and y is the dependent variable. The coefficients of this regression line are obtained so that the sum of the squared errors between the predicted values estimated for a training dataset and their real response values is minimized. These coefficients can be obtained by least square minimization or by maximum likelihood methods [26]. The closed formulation for these parameters is given by $\beta=(X^tX)^{-1}X^tY$, where X is the matrix with the regressors of each datum of the training set as rows and Y is the vector of the associated responses. This technique has been shown to be similar to linear discriminant analysis when the class label is used as dependent variable [27].
- Stepwise linear regression (SLR) [28] might be considered as a combination of linear regression and feature selection. Usually regarded as a generic method with many possible variants in the selection procedure, in most cases it is built upon forward selection. In this study, one of the most popular versions of this technique was used [29]. It consists in adding, at each step, the regressor that shows the highest correlation with the dependant variable or prediction error. After each addition, all variables in the selected set are checked to ensure that their contribution is significant enough. Otherwise they are removed. An interesting advantage of the stepwise method with respect to the basic LR is that, if a similar or higher classification accuracy is obtained with a reduced set of features, there could be the clinical benefit of reducing the time required for each patient.
- Decision trees (DT). A decision tree is a non-linear classification algorithm. Decision trees sequentially divide the data into different subgroups so that all or almost all the elements in a subgroup belong to the same class. Different criteria can be used for the data partitioning. In this paper, maximum deviance reduction was used. A more detailed description of decision trees can be found in Breiman [30].
- Elastic net (Lars-en) [31] is a variant of SLR that usually improves its performance. Also based on forward selection, this method modifies the optimization function by adding constraints to the L1 and L2 norms of the vector of coefficients. This technique minimizes $|y-X\beta|^2+\lambda_1|\beta|_1+\lambda_2|\beta|_2$, where $|\beta|_1=\sum_{j=1}^{p}|\beta_j|$ and $|\beta|_2=\sum_{j=1}^{p}\beta_j^2$. The response is centered and the predictors are standardized before applying the technique. This rather simple modification has proven to be very powerful, and can be understood as a generalization of other two popular methods: least angle regression (Lars) [32] and least absolute shrinkage and selection operator (Lasso) [33].
- Support vector machines (SVM) [34] are non-linear classifiers that have shown outstanding performance in several classification tasks [35]. If the m-dimensional data $\vec{x}_i$, $i=1,\ldots,M$, which belong to class y_i ($y_i\in\{-1,1\}$) are linearly separable, SVMs compute a decision function $D(\vec{x})=w^t\vec{x}+b$, such that, for $i=1,\ldots,M$,

$$w^t\vec{x}_i+b=\begin{cases}\geq 1 & \text{for } y_i=1\\ \leq -1 & \text{for } y_i=-1\end{cases} \tag{1}$$

w being an m-dimensional vector and b a bias term.

If data were linearly separable, the hyperplanes $D(\vec{x})=w^t\vec{x}+b=a$, $-1<a<1$, would achieve perfect classification. From the infinite number of decision functions that satisfy Eq. (1), SVMs select the one that maximizes the distance between the separating hyperplane and the nearest training datum to the selected hyperplane. This distance is called the margin of the classifier.

However, the assumption that the data are linearly separable is rarely fulfilled. In this case, non-negative slack variables ξ_i, $i=1,\ldots,M$, are introduced and the optimization problem consists in minimizing

$$\frac{1}{2}\|w\|^2+C\sum_{i=1}^{M}\xi_i \tag{2}$$

subject to the constraints:

$$y_i(w^t\vec{x}_i+b)\geq 1-\xi_i \quad \text{for } i=1,\ldots,M \tag{3}$$

In order to obtain non-linear separability, the original data are usually mapped implicitly into a high dimensional space. In this extended case, the optimization consists of maximizing:

$$\sum_{i=1}^{M}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{M}\alpha_i\alpha_j y_i y_j K(\vec{x}_i,\vec{x}_j) \tag{4}$$

subject to the constraints

$$\sum_{i=1}^{M}y_i\alpha_i=0$$

$$\xi_i\geq 0,\quad i=1,\ldots,M \tag{5}$$

where $K(\vec{x}_i, \vec{x}_j)$ is a kernel. There are many ways to define it, the most popular being the Gaussian, which is used in this paper:

$$k(\vec{x}_i, \vec{x}) = e^{-\|\vec{x}_i - \vec{x}\|^2/(2\sigma^2)} \tag{6}$$

5. Results

In order to discriminate between SA and non-SA, we used data from 883 subjects. Participants were 18 years or older and provided written informed consent before participating in the study. The cases included 347 SA (220 women and 127 men) admitted to two university hospitals in Madrid, Spain, between 1999 and 2003. A suicide attempt was defined as a self-destructive behavior with the intention of ending one's life, independent of the resulting damage [36]. Mean age of SA was 37.1. Non-SA (n=536) included 124 psychiatric inpatients (73 women and 51 men) hospitalized for a reason different from suicidal behavior and without a history of suicidal behavior, and 412 healthy controls (blood donors) (153 women and 259 men) from the same hospitals. Mean age of controls was 36.44. The study was approved by the appropriate ethics committee.

Three experiments were conducted. The first two were aimed at (1) assessing the classification accuracy that can be achieved and (2) finding the more discriminative items from a classification point of view, when the SRRS and IPDE-SQ questionnaires are used separately. The third experiment investigated if a better classification accuracy might be obtained by combining both questionnaires.

In all experiments, data were randomly divided into three sets: training set, evaluation set and test set. The training set (179 non-SA, 115-SA) was used to build the previously described classifiers. Because the accuracy of classifiers depends on different parameters, the evaluation set (179 non-SA, 115-SA) was used to determine the optimal settings.[1] These parameters were set to the values that maximize the accuracy in the evaluation set and kept fixed for the test set (178 non-SA, 115-SA) which was used to obtain the results that are reported for the different classifiers. In all cases, reported results correspond to average statistics from 100 repetitions of the process just described.

5.1. IPDE-SQ

Before conducting this analysis, two items from the IPDE-SQ were removed. Item 25 (I've never threatened suicide or injured myself on purpose) was excluded to avoid a tautological problem. Item 49 (I often seek advice or reassurance about everyday decisions) was removed due to a high rate of non-responding subjects, which considerably diminished the statistical power.

Classification results are shown in Table 1. All techniques rendered good results. Lars-en and SVM attained slightly better performance compared with the other three techniques. Moreover, it was observed that Lars-en and SLR techniques used an average of 30.4 and 10.3 items, respectively. The other two classifiers used all items. In order words, Lars-en and SLR techniques obtained similar classification results using less items, therefore being less time-consuming.

The 10 items receiving the highest absolute average weights are displayed in Table 2. The most determining personality items to accurately classify SA were "I feel often empty inside" and "I have tantrums or angry outburst" from borderline PD, which agrees with previous literature. Borderline PD is one of the most frequent diagnosis among suicide attempters [37]. In accordance with our previous research and literature [38], the present study also suggests that some personality criteria from different PD across clusters, and not only borderline PD, are associated with an increased risk of suicidal behavior. On the other hand, those subjects presenting two IPDE-SQ items, namely "I usually get fun and enjoyment out of life" and "people have a high opinion on me" were less likely classified as SA. In other words, these items might be protective of suicidal behavior.

Table 1
Average classification accuracy (acc), specificity (spc) and sensitivity (sens) using the IPDE.

Technique	LR (%)	SLR (%)	DT (%)	Lars-en (%)	SVM (%)
Eval.					
acc.	79.2	80.8	78.7	83.7	82.2
spc.	83.2	84.9	83.5	87.9	86.3
sens.	72.9	74.4	71.2	77.1	76.0
Test					
acc.	77.3	79.0	77.4	80.4	80.3
spc.	81.5	83.0	82.6	85.2	84.4
sens.	70.7	72.7	69.3	73.0	73.9

Table 2
IPDE-SQ items receiving highest absolute average weights using the Lars-en technique.

1. I feel often empty inside (Borderline PD)	1.77
2. I usually get fun and enjoyment out of life (Schizoid PD)	−1.13
3. I have tantrums or angry outburst (Borderline PD)	0.78
4. I have been the victim of unfair attacks on my character or reputation (Paranoid PD)	0.60
5. I can't decide what kind of person I want to be (Borderline PD)	0.49
6. I usually feel uncomfortable or helpless when I'm alone(Dependent PD)	0.45
7. I think my spouse (or lover) may be unfaithful to me (Paranoid PD)	0.45
8. My feelings are like the weather: they're always changing (Histrionic PD)	0.44
9. People have a high opinion on me (Narcissistic PD)	−0.37
10. I take chances and do reckless things (Antisocial PD)	0.34

5.2. SRRS

Some LE were rare in our sample. For instance, only 0.5% people presented with "Jail term or probation", and none of them showed "change in religious activities", "minor financial loan", or "change in schools". By considering these LE in this analysis, we would have included some LE with limited clinical interest, and the covariance matrix would be singular, thus causing problems in the parameter estimations. LE with a frequency lower than 5% were removed from the study, which resulted in a set of 14 items kept in our analysis. The 14 LE considered after the purging process, and the percentage of SA among people who satisfied this LE are shown in Table 3.

Classification results are shown in Table 4. Again, all techniques rendered satisfactory classification results.

The average weights of the items using the Lars-en technique are shown in Fig. 1. The most relevant LE to classify SA were "changing number of arguments with spouse or life partner", "marital separation", "personal injuries or illness", in decreasing order of importance. Both marital separation and somatic illness are associated with suicide [19]. Separated and ill people are more

[1] These parameters were the σ and C parameters for the SVM, the number of features and the regularization parameters for Lars-en, priors and minimal number of observations per tree leaf for DT and thresholds for LR and SLR.

Table 3
SRSS items used to classify SA.

Description	SA satisfying the LE/subjects satisfying the LE (percentage)
Death of a close family member	48/96 (50%)
Revision of personal habits	50/63 (79.7%)
Change in health of immediate family member	44/199 (22.1%)
Change in financial state	64/99 (64.7%)
Gain a new family member	108/165 (65.5%)
Change in residence	65/123 (52.9%)
Marital separation	99/133 (74.4%)
Fired at work	31/51 (60.8%)
Change to different line of work	32/63 (50.8%)
Change in number of arguments with spouse or life partner	146/171 (85.4%)
Change in work hours or conditions	62/134 (46.3%)
Personal injury or illness	84/121 (69.4%)
Change in social activities	43/71 (60.6%)
Begin or end studies	24/61 (39.3%)

Table 4
Average classification accuracy (acc), specificity (spc) and sensitivity (sens) using the SRRS.

Technique	LR (%)	SLR (%)	DT (%)	Lars-en (%)	SVM (%)
Eval.					
acc.	81.1	80.6	80.1	82.0	80.9
spc.	83.7	83.3	80.3	85.0	87.7
sens.	76.9	76.3	79.8	77.3	70.3
Test					
acc.	79.1	79.3	78.2	79.1	78.8
spc.	81.9	82.0	78.5	82.1	86.2
sens.	74.8	75.0	77.6	74.4	67.4

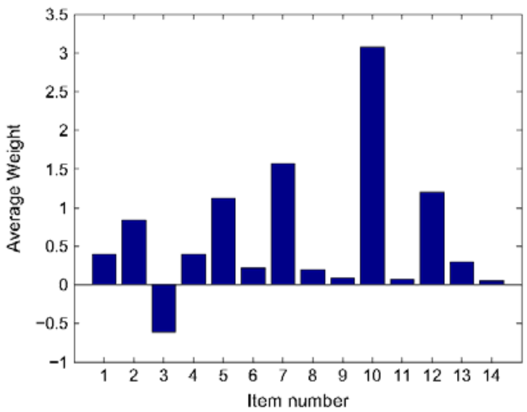


Fig. 1. Average weights of the items used by the Lars-en technique using the SRRS.

likely to show suicidal behavior [39]. The Lars-en technique used an average of 10.8 LE, whereas SLR used an average of 6.3 LE.

5.3. IPDE-SQ plus SRRS

Classification results are displayed in Table 5. As expected, the combination of both scales attained better classification results. The Lars-en technique nearly classified 84% of SA. The items with the highest average weights are the same that the ones identified in the two preceding analyses. On average, Lars-en and SLR used 34.9 and 14 items, respectively.

Table 6 displays the computational time of each of the different techniques. Lars-en was the slowest technique during the training (approximately 20 min). This time is affordable for a training step. More importantly, the fact that the classification is obtained in real time indicates that all the methods can be used in clinical settings.

Table 5
Average classification accuracy (acc), specificity (spc) and sensitivity (sens) using the IPDE and SRRS questionnaires together.

Technique	LR (%)	SLR (%)	DT (%)	Lars-en (%)	SVM (%)
Eval.					
acc.	81.6	83.6	81.1	86.4	84.4
spc.	84.6	87.2	85.6	89.5	87.0
sens.	77.0	78.1	74.0	81.6	80.3
Test					
acc.	80.2	82.3	79.3	83.6	82.7
spc.	83.1	85.7	83.9	87.0	85.9
sens.	75.7	76.9	72.1	78.3	77.7

Table 6
Computational time (in seconds).

Method	Training	Test
LR	$10^{\,6}$	$10^{\,6}$
SLR	0.062	$10^{\,6}$
DT	77.17	$10^{\,6}$
Lars-en	1267.54	$10^{\,6}$
SVM	184.9	0.09

6. Conclusion

Previous research has unfortunately failed to provide clinicians with a tool to accurately identify subjects at risk of suicide. This study is a first step toward the development of predictive models of suicidal behavior with reasonable accuracy, sensitivity, and specificity. Compared with previous efforts, our results offer acceptable accuracy, specificity, and sensitivity in classifying SA [21,22], and are easy to implement in the clinical arena.

From a clinical point of view, it is important to stress that both Lars-en and SLR techniques considerably reduced the number of items. Lars-en and SLR techniques were capable of accurately classifying SA using 35 and 14 items (both scales combined), respectively. Thus, both techniques allow for a more time-efficient use of the IPDE-SQ and SRRS. Clinicians may decide to use the items selected by either technique depending on the setting where the evaluation is carried out. For instance, items selected by SLR might be preferred in the psychiatric emergency department, while items selected by Lars-en could be used in a setting where time pressure is not a determinant factor (e.g. acute inpatient units).

Finally, marital discord and two borderline PD items (“I feel often empty inside” and “I have tantrums or angry outburst”) were the best discriminative items.

References

[1] I.M. Hunt, N. Kapur, J. Robinson, J. Shaw, S. Flynn, H. Bailey, et al., Suicide within 12 months of mental health service contact in different age and diagnostic groups: national clinical survey, Br. J. Psychiatry 188 (2006) 135–142.

[2] World Health Organization (WHO), World Report on Violence and Health, WHO Publications, Geneva, 2002.

[3] E.A. Holmes, C. Crane, M.J.V. Fennell, J.M.G. Williams, Imagery about suicide in depression "Flash-forwards"? J. Behav. Ther. Exp. Psychiatry 38 (4) (2007) 423–434.
[4] J. Coreil, C.A. Bryant, J.N. Henderson, Social and Behavioral Foundations of Public Health, Sage Publications, CA, USA, 2001.
[5] K.R. Jamison, Suicide and bipolar disorder, J. Clin. Psychiatry 61 (2000) 47–51.
[6] G.K. Brown, T.T. Have, G.R. Henriques, S.X. Sie, J.E. Hollander, A.T. Beck, Cognitive therapy for the prevention of suicide attempts, J. Am. Med. Assoc. 294 (2005) 563–570.
[7] G. Isaacson, Suicide prevention a medical breakthrough? Acta Psychiatr. Scand. 102 (2000) 103–117.
[8] W. Coryell, M. Schlesser, The dexamethasone suppression test and suicide prediction, Am. J. Psychiatry 158 (5) (2001) 748–753.
[9] M. Sarchiapone, V. Carli, L. Janiri, M. Marchetti, C. Cesaro, A. Roy, Family history of suicide and personality, Arch Suicide Res. 13 (2) (2009) 178–184.
[10] P.R. Duberstein, Y. Conwell, K.R. Conner, S. Eberly, E.D. Caine, Suicide at 50 years of age and older: perceived physical illness, family discord and financial strain, Psychol. Med. 34 (1) (2004) 137–146.
[11] M.E. Heikkinen, E.T. Isometsa, H.M. Aro, S.J. Sarna, J.K. Lonnqvist, Age-related variation in recent life events preceding suicide, J. Nerv. Ment. Dis. 183 (5) (1995) 325–331.
[12] J.H. Patton, M.S. Stanford, E.S. Barratt, Factor structure of the Barratt impulsiveness scale, J. Clin. Psychol. 51 (6) (1995) 768–774.
[13] J.J. Mann, C. Waternaux, G.L. Haas, K.M. Malone, Toward a clinical model of suicidal behavior in psychiatric patients, Am. J. Psychiatry 156 (2) (1999) 181–189.
[14] N. Horesh, J. Sever, A. Apter, A comparison of life events between suicidal adolescents with major depression and borderline personality disorder, Compr. Psychiatry 44 (4) (2003) 277–283.
[15] H. Blasco-Fontecilla, E. Baca-Garcia, K. Dervic, M.M. Perez-Rodriguez, M.D. Saiz-Gonzalez, J. Saiz-Ruiz, et al., Severity of personality disorders and suicide attempt, Acta Psychiatr. Scand. 119 (2) (2009) 149–155.
[16] S.S. Welch, M.M. Linehan, High-risk situations associated with parasuicide and drug use in borderline personality disorder, J. Pers. Disord. 16 (6) (2002) 561–569.
[17] J. Cooper, L. Appleby, T. Amos, Life events preceding suicide by young people, Soc. Psychiatry Psychiatr. Epidemiol. 37 (6) (2002) 271–275.
[18] E. Baca-Garcia, C.P. Parra, M.M. Perez-Rodriguez, C. Diaz-Sastre, R. Reyes Torres, J. Saiz-Ruiz, et al., Psychosocial stressors may be strongly associated with suicide attempts, Stress Health 23 (2007) 191–198.
[19] K. Kolves, A. Varnik, B. Schneider, J. Fritze, J. Allik, Recent life events and suicide: a case–control study in Tallinn and Frankfurt, Soc. Sci. Med. 62 (11) (2006) 2887–2896.
[20] J.T. Cavanagh, D.G. Owens, E.C. Johnstone, Life events in suicide and undetermined death in south-east Scotland: a case–control study using the method of psychological autopsy, Soc. Psychiatry Psychiatr. Epidemiol. 34 (12) (1999) 645–650.
[21] A.D. Pokorny, Prediction of suicide in psychiatric patients. Report of a prospective study, Arch. Gen. Psychiatry 40 (3) (1983) 249–257.
[22] H. Hendin, R.K. Al Jurdi, P.R. Houck, S. Hughes, J.B. Turner, Role of intense affects in predicting short-term risk for suicidal behavior, J. Nerv. Ment. Dis. 198 (2010) 220–225.
[23] J.L. Gonzalez de Rivera, A. Morera, La valoración de sucesos vitales: adaptación española de la escala de Holmes y Rahe, Psiquis 4 (1983) 7–11.
[24] T.H. Holmes, R.H. Rahe, The social readjustment rating scale, J. Psychosom. Res. 11 (1967) 213–218.
[25] V. Egan. E. Austin. D. Elliot. D. Patel. P. Charlesworth. Personality traits. personality disorders and sensational interests in mentally disordered offenders, Crim. Psychol. 8 (1) (2003) 51–62.
[26] S. Chatterjee, A.S. Hadi, Influential observations, high leverage points, and outliers in linear regression, Stat. Sci. 1 (3) (1986) 379–416.
[27] J. Ye, Least squares linear discriminant analysis, in: Proceedings of the 24th International Conference on Machine Learning, 2007, pp. 1087–1093.
[28] D.G. Kabe, Stepwise multivariate linear regression, J. Am. Stat. Assoc. 58 (303) (1963) 770–773.
[29] P. Pudil, F.J. Ferri, J. Novovicova, J. Kittler, Floating search methods for feature selection with nonmonotonic criterion functions, in: Proceedings of the 12th IAPR International Conference on Computer Vision and Image Processing, vol. 2, 1994, pp. 279–283.
[30] L. Breiman, J. Friedman, R. Olshen, C. Stone, Classification and Regression Trees, Wadsworth, 1984.
[31] H. Zou, T. Hastie, Regularization and variable selection via the elastic net, J. R. Stat. Soc. B 67 (2) (2005) 301–320.
[32] B. Efron, T. Hastie, I. Johnstone, R. Tibshirani, Least angle regression, Ann. Stat. 32 (2) (2004) 407–499.
[33] R. Tibshirani, Regression shrinkage and selection via the lasso, J. R. Stat. Soc. B 58 (1) (1996) 267–288.
[34] J. Shawe-Taylor, C. Nello, Kernel Methods for Pattern Analysis, Cambridge University Press, 2004.
[35] I. Guyon, J. Weston, S. Barnhill, V. Vapnik, Gene selection for cancer classification using support vector machines, Mach. Learn. 46 (1) (2002) 389–422.
[36] P.W. O'Carroll, A.L. Berman, R.W. Maris, E.K. Moscicki, B.L. Tanney, M.M. Silverman, Beyond the Tower of Babel: a nomenclature for suicidology, Suicide Life Threat Behav. 26 (3) (1996) 237–252.
[37] D.W. Black, N. Blum, B. Pfohl, N. Hale, Suicidal behavior in borderline personality disorder: prevalence, risk factors, prediction, and prevention, J. Pers. Disord. 18 (3) (2004) 226–239.
[38] H. Blasco-Fontecilla, E. Baca-Garcia, K. Dervic, M.M. Perez-Rodriguez, J. Lopez-Castroman, J. Saiz-Ruiz, et al., Specific features of suicidal behavior in patients with narcissistic personality disorder, J. Clin. Psychiatry 70 (11) (2009) 1583–1587.
[39] K. Kolves, N. Ide, D. De Leo, Suicidal ideation and behaviour in the aftermath of marital separation: gender differences, J. Affect. Disord. 120 (1–3) (2010) 48–53.

David Delgado-Gomez received the M.Sc. degree in Mathematics from the Autonomous University of Madrid, in 1999 and the Ph.D. degree in applied Mathematics from the Technical University of Denmark in 2005. He spent three years as post-doc in the computational imaging lab at the Pompeu Fabra University of Barcelona. Currently, he is working as a Full-Time Professor in the Department of Statistics at the Carlos III University of Madrid, where he combines research and teaching. His research areas include applied multivariate statistics, machine learning, and pattern recognition. He has published several articles in different fields such as biometrics, dermatology, medical image analysis, psychiatry and psychology.

Hilario Blasco-Fontecilla, M.D., Ph.D., is Assistant Professor of victimology at Universidad Europea de Madrid and Assistant Professor of Psychiatry at Universidad Autonoma de Madrid. He is an expert on personality disorders and suicide. He received competitive funding from the Castilla del Pino Foundation to attending a one year course leading to a Diploma in Child and Adolescent Psychiatry at the Institute of Psychiatry, London, in 2003 under the supervision of Dr. Anula Nikapota and Dr. Eric Taylor. He is the recipient of a Rio Hortega Grant (Spanish Ministry of Health, competitive funding) at Jimenez Diaz Foundation, Madrid, which is a three year contract and equivalent to a research fellowship. He has published more than 30 chapters and a monography on personality disorders and suicidal behavior in a US editorial. He has also published more than 20 articles indexed at PUBMED in some of the top journals of the field such as The Journal of Clinical Psychiatry, The Journal of Clinical Psychopharmacology or Schizophrenia Research.

Federico M. Sukno was born in 1975 in Rosario, Argentina. He obtained the degree in Electrical Engineering at La Plata National University (Argentina, 2000) and a Ph.D. in Biomedical Engineering at Zaragoza University (Spain, 2008). After graduation, and before starting the Ph.D., he worked for about 3 years in the industry, as a field engineer in telecommunications and industrial control. Between 2005 and 2010, and in parallel with research activities within the research group for Computational Imaging and Simulation Technologies in Biomedicine (CISTIB), he was also an assistant professor at Pompeu Fabra University (Spain), where he has tough courses on Analog electronics and Information theory and coding for 5 years. Since November 2010 he is a post-doctoral researcher working jointly at Dublin City University and the Royal College of Surgeons in Ireland within the Face3D project, addressing the analysis of three-dimensional facial dysmorphology.

Maria Socorro Ramos-Plasencia is an expert in linguistics. She obtained the M.Sc. in linguistics from La Laguna University in 1998. Currently, she is starting her Ph.D. studies in the topic of suicidal behavior profiling.

Enrique Baca-Garcia is a well-trained psychiatrist with expertise in the areas of genetics and reproductive hormone influences on behavior and suicidal behavior. He has published extensively on these topics. Dr. Baca-Garcia has over 80 peer reviewed publications, several in top journals like Molecular Psychiatry, American Journal of Psychiatry, Neuropsychopharmacology, Biological Psychiatry, British Journal of Psychiatry, and Schizophrenia Research, to name a few. He has received two NARSAD Awards. Indeed, he was the first Spaniard Young Investigator awarded by NARSAD. He has mentored or is mentoring over 20 Ph.D. theses by Psychiatry residents. He has received more than 20 honors and awards including the AEP (Association of European Psychiatrists) Research Prize.

2.1.3 Combinando escalas para caracterizar la conducta suicida

Con el tercer y cuarto artículo se completa la investigación relacionada con la valoración de riesgo suicida (Blasco-Fontecilla, Delgado-Gomez, Ruiz-Hernandez, Aguado, Baca-García, & Lopez-Castroman, 2012; Artieda-Urrutia et. al., 2015).

En el tercer artículo la pregunta a responder es la que surge de manera inmediata tras la realización de los dos trabajos anteriores: ¿Cuál sería la mejor clasificación que pudiéramos alcanzar si combinásemos todas las escalas disponibles? Como se intuye de los resultados obtenidos en el ejercicio anterior, es necesario utilizar un algoritmo que permita seleccionar las variables (ítems) más adecuadas. Por tanto, en este artículo únicamente consideramos el algoritmo Lars-en. Las escalas consideradas fueron las anteriormente utilizadas (BIS-11, IPDE-SQ, Holmes-Rahe) junto con la escala de agresividad de Brown y la inclusión del sexo y la edad del participante. La escala de agresividad de Brown fue incluida porque muchos trabajos han relacionado agresividad y suicido (Gvion & Apter, 2011). De igual forma, sexo y edad han sido frecuentemente relacionados con suicidio (Canetto, 1998; Shah, 2007). En un estudio previo, se observó que las edades de la muestra de estudio se pueden agrupar en dos clases utilizando una mezcla de Gaussianas (Blasco-Fontecilla et. al., 2012). El principal resultado de este estudio fue que era posible obtener una clasificación de 85.3% cuando se realizaba selección de ítems. Como producto derivado de este estudio, se propuso una escala consistente en 27 ítems. Como se intuía de los dos artículos precedentes, los ítems con un mayor peso en la valoración eran los pertenecientes a la escala IPDE-SQ y Holmes–Rahe, complementados con ítems de impulsividad y agresión. La edad también se retuvo como factor importante. Un aspecto a reseñar es que, a diferencia de las otras escalas existentes en la que una respuesta positiva a cada uno de los ítems aumenta el riesgo de suicidio, la escala desarrollada contiene ítems que reducen el riesgo suicida. Estos ítems, a los que podemos llamar protectores, incluyen autocontrol, planificación, confianza hacia uno mismo o autoprotección.

En el segundo artículo (Artieda-Urrutia et al., 2015) de este apartado se pretende conseguir una versión optimizada y más breve de la prueba de 27 ítems generada en el trabajo anterior. Se decidió aplicar una estrategia diferente de selección de ítems, que había resultado solo muy ligeramente menos eficiente en el último trabajo comentado: la regresión lineal por pasos con selección de variables retrospectivas. La nueva prueba, de solo 6 ítems dicotómicos, se aplica en un minuto, no alcanzó una precisión alta pero sus indicadores en la curva ROC fueron adecuados (sensibilidad del 80%, especificidad del 75% y área bajo la curva ROC del 88%, para discriminar entre los que habían tenido y no tenido intentos de suicidio). Estos resultados se obtuvieron en un estudio de validación cruzada que evita el sobreajuste.

A continuación se reproducen ambos artículos (Blasco-Fontecilla, Delgado-Gómez, Ruiz-Hernández, Aguado, Baca-García, & López-Castroman, 2012; Artieda-Urrutia et. al., 2015).

Journal of Psychiatric Research 46 (2012) 1272–1277

Contents lists available at SciVerse ScienceDirect

Journal of Psychiatric Research

journal homepage: www.elsevier.com/locate/psychires

Combining scales to assess suicide risk

Hilario Blasco-Fontecilla [a,*], David Delgado-Gomez [b], Diego Ruiz-Hernandez [c], David Aguado [d], Enrique Baca-Garcia [e,f], Jorge Lopez-Castroman [e,g]

[a] Department of Psychiatry, Puerta de Hierro Hospital, CIBERSAM, Calle Manuel de Falla 1, 28222 Majadahonda, Spain
[b] Department of Statistics, Carlos III University, Calle Madrid 126, Getafe 28903, Spain
[c] Department of Statistics, CUNEF University, Serrano Anguita, 8, 28004 Madrid, Spain
[d] Institute of Knowledge Engineering, Autonoma University, 28049 Madrid, Spain
[e] Department of Psychiatry, IIS-Jimenez Diaz Foundation, CIBERSAM, 28040 Madrid, Spain
[f] Department of Psychiatry, Columbia University Medical Center, 1051 Riverside Drive, New York, USA
[g] Inserm U888 & CHU Montpellier, Montpellier, France

ARTICLE INFO

Article history:
Received 20 November 2011
Received in revised form
22 May 2012
Accepted 19 June 2012

Keywords:
Lars-en algorithm
Suicidal behavior
Personality traits
Life events

ABSTRACT

Objectives: A major interest in the assessment of suicide risk is to develop an accurate instrument, which could be easily adopted by clinicians. This article aims at identifying the most discriminative items from a collection of scales usually employed in the assessment of suicidal behavior.
Methods: The answers to the Barrat Impulsiveness Scale, International Personality Disorder Evaluation Screening Questionnaire, Brown–Goodwin Lifetime History of Aggression, and Holmes & Rahe Social Readjustment Rating Scale provided by a group of 687 subjects (249 suicide attempters, 81 non-suicidal psychiatric inpatients, and 357 healthy controls) were used by the Lars-en algorithm to select the most discriminative items.
Results: We achieved an average accuracy of 86.4%, a specificity of 89.6%, and a sensitivity of 80.8% in classifying suicide attempters using 27 out of the 154 items from the original scales.
Conclusions: The 27 items reported here should be considered a preliminary step in the development of a new scale evaluating suicidal risk in settings where time is scarce.

1. Introduction

Suicide is a major health issue. One suicide is completed every 40 seconds, leading to approximately one million deaths every year worldwide (WHO, 2002). Moreover, suicide is the third most important cause of death worldwide among people aged 15–44 (Holmes et al., 2007). Notwithstanding human costs, the economic burden of suicidal behavior has been estimated annually in $33 billion in the United States (Coreil et al., 2001). Fortunately, suicidal behavior might be prevented to a great extent (Jamison, 2000). Treating subjects at risk with the appropriate preventive measures, such as cognitive behavior therapies (Brown et al., 2005) can reduce suicide rates up to 25% (Isaacson, 2000). More recently, a 75% reduction of suicide rates has been reported in a large depression care program (Hampton, 2010). In order to detect subjects at risk, researchers have investigated the factors underlying suicidal behavior. The most relevant risk factors are major depression (Mann et al., 1999b), high impulsiveness (Patton et al., 1995), aggressiveness (Mann et al., 1999b), personality disorders (Mann et al., 1999a), life events (Kolves et al., 2006), and social-demographic factors (Smith et al., 1988).

Unfortunately, most of these studies did not measure the effectiveness of the risk factors to identify subjects at risk. They just tested if there was a statistically significant relationship between the studied variable (e.g. high impulsiveness) and suicidal behavior. Therefore, the clinical usefulness of these studies is limited. One notable exception is the seminal Pokorny's article (Pokorny, 1983). Pokorny applied discriminant analysis to several features including, among others, socio-demographic variables, the 24 items of the brief psychiatric rating scale, and the items included in the nurses' observation scale for inpatient evaluation. Although it was an innovative approach, his results showed a weak performance, with accuracy, sensitivity and specificity levels below 70%. More recently, Hendin (Hendin et al., 2010) slightly improved these results achieving an accuracy of 71.67% with a specificity of 74% and a sensitivity of 60%. The improvement was basically due to the use of a different set of predictive variables, as they used a simple

* Corresponding author. Villalba MHC, Department of Psychiatry, Puerta de Hierro Hospital, CIBERSAM, Calle de Los Madroños 5, 28400 Collado Villalba, Spain. Tel.: +34 918505161; fax: +34 918514707.
E-mail address: hmblasco@yahoo.es (H. Blasco-Fontecilla).

0022-3956/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jpsychires.2012.06.013

Table 1
Sociodemographic characteristics of healthy controls, psychiatric inpatients, and suicide attempters.

Sociodemographic characteristics	Suicide attempters ($N = 249$) n (%)	Healthy controls ($N = 356$) n (%)	Psychiatric inpatients ($N = 81$) n (%)	Stats	df	P value
Sex (female)	157 (63.0)	131(36.7)	54 (66.7)	27.56	2	<0.001
Age (mean ± SD)	37.2 ± 14.5	42.1 ± 13.0	34.6 ± 10.7	13.18	2	<0.001
Marital status						
Single	113 (45.4)	172 (48.3)	42 (51.8)	46.08	4	<0.001
Married/cohabiting	91 (36.5)	174 (48.8)	26 (32.1)			
Separated/widowed	45 (18.1)	10 (2.7)	13 (16.0)			
Years of education						
< 8	72 (29.5)	73 (20.6)	22 (27.5)	14.97	4	0.005
9–12	114 (46.7)	147 (41.5)	32 (40.0)			
>12	58 (23.8)	134 (37.9)	26 (32.5)			
Employment status						
Unemployed	66 (27.0)	32 (9.3)	29 (35.8)	155.46	4	<0.001
Employed	117 (48.1)	308 (89.7)	34 (48.0)			
Disabled/retired	61 (24.9)	4 (1.0)	18 (22.2)			

Note: x^2 tests were applied for all comparisons except age (ANOVA).

classifier consisting on the sum of the individual scores associated to each variable. This study suggested that the use of a more suitable set of predictive variables together with the use of more sophisticated classifiers might improve the classification accuracy of people at risk of suicidal behavior. This intuition was confirmed by us (Delgado-Gomez et al., 2011) in a study aimed at discriminating between suicide attempters (SA) and non-SA. In this study, we used two personality scales as predictors, and a collection of modern classification techniques such as linear discriminant analysis, Fisher linear discriminant analysis, boosting, and support vector machines (SVM). The best results were obtained with SVM, which achieved a classification accuracy of 80.3%, with a specificity of 86.8% and a sensitivity of 76.1%. Recently, Stefansson et al. (Stefansson et al., 2010) have shown that the prediction of suicide can be improved by means of an appropriate selection of the items. However, their results were obtained *ad hoc*. Therefore, we have applied the Lars-en algorithm in order to automatically select the most discriminative items (Delgado-Gomez et al., 2012). Using the selected items of scales measuring life events and personality disorders, it was possible to achieve a classification accuracy of 83%. A question that remains open is the accuracy that could be obtained if the Lars-en algorithm were applied to a set of items assessing the most relevant risk factors for suicidal behavior.

The present study extends our previous findings and is conceived as a further step toward the development of more precise and reliable measures of suicide risk (García-Nieto et al., 2012). The major goal is to maximize the classification accuracy applying simultaneously Lars-en to sociodemographic factors and items from four scales measuring some of the most relevant risk factors for suicidal behavior (impulsive aggression, life events, and personality disorders). As a by-product, we developed a small set of items to classify subjects as SA or non-SA. This set of items might help to develop a tool to support clinical decisions with regard to suicide risk in settings where time is scarce.

2. Methods

2.1. Participants

To accomplish our objectives, data from 687 subjects were used. Participants were 18 years or older and provided written informed consent before participating in the study. Subjects that showed incapability to provide informed consent were excluded (e.g. presence of a life-threatening medical condition, significant organic brain disease). The cases included 249 first-time SA (157 women and 92 men) admitted to two university hospitals in Madrid, Spain, between 1999 and 2003. Non-SA ($n = 438$) included 81 psychiatric inpatients (54 women and 27 men) without current or past history of suicidal behavior and 357 healthy controls (blood donors; 131 women and 226 men) recruited in the same hospitals. Healthy controls had neither Axis I diagnoses nor a history of suicidal behavior. The appropriate ethics committee approved the study. The study was carried out in accordance with the latest version of the Declaration of Helsinki.

Mean age (±SD) of the SA, psychiatric controls, and healthy controls were 37.2 (±14.5), 42.1 (±13.0), and 34.6 (±10.7), respectively ($F = 13.18$; df = 2; $p < 0.001$). We used the Mini International Neuropsychiatric Interview (MINI), a short, easy to administrate, and efficient structured diagnostic interview to assess Axis I disorders in psychiatric inpatients and SA (Sheehan et al., 1998). Tables 1 and 2 show information with regard to sociodemographic and clinical factors of the study groups.

We also used the Risk-Rescue Rating Scale (RRRS), a 10-item interviewer-administered questionnaire that provides an estimate of the seriousness of a SA (Weisman and Worden, 1972). The first five items of the scale describe risk factors of a SA. The Risk Rating ranges from 5 (5–6 points indicate "low risk") to 15 (13–15 points indicate "high risk"). 81.3% of the SA were classified as low risk SA. Furthermore, item 15 of the Beck Suicidal Intent Scale (SIS) was implemented to elucidate the degree of premeditation (Beck et al., 1974). In 63% of SA there was no premeditation at all (impulsive SA) and only 23.7% thought on suicide more than three hours before the attempt.

Table 2
Comparison of suicide attempters ($n = 244$) and psychiatric inpatients ($n = 81$) on some mental disorders (MINI).

Axis I disorders (any mental disorder)	Suicide attempters ($N = 244$) n (%)	Psychiatric inpatients ($N = 81$) n (%)	x^2	df	P value
Psychotic disorder (current)	19 (7.8)	22 (27.5)	21.04	1	<0.001
Major depressive episode (current)	132 (54.3)	23 (28.4)	16.36	1	<0.001
Major depressive episode (recurrent)	31 (33.3)	7 (8.9)	14.86	1	<0.001
Dysthymia	25 (10.3)	4 (4.9)	2.13	1	0.144
Panic disorder (current)	15 (6.2)	6 (7.6)	0.20	1	0.657
Generalized anxiety disorder	32 (13.2)	21 (25.9)	7.23	1	0.007
Alcohol dependence	27 (11.1)	10 (12.3)	0.10	1	0.753
Drug dependence	19 (7.8)	3 (3.7)	1.63	1	0.202

2.2. Measures

We selected age and sex as socio-demographic features and four clinical scales. Both age and sex are well-known risk factors of suicidal behavior. Suicide attempts are 2–3-fold more frequent among women than among men (Canetto and Sakinofsky, 1998). In addition, age is reported as a risk factor for future attempts in most studies (Christiansen and Jensen, 2007), and an age pattern has been described for the first suicide attempt (Slama et al., 2009). The selected scales assess central features of suicidal behavior: personality, life events and the impulsive aggression construct. Personality and life events are nuclear factors to explain suicidal behavior in the context of the stress-diathesis model (Mann et al., 1999b). Impulsivity and aggression personality traits have been particularly associated with suicidal behavior (Baca-Garcia et al., 2005, 2006; Perroud et al., 2011).

- The 11th version of the Barrat Impulsiveness Scale (BIS-11) is a widely used measure of impulsiveness. In the present study we used the Spanish version (Oquendo et al., 2001). The BIS-11 is a 30-item self-reported Likert scale that comprises three subscales to assess cognitive, motor, and non-planned impulsiveness. Items are scored from 1 (rarely/never) to 4 (almost always/always).
- The International Personality Disorder Evaluation Screening Questionnaire (IPDE-SQ). The IPDE-SQ (Loranger et al., 1994) is a screening psychometric instrument designed to identify relevant traits and behaviors in the assessment of personality disorders according to the main international classifications of mental disorders. This questionnaire examines the presence in adults (if apparent for at least five years) of diagnostic criteria for any personality disorder and comprises 77 True/False self-report items.
- The Holmes and Rahe Social Readjustment Rating Scale (SRRS). Life events within two years preceding a suicide attempt were coded according to the standardized and adapted Spanish version (Gonzalez de Rivera and Morera, 1983) of the SRRS (Holmes and Rahe, 1967). The SRRS includes 43 life events ranked according to the degree of severity. Death of spouse is considered the most severe item while minor violations of the law is the mildest.
- The Brown–Goodwin Lifetime History of Aggression (BGHA). The BGHA scale is a 11-item questionnaire measuring how many times different types of aggressive behavior occurred across childhood, adolescence, and adulthood (Brown et al., 1979). It includes 11 questions over a large range of aggressive behaviors. Different scores are registered for childhood, adolescence and adulthood. Subjects were requested to consider each of the 11 aggressive behaviors into a 4-point Likert scale (0 = never; 1 = rarely; 2 = occasionally; 3 = frequently).

2.3. Data analysis

An experiment was conducted with two objectives, namely, maximizing the classification accuracy of SAs, and selecting the most discriminative set of items. Our analyzes followed three steps. Initially, we explored the classification accuracy that could be obtained by automatically selecting the best items of the previously described scales and socio-demographic factors. Then, we determined the items that best discriminated between SA and non-SA. Finally, the classification accuracy achieved using the selected items was compared with the obtained by each individual scale.

In order to reach our objectives, we used elastic net (Lars-en) (Zou and Hastie, 2005), which is a variant of stepwise linear regression that usually improves its performance. Also based on forward selection, this method modifies the optimization function by adding constraints to the L1 and L2 norms of the vector of coefficients. This technique minimizes

$$|y - X\beta|^2 + \lambda_1|\beta|_1 + \lambda_2|\beta|_2$$

where

$$|\beta|_1 = \sum_{j=1}^{p} |\beta_j|$$

and

$$|\beta|_2 = \sum_{j=1}^{p} \beta_j^2$$

The response is centered and the predictors are standardized before applying the technique. This rather simple modification has proven very powerful, and can be understood as a generalization of other two popular methods: least angle regression (Lars) (Efron et al., 2004) and least absolute shrinkage and selection operator (Lasso) (Tibshirani, 1996).

In order to conduct the experiment, the data set was randomly divided into three sets: training set, evaluation set and test set. Each set was composed of 146 non-SA and 82 SA. The training set was used to train the Lars-en algorithm. Because the classifier accuracy depends on the selected variables and the threshold, the evaluation set was used to determine the optimal configuration parameters. These parameters were first adjusted according to the values that maximized the accuracy in the evaluation set and then used for the test set. Average results after 100 repetitions of this process are reported below.

The scales included in this study were presented to all the subjects. However, some items were removed before starting the experiment. Regarding the IPDE-SQ, item 25 ("I have never threatened suicide or injured myself on purpose") reflects suicidal behavior and therefore was excluded from the analysis. Item 49 ("I often seek advice or reassurance about everyday decisions") was also removed due to the existence of non-responding subjects. There were no missing values in the remaining items. Regarding the SRRS scale, we found that certain life events were extremely rare in our sample. For instance, only 0.5% people presented with "jail term or probation", and none of them showed "change in religious activities", "minor financial loan", or "change in schools". Uncommon life events have limited clinical interest and furthermore, if included in the analysis, the covariance matrix would be singular and the estimation of parameters would have been problematic. Therefore life events with a frequency lower than 5% were removed from the study. The remaining set of variables for the analysis consisted of 154 variables (30 BIS-11 items, 75 IPDE-SQ items, 33 BGHA items, 14 SRRS items, age and sex).

3. Results

Classification results are displayed in Table 3. The items selected by the Lars-en algorithm attained an average accuracy of 85.3% in classifying SA.

For a better understanding of the predictive capacity of these scales, Fig. 1 shows the respective average receiver operating characteristic (ROC) curves together with the average area under the curves.

Table 3
Average classification accuracy, specificity, and sensitivity (±standard errors) of each assessment scale and using the most discriminative items from the four scales and socio-demographic factors.

Scales		Brown–Goodwin	BIS	SRRS	IPDE-SQ	Four scales and socio-demographic factors[a]
Evaluation	Accuracy	77.7 ± 0.02	81.0 ± 0.02	83.9 ± 0.02	85.0 ± 0.02	88.6 ± 0.01
	Specificity	92.4 ± 0.04	88.6 ± 0.03	88.8 ± 0.04	89.9 ± 0.03	92.7 ± 0.03
	Sensitivity	51.8 ± 0.08	67.6 ± 0.07	75.3 ± 0.08	76.3 ± 0.05	81.5 ± 0.05
Test	Accuracy	74.8 ± 0.02	78.4 ± 0.02	79.8 ± 0.06	81.6 ± 0.02	85.3 ± 0.02
	Specificity	89.5 ± 0.05	86.5 ± 0.04	84.3 ± 0.13	86.7 ± 0.04	89.8 ± 0.04
	Sensitivity	48.9 ± 0.08	64.2 ± 0.02	71.9 ± 0.10	72.6 ± 0.07	77.3 ± 0.07

[a] Note that the Lars-en algorithm can select a different set of items in each of the 100 repetitions. For instance, in one repetition it can select 35 items, 22 in another, and so on.

Once we observed that a suitable accuracy could be obtained, it was necessary to decide which items should be part of an accurate scale for measuring risk of suicidal behavior. In order to do this, each item was assigned to an index that indicates its relevance. The index is defined as the absolute average weight associated to the item in the projections obtained by the Lars-en algorithm in the 100 repetitions of the experiment. The scale was constructed by selecting the n most relevant items that maximized the average classification accuracy in the evaluation set. Fig. 2 displays the average accuracy obtained in the evaluation set when the n more relevant items were included. The maximum was attained using the 27 most relevant items.

Table 4 displays the selected items and their weights. The total score of a particular subject is obtained summing (or subtracting, if negative) each item value (0 or 1 in the SRRS and IPDE-SQ; 1 to 4 in the BIS-11; and 0 to 3 in the BGHA scale) multiplied by each item's weight. For instance, a 30 years-old subject who is in the process of marital separation (SRRS), but considers that he/she usually get fun and enjoyment out of life (IPDE-SQ) and always plan for job security (BIS-11) would score 30 (years) $\times$ 0.2 (weight) + 1 $\times$ 15 (weight) + 1 $\times$ (−8.1) (weight) + 3 $\times$ (−2.6) (weight) = 5.1 (see Table 4 for weight's information). Using this set of 27 items, we obtained an average accuracy of 86.4%, a specificity of 89.6%, and a sensitivity of 80.8% in classifying suicide attempters. These results are similar to those reported in Table 3, thus giving further support to the item selection by the Lars-en algorithm. Of note, the proposed scale did not include sex. Two facts might explain this result. First, sex was the 30th more relevant feature. As it can be appreciated in Fig. 2, the difference in the fitness function when selecting 27 or 30 items is minimal. Secondly, the effect of sex might be captured in the scale through other items. For instance, the selected item "revision of personal habits" which is closely associated with suicidal behavior in females.

Finally, we tested if the scores obtained by the SA and non-SA groups were significatively different. Then, we calculated the scores of the whole sample using the proposed scale. The Lilliefors goodness-of-fit test of composite normality showed that the projections of the SA and the non-SA (psychiatric inpatients and blood donors) groups followed Gaussian distributions. Mean (±SD) total scores of blood donors, psychiatric inpatients, and suicide attempters were 8.4 (±14.6), 37.1 (±21.9), 75.9 (±28.7), respectively. In order to provide a visualization of these groups, Fig. 3 adjusts a mixture of Gaussians to the histogram of the scores. Intuitively, it is appreciated that there are differences between the three groups. An ANOVA test with Bonferroni correction verified this assumption ($F = 717.23$; df = 2; $p < 0.001$).

4. Discussion

In this article, the Lars-en algorithm analyzes the accuracy that can be obtained in the classification of individuals as SA or non-SA, by selecting the most discriminating items of well-known psychiatric scales. Our results indicate that SA can be accurately classified using a set of 27 items. This set of items showed an average accuracy of 86.4%, a specificity of 89.6%, and a sensitivity of 80.8% in classifying SA.

According to their weights, the items most closely associated to SA status came from the IPDE-SQ (personality traits) and SRRS (life events). This is in accordance with the stress-diathesis model of

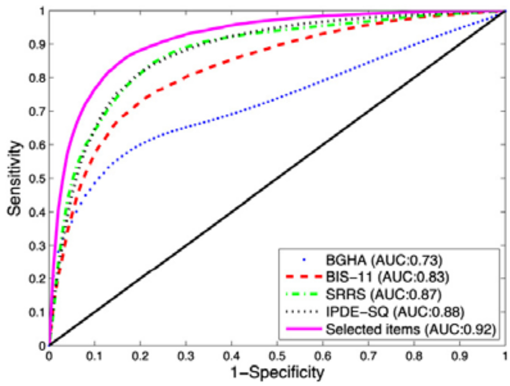


Fig. 1. Receiver operating characteristic (ROC) curves using the different scales.

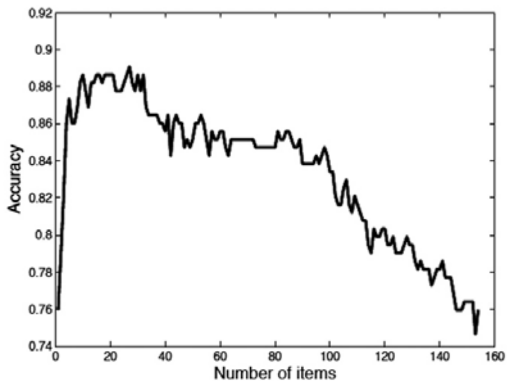


Fig. 2. Average accuracy on the evaluation set based on the number of relevant items included.

Table 4
Weight of the items selected by Lars-en.

Scale	Item	Weight[a]
SRRS	Change in frequency of arguments	29.4
SRRS	Revision of personal habits	22.7
SRRS	Marital separation	15.0
IPDE-SQ	I often feel empty inside	13.6
SRRS	Gain of new family member	10.7
BGHA	Adult self-harm	9.8
SRRS	Personal injury or illness	9.8
IPDE-SQ	I worry about being alone and having to care for myself	7.1
IPDE-SQ	I have tamtrums or angry outburst	6.2
IPDE-SQ	I've been the victim of unfair attacks on my character or reputation	5.9
IPDE-SQ	I can't decide what kind of persona I want to be	5.3
IPDE-SQ	I think my spouse (or lover) may be unfaithful to me	4.7
IPDE-SQ	I usually feel unconfortable or helpless when I'm alone	4.2
IPDE-SQ	I won't get involved with people until I'm certain they like me	3.9
IPDE-SQ	I have little or no desire to have sex with anyone	3.3
BIS-11	I spend or charge more than I earn	2.6
IPDE-SQ	People think I'm odd or eccentric	2.3
IPDE-SQ	I go to extremes to try to keep people from leaving me	1.5
IPDE-SQ	My feelings are like the weather, they are always changing	1.0
BIS-11	I act on impulse	0.4
Socio-demographic factors	Age	0.2
BIS-11	I am self-controlled	0.1
BIS-11	I plan trips well ahead of time	−2.4
BIS-11	I plan for job security	−2.6
IPDE-SQ	People have a high opinion on me	−3.4
IPDE-SQ	I usually get fun and enjoyment out of life	−8.1
BGHA	Armed aggression to others	−22.1

[a] The intercept in the regression was −120.82.

suicide (Mann et al., 1999b) and literature suggesting that suicide is related to personality and triggered by dramatic life events (Blasco-Fontecilla et al., 2010). Additionally, 26% (7/27) of the items included in the scale were related to the BIS-11 and BGHA scales in agreement with the expected association of impulsive aggression

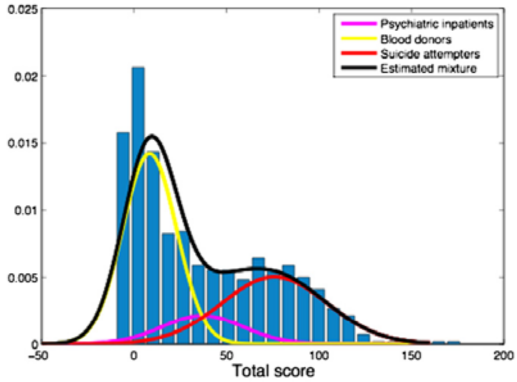


Fig. 3. Mixture of Gaussians of the three populations of the study.

traits with suicidal behavior (Giegling et al., 2009). If we analyze individually the items that received the highest weights, we can observe that "change in the number of arguments" and "marital separation" were two of the most relevant life events, which probably reflects the relevance of distressful life events in social and marital domains for SAs. All these findings, along with the fact that the most relevant personality items according to their relative weights indicated emptiness ("I often feel empty inside"), unhappiness ("I usually get fun and enjoyment out of life"), and dependency needs ("I worry about being alone and having to care for myself"), suggest that individuals unable to cope with problematic relationships with other people are at risk of attempting suicide.

Another interesting finding is that 18% (5/27) of the selected items had negative weights. Subjects endorsing items with negative weights such as "I usually get fun and enjoyment out of life" (IPDE-SQ), "I plan trips well ahead of time" (BIS-11), or "Armed aggression to others" (BGHA) were less likely to be a SA. This is in accordance with the scarce available literature. For instance, an inverse association between suicide rates and happiness or life satisfaction has been reported (Bray and Gunnell, 2006). A more recent study also reported an inverse association between well-being and suicide intent in a sample of 469 SA (Sisask et al., 2008). The negative weight of the items "I plan trips well ahead of time" and "I plan for job security" is coherent with the fact that most suicide attempts in our sample were impulsive (only 23.7% thought on suicide more than three hours before the attempt). "Armed aggression to others" was the factor with the largest negative weight, which might be compatible with the classical Freudian assumption that the externalization of aggression could protect from suicidal tendencies (Freud, 1947), although this is a controversial, poorly studied topic (Ferreira de Castro et al., 1986). Therefore, a sense of happiness, planning ahead, and externalization of aggression could be considered protective factors by clinicians evaluating suicidal risk.

Major strengths of the present study are the relatively large sample and the use of an efficient algorithm (Lars-en) that selects the most suitable items to assess suicidal risk. Our study presents some limitations and should just be considered a preliminary step in the development of a new scale for suicide risk assessment. The findings reported here require replication in other samples and different settings. The resulting set of items should be eventually validated and compared to specific scales that assess suicidal risk. Although our results suggest that suicide attempters can be accurately detected without information about Axis I disorders (which are consistently associated with suicidal behavior), including this information might increase the accuracy of our results.

5. Conclusion

The reduced number of items selected by the Lars-en algorithm in this study suggests that this set of items could be used as a quick, feasible but accurate instrument to assist clinicians in the evaluation of suicide risk. For instance, it might help primary care physicians in deciding which patients are at risk of suicide and should be referred to a psychiatrist or even hospitalized. It might also assist psychiatrists in evaluating short-term suicide risk in the emergency departments.

Contributors

Dr. Baca-Garcia designed the study and wrote the protocol. Dr. Blasco-Fontecilla and Dr. Lopez-Castroman managed the literature searches. Dr. Delgado-Gomez, Dr. Aguado and Dr. Ruiz-Hernandez undertook the statistical analysis. Dr. Blasco-Fontecilla and Dr.

Delgado-Gomez wrote the first draft of the manuscript. All authors contributed to and have approved the final manuscript.

Conflict of interest

None author report any conflict of interest.

Acknowledgments

This article was supported by the National Alliance for Research on Schizophrenia and Affective Disorders (NARSAD), Fondo de Investigacion Sanitaria (FIS) PI060092, Fondo de Investigacion Sanitaria FIS RD06/0011/0016, ETES (PI07/90207), the Conchita Rabago Foundation, and the Spanish Ministry of Health, Instituto de Salud Carlos III, CIBERSAM (Intramural 521 Project, P91B; SCO/3410/2004). Dr. Blasco-Fontecilla acknowledges the Spanish Ministry of Health (Rio Hortega CM08/00170), Alicia Koplowitz Foundation, and Conchita Rabago Foundation for funding his post-doctoral stage at CHRU, Montpellier, France. The work by Jorge Lopez-Castroman was supported by a FondaMental Foundation grant.

References

Baca-Garcia E, Diaz-Sastre C, García-Resa E, Blasco H, Braquehais D, Oquendo M, et al. Suicide attempts and impulsivity. European Archives of Psychiatry and Clinical Neurosciences 2005;255:152–6.
Baca-Garcia E, Oquendo MA, Saiz-Ruiz J, de Leon J, Mann JJ. Differences in aggression in New York and Madrid: impact on suicidal behavior. Journal of Clinical Psychiatry 2006;67:75–80.
Beck RW, Morris JB, Beck AT. Cross-validation of the Suicidal Intent Scale. Psychological Reports 1974;34.
Blasco-Fontecilla H, Baca-Garcia E, Duberstein P, Perez-Rodriguez M, Dervic K, Saiz-Ruiz J, et al. An exploratory study of the relationship between diverse life events and personality disorders in a sample of suicide attempters. Journal of Personality Disorders 2010;24:774–85.
Bray I, Gunnell D. Suicide rates, life satisfaction and happiness as markers for population mental health. Social Psychiatry and Psychiatric Epidemiology 2006;41:333–7.
Brown GK, Have TT, Henriques GR, Sie SX, Hollander JE, Beck AT. Cognitive therapy for the prevention of suicide attempts. Journal of the American Medical Association 2005;294.
Brown GL, Goodwin FK, Ballenger JC, Goyer PF, Major LF. Aggression in humans correlates with cerebrospinal fluid amine metabolites. Psychiatry Research 1979;1:131–9.
Canetto SS, Sakinofsky I. The gender paradox in suicide. Suicide and Life Threatening Behavior 1998;28:1–23.
Christiansen E, Jensen BF. Risk of repetition of suicide attempt, suicide or all deaths after an episode of attempted suicide: a register based survival analysis. The Australian and New Zealand Journal of Psychiatry 2007;41:257–65.
Coreil J, Bryant CA, Henderson JN. Social and behavioral foundations of public health. Sage Publications; 2001.
Delgado-Gomez D, Blasco-Fontecilla H, Alegria AA, Legido-Gil T, Artes-Rodriguez A, Baca-Garcia E. Improving the accuracy of suicide attempter classification. Artificial Intelligence in Medicine 2011;52:165–8.
Delgado-Gomez D, Blasco-Fontecilla H, Sukno F, Ramos-Plasencia S, Baca-Garcia E. Suicide attempters classification: toward predictive models of suicidal behavior. Neurocomputing 2012;92:3–8.
Efron B, Hastie T, Johnstone I, Tibshirani R. Least angle regression. The Annals of Statistics 2004;32:407–99.
Ferreira de Castro E, Albino L, Martins I. Relation between suicide and homicide in Portugal from 1970 to 1982. Acta Psychiatrica Scandinavica 1986;74:425–32.
Freud S. The ego and the id. London: The Hogarth Press; 1947.
García-Nieto R, Parra Uribe I, Palao D, Lopez-Castroman J, Sáiz PA, García-Portilla MP, et al. Brief suicide questionnaire. Inter-rater reliability. Revista de Psiquiatria y Salud Mental 2012;5:24–36.
Giegling I, Olgiati P, Hartmann AM, Calati R, Moller HJ, Rujescu D, et al. Personality and attempted suicide. Analysis of anger, aggression and impulsivity. Journal of Psychiatric Research 2009;43:1262–71.
Gonzalez de Rivera JL, Morera A. La valoración de sucesos vitales: adaptación española de la escala de Holmes y Rahe. Psiquis 1983;4:7–11.
Hampton T. Depression care effort brings dramatic drop in large HMO population's suicide rate. Journal of the American Medical Association 2010;303:1903–5.
Hendin H, Al Jurdi RK, Houck PR, Hughes S, Turner JB. Evidence for significant improvement in prediction of acute risk for suicidal behavior. Journal of Nervous and Mental Disease 2010;198:604–5.
Holmes EA, Crane C, Fennell MJ, Williams JM. Imagery about suicide in depression—"flash-forwards"? Journal of Behavior Therapy and Experimental Psychiatry 2007;38:423–34.
Holmes TH, Rahe RH. The social readjustment rating scale. Journal of Psychosomatic Research 1967;11:213–8.
Isaacson G. Suicide prevention-a medical breakthrough? Acta Psychiatrica Scandinavica 2000;102.
Jamison KR. Suicide and bipolar disorder. Journal of Clinical Psychiatry 2000;61:47–51.
Kolves K, Varnik A, Schneider B, Fritze J, Allik J. Recent life events and suicide: a case-control study in Tallinn and Frankfurt. Social Science & Medicine 2006;62:2887–96.
Loranger AW, Sartorius N, Andreoli A, Berger P, Buchheim P, Channabasavanna SM, et al. The international personality disorder examination. The World Health Organization/alcohol, drug abuse, and mental health administration international pilot study of personality disorders. Archives of General Psychiatry 1994;51:215–24.
Mann AH, Raven P, Pilgrim J, Khanna S, Velayudham A, Suresh KP, et al. An assessment of the standardized assessment of personality as a screening instrument for the international personality disorder examination: a comparison of informant and patient assessment for personality disorder. Psychological Medicine 1999a;29:985–9.
Mann JJ, Waternaux C, Haas GL, Malone KM. Toward a clinical model of suicidal behavior in psychiatric patients. American Journal of Psychiatry 1999b;156:181–9.
Oquendo M, Baca-Garcia E, Graver R, Morales M, Montalvan V, Mann JJ. Spanish adaptation of the Barratt Impulsiveness Scale (BIS-11). European Journal of Psychiatry 2001;15:147–55.
Patton JH, Stanford MS, Barratt ES. Factor structure of the barratt impulsiveness scale. Journal of Clinical Psychology 1995;51:768–74.
Perroud N, Baud P, Mouthon D, Courtet P, Malafosse A. Impulsivity, aggression and suicidal behavior in unipolar and bipolar disorders. Journal of Affective Disorders 2011;134:112–8.
Pokorny AD. Prediction of suicide in psychiatric patients. Report of a prospective study. Archives of General Psychiatry 1983;40:249–57.
Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. Journal of Clinical Psychiatry 1998;59(Suppl. 20):22–33.
Sisask M, Varnik A, Kolves K, Konstabel K, Wasserman D. Subjective psychological well-being (WHO-5) in assessment of the severity of suicide attempt. Nordic Journal of Psychiatry 2008;62:431–5.
Slama F, Courtet P, Golmard JL, Mathieu F, Guillaume S, Yon L, et al. Admixture analysis of age at first suicide attempt. Journal of Psychiatric Research 2009;43:895–900.
Smith JC, Mercy JA, Conn JM. Marital status and the risk of suicide. American Journal of Public Health 1988;78:78–80.
Stefansson J, Nordstrom P, Jokinen J. Suicide Intent Scale in the prediction of suicide. Journal of Affective Disorders 2010.
Tibshirani R. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society Series B 1996;58:267–88.
Weisman AD, Worden JW. Risk-rescue rating in suicide assessment. Archives of General Psychiatry 1972;26:553–60.
WHO. World report on violence and health. Geneva: World Health Organization; 2002.
Zou H, Hastie T. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society Series B 2005;67:301–20.

+Model
RPSM-277; No. of Pages 8
ARTICLE IN PRESS

Rev Psiquiatr Salud Ment (Barc.). 2015;xxx(xx):xxx–xxx

ELSEVIER

Revista de Psiquiatría
y Salud Mental

www.elsevier.es/saludmental

ORIGINAL ARTICLE

A Short Personality and Life Event (S-PLE) scale for detection of suicide attempters

Paula Artieda-Urrutia[a], David Delgado-Gómez[b], Diego Ruiz-Hernández[c], Juan Manuel García-Vega[a], Nuria Berenguer-Elias[a], Maria A. Oquendo[d], Hilario Blasco-Fontecilla[a,e,*]

[a] Department of Psychiatry, IDIPHIM, Puerta de Hierro University Hospital, Majadahonda, Spain
[b] Department of Statistics, Universidad Carlos III de Madrid, Spain
[c] Department of Quantitative Methods, CUNEF, Madrid, Spain
[d] Molecular Imaging and Neuropathology Division, New York State Psychiatric Institute, Columbia University, New York, USA
[e] Autonoma University, CIBERSAM, Madrid, Spain

Received 26 January 2015; accepted 20 February 2015

KEYWORDS
Suicide attempters;
Personality;
Life events;
Emptiness

Abstract
Objective: To develop a brief and reliable psychometric scale to identify individuals at risk for suicidal behaviour.
Method: Design: case–control study; sample and setting: 182 individuals (61 suicide attempters, 57 psychiatric controls, and 64 psychiatrically healthy controls) aged 18 or older, admitted to the emergency department at Puerta de Hierro University Hospital in Madrid, Spain; measures: all participants completed a form including their socio-demographic, clinical characteristics, and the Personality and Life Events (PLE) scale (27 items). To assess Axis I diagnoses, all psychiatric patients (including suicide attempters) were administered the Mini International Neuropsychiatric Interview (MINI); statistical analyses: descriptive statistics were computed for the socio-demographic factors. Additionally, chi-square independence tests were applied to evaluate differences in socio-demographic and clinical variables, and the PLE scale between groups. A stepwise linear regression with backward variable selection was conducted to build the Short Personality Life Event (S-PLE) scale. In order to evaluate the accuracy, a ROC analysis was conducted. The internal reliability was assessed using Cronbach's α, and the external reliability was evaluated using a test–retest procedure.
Results: The S-PLE scale, composed of just six items, showed good performance in discriminating between medical controls, psychiatric controls and suicide attempters in an independent sample. For instance, the S-PLE scale discriminated between past suicide and past non-suicide attempters with sensitivity of 80% and specificity of 75%. The area under the ROC curve was

* Corresponding author.
E-mail addresses: hmblasco@yahoo.es, hblasco@idiphim.org (H. Blasco-Fontecilla).

http://dx.doi.org/10.1016/j.rpsm.2015.02.004

Please cite this article in press as: Artieda-Urrutia P, et al. A Short Personality and Life Event (S-PLE) scale for detection of suicide attempters. Rev Psiquiatr Salud Ment (Barc.). 2015. http://dx.doi.org/10.1016/j.rpsm.2015.02.004

88%. A factor analysis extracted only one factor, revealing a single dimension of the S-PLE scale. Furthermore, the S-PLE scale provides values of internal and external reliability between poor (test–retest: 0.55) and acceptable (Cronbach's α: 0.65) ranges.

Administration time is about 1 min.

Conclusions: The S-PLE scale is a useful and accurate instrument for estimating the risk of suicidal behaviour in settings where the time is scarce.

PALABRAS CLAVE
Personas que intentan suicidarse;
Personalidad;
Acontecimientos vitales;
Sensación de vacío

Escala abreviada de personalidad y acontecimientos vitales (S-PLE) para la detección de los intentos de suicidio

Resumen

Objetivo: Desarrollar una escala breve y fiable para identificar a las personas en riesgo de conducta suicida.

Método: Diseño: estudio de casos y controles; muestra y centro: 182 individuos (61 personas que intentaron suicidarse, 57 controles psiquiátricos, y 64 controles sanos mentalmente) de 18 años o más, valorados en la unidad de urgencias del Hospital Universitario Puerta de Hierro de Madrid, España; mediciones: todos los participantes rellenaron un formulario que incluía sus características socio-demográficas y clínicas, y la escala de Personalidad y Acontecimientos Vitales (PLE) (27 cuestiones). Para evaluar los diagnósticos del Eje I, a todos los pacientes psiquiátricos (incluyendo a las personas que intentaron suicidarse) se les realizó la Entrevista Neuropsiquiátrica Internacional (MINI); análisis estadísticos: se usaron estadísticos descriptivos para los factores socio-demográficos. Además, se usó la χ^2 para evaluar las diferencias de las variables socio-demográficas y clínicas, y de la escala PLE entre grupos. Para construir la escala abreviada de personalidad y acontecimientos vitales (S-PLE) se llevó a cabo una regresión lineal por pasos con selección de variables retrospectivas. A fin de evaluar la precisión se realizó un análisis de ROC. Se evaluó la fiabilidad interna utilizando la α de Cronbach, y la fiabilidad externa utilizando un test-retest.

Resultados: La escala S-PLE, que se compone únicamente de seis cuestiones, permitió diferenciar adecuadamente entre controles sanos, controles psiquiátricos y los pacientes con intentos de suicidio, en una muestra independiente. Por ejemplo, la escala S-PLE permitió diferenciar a las personas que intentaron suicidarse y a las que no tenían antecedentes de conducta suicida, con una sensibilidad del 80% y una especificidad del 75%. El área bajo la curva ROC fue del 88%. Un análisis factorial extrajo solamente un factor, lo que revela la dimensión única de la escala S-PLE. Además, la escala S-PLE tiene valores de fiabilidad interna y externa que se incluyen dentro de los rangos débil (test-retest: 0,55) y aceptable (α de Cronbach: 0,65).

El tiempo de aplicación de la PLE es de alrededor de un minuto.

Conclusiones: La escala S-PLE es un instrumento útil y preciso para calcular el riesgo de conducta suicida en dispositivos donde escasea el tiempo.

Introduction

Suicide prevention is a major public health concern.[1-3] According to the World Health Organization (WHO), suicide accounts approximately for 1.4% of the Global Burden of Disease.[4] It is estimated that about 800,000 people will suicide each year, and at least 10–20 times more will make non-lethal attempts annually by the year 2020.[5] Moreover, in the United States only, the annual estimated economic costs of suicide are about $33 billion per year.[6]

Despite these staggering figures, it is possible to partially prevent suicidal behaviour with specific interventions in high-risk populations.[7,8] For instance, some investigators have pointed out that access to adequate treatment might reduce suicide rates up to 25%.[9] More recently, Hampton[10] reported a 75% reduction of suicide rates by implementing a depression care programme.

One of the most difficult tasks in preventing suicide is the detection of individuals at risk.[11] Among the reasons behind this problem is the fact that most suicide attempters do not reveal their suicidal thoughts and plans to their physicians.[12,13] Moreover, although several risk factors such as major depression,[14] a previous suicide attempt,[15] or recent life events[16] increase the risk of suicide, previous efforts to find a tool that accurately identifies people at risk of suicidal behaviour have yet to yield good results.[17,18] For instance, Bolton et al.[19] conducted a study where the SAD PERSONS scale could not accurately predict suicide attempts among individuals requiring psychiatric services in emergency departments. In another study, the Suicide

Please cite this article in press as: Artieda-Urrutia P, et al. A Short Personality and Life Event (S-PLE) scale for detection of suicide attempters. Rev Psiquiatr Salud Ment (Barc.). 2015. http://dx.doi.org/10.1016/j.rpsm.2015.02.004

Intent Scale (SIS) was not able to predict which self-harming individuals would ultimately die by suicide.[20]

These findings highlight the importance of developing accurate tools to identify and prevent suicidal behaviour. With this aim, Blasco-Fontecilla et al.[21] developed the Personality and Life Event (PLE) scale. The PLE scale consists of 27 of the most discriminative items from a collection of questionnaires usually employed in the assessment of suicidal behaviour. The 27 items were selected using the Lars-en algorithm.[22] The PLE scale showed excellent accuracy (86.4%), sensitivity (80.8%), and specificity (89.6%) in classifying suicide attempters.[21]

The aim of this article is to develop and test a shorter version of the PLE scale that can be used in setups where time is scarce such as emergency departments and overloaded outpatient clinics. Moreover, the development of this short scale is also relevant because of the increase of online e-health applications: nowadays, both researchers and practitioners collect patient's data by means of mobile phone applications.[23] The availability of short scales is vital considering that attrition is one of the main problems faced when collecting this type of data.[24] Our hypothesis is that a shorter version of the PLE scale can accurately identify individuals at risk for suicidal behaviour.

Material and methods

Samples and procedure

Participants received no incentives. Initially, we recruited 236 individuals, but 54 (22.5%) were excluded from our analyses because they had missing values. The final sample in this study consisted of 182 individuals (61 suicide attempters, 57 psychiatric controls, and 64 psychiatrically healthy controls) aged 18 or older, admitted to the emergency department at Puerta de Hierro University Hospital in Madrid, Spain, between 1st of June and 1st of December, 2013. All participants were evaluated within the first 24 h after admission. The assessments were made by residents in psychiatry with specific training to evaluate suicide attempts.

The suicide attempter group comprised 18 men and 43 women. The psychiatric controls were 21 men and 36 women presenting to the emergency department for any psychiatric reason other than suicidal behaviour. These individuals denied a history of previous suicidal attempts. The psychiatrically healthy controls (henceforth referred to simply as medical controls) were 26 men and 38 women, presenting to the emergency department because of a medical (non-psychiatric) reason. Medical controls were free of psychiatric conditions. In all three groups, individuals who were unable to decide or understand the PLE scale for any reason were excluded. All included patients provided written informed consent to take part in the study. The Puerta de Hierro University Hospital Ethics Committee approved the study.

Measures

All participants completed a form including their sociodemographic and clinical characteristics. In order to facilitate the use of the PLE scale, all the items were dichotomized (yes/no). To assess Axis I diagnoses, all psychiatric patients (including suicide attempters) were administered the Mini International Neuropsychiatric Interview (MINI).[25] The MINI is an efficient, short, and easy to administrate structured diagnostic interview.[21,25] Medical controls were determined not to have a personal history of psychiatric illnesses, as verified by direct questioning and review of the electronic medical record.

Statistical analysis

Descriptive statistics were computed for the sociodemographic factors. Additionally, Chi-square independence tests were applied to evaluate differences in sociodemographic and clinical variables between groups. The Chi-square independence test was also conducted to assess the strength of the relationship between each item in the PLE scale and the type of individual (suicide attempter, psychiatric control or medical control).

A stepwise linear regression with backward variable selection was conducted to build the short version of the PLE scale. In order to evaluate the accuracy of the developed scale, a ROC analysis was conducted. Typically, works published in the areas of psychology and psychiatry conduct this type of analysis using the same data set for both, estimation of the parameters of the classifiers and accuracy assessment. The consequence of this practice is that the reported results tend to be better than what is observed in practice. For example, Support Vector Machines are capable of attaining a perfect classification in a given set. However, whenever this classifier is applied to a different data set, its performance decreases dramatically. This phenomenon is known as overfitting. In order to avoid this problem, the current ROC analysis is performed according to the usual practice in the pattern recognition community, i.e., the data are separated in two disjoint sets, namely, training and test sets. The training set is used to estimate the parameters of the linear regression, and the test set is used to calculate the ROC curves. With the aim of obtaining more relevant results, 100 cross-validations were conducted. In each cross-validation, the training set was composed by 32 medical controls, 29 psychiatric controls, and 31 suicide attempters randomly selected from our database. The remaining individuals (32 medical controls, 28 psychiatric controls, and 30 suicide attempters) were used as the test set.

Finally, after the performance of the proposed scale was assessed, a factor analysis with varimax rotation was performed, and the internal and external reliability were computed in order to elucidate the psychometric properties of the proposed scale. The internal reliability was assessed using Cronbach's α, and the external reliability was evaluated using a test–retest procedure, and the intraclass correlation coefficient. The interval for the test–retests ranged between three and six months. Basically, between 3 and 6 months after the initial interview, all participants were contacted by telephone, and asked to answer the PLE scale again. Due to refusal to participate in the retest, impossibility to contact the individual, or scarce time since the first interview (less than three months scheduled), we used a subsample of 85 individuals (47%) to make the test–retest analyses.

Please cite this article in press as: Artieda-Urrutia P, et al. A Short Personality and Life Event (S-PLE) scale for detection of suicide attempters. Rev Psiquiatr Salud Ment (Barc.). 2015. http://dx.doi.org/10.1016/j.rpsm.2015.02.004

Table 1 Comparison of suicide attempters ($n=61$), psychiatric controls ($n=57$), and medical controls ($n=64$) on socio-demographic variables.

Variable	Suicide attempter ($n=61$) N (%)	Psychiatric control ($n=57$) N (%)	Medical control ($n=64$) N (%)	Stats*	Df	p-value
Age (years)	36.9 (SD:11.83)	44.1 (SD:13.48)	41.1 (SD:15.19)	4.18	2;179	0.02
Sex (female)	43 (70.5%)	36 (63.2%)	38 (59.4%)	1.727	2	0.42
Years of education				8.472	4	0.08
<8	4 (6.6%)	10 (19.6%)	5 (7.8%)			
9–12	37 (60.6%)	22 (43.1%)	30 (46.9%)			
>12	20 (32.8%)	19 (37.3%)	29 (45.3%)			
Socio-economic status				4.986	4	0.29
Low (<500) and Low-Middle (500–1500)	33 (62.3%)	26 (59.1%)	28 (47.5%)			
Middle (1500–2000)	11 (20.8%)	6 (13.6%)	11 (18.6%)			
Middle-High (>2000)	9 (16.9%)	12 (27.3%)	20 (33.9%)			
Living arrangements				10.82	4	0.03
Family of origin	22 (36.1%)	10 (17.5%)	9 (14.1%)			
Own family	31 (50.8%)	37 (64.9%)	47 (73.4%)			
Alone and others	8 (13.1%)	10 (17.6%)	8 (12.5%)			
Race				4.210	4	0.38
Caucasian	53 (86.9%)	53 (93.0%)	55 (85.9%)			
Hispanic	7 (11.5%)	3 (5.3%)	5 (7.8%)			
Other race	1 (1.6%)	1 (1.8%)	4 (6.2%)			

* χ^2 tests were applied for all comparisons except age (ANOVA).

Results

Sample characteristics

Table 1 shows the socio-demographic data. Apart from age and living arrangements, there no statistically significant differences between groups in terms of socio-demographics.

Table 2 shows the comparison between the scores obtained by the suicidal attempters and the psychiatric controls in the MINI scale. There were no statistically significant differences in Axis I disorders when comparing suicide attempters and psychiatric controls.

Chi-square independence tests

A chi-square test assessed the relationship between each item and group membership (see Table 3).

The items ''Gain of a New Family Member'', ''Personal Injury or Illness'', ''I Plan for Job Security'' or ''Armed Aggression to Others'' were not related to suicide attempter status. Similarly, ''I Have Little or No Desire to have sex with Anyone'' and ''I Spend or Charge More that I Earn'' showed a weak relation to suicide attempter status. These results support the idea of shortening the PLE scale.

Table 2 Comparison of suicide attempters and psychiatric controls on psychiatric disorders (MINI scale).

MINI	Suicide attempters* $n=61$(%)	Psychiatric controls* $n=57$ (%)	χ^2 test (p-value)
Major depressive episode (current)	45/60 (75.0%)	39/57 (68.4%)	0.42
Major depressive episode (past)	21/58 (36.2%)	19/57 (33.3%)	0.75
Dysthymia	7/60 (11.7%)	3/57 (5.3%)	0.22
Panic disorder (current)	6/59 (10.2%)	5/57 (8.8%)	0.79
Panic disorder with agoraphobia	1/59 (1.7%)	4/57 (7.0%)	0.16
Alcohol dependence	12/58 (20.7%)	8/58 (13.8%)	0.34
Drug dependence	4/59 (6.8%)	3/57 (5.3%)	0.73
Generalized anxiety disorder	24/58 (41.4%)	22/57 (38.6%)	0.76
Psychotic disorder (Current)	2/59 (3.4%)	2/57 (3.5%)	0.97

* Cells in the first two columns show the ratio of cases with the attribute with respect to the available data. Missing data are no-answers to each particular item.

Please cite this article in press as: Artieda-Urrutia P, et al. A Short Personality and Life Event (S-PLE) scale for detection of suicide attempters. Rev Psiquiatr Salud Ment (Barc.). 2015. http://dx.doi.org/10.1016/j.rpsm.2015.02.004

Table 3 Comparison of suicide attempters ($n = 61$), psychiatric controls ($n = 57$), and medical controls ($n = 64$) on the items of the PLE scale.

	PLE item	Suicide attempters $n = 61$ (%)	Psychiatric control $n = 57$ (%)	Medical control $n = 64$ (%)	χ^2 test (p-value)
1	Change in frequency of arguments	41 (67.2)	31 (54.4)	11 (17.2)	0.00
2	Change in personal habits	39 (63.9)	30 (52.6)	17 (26.6)	0.00
3	Marital separation	18 (29.5)	7 (12.3)	0 (0.0)	0.00
4	I often feel empty inside	55 (90.2)	37 (64.9)	5 (7.8)	0.00
5	Gain of new family member	6 (9.8)	7 (12.3)	10 (15.6)	0.61*
6	Adult self-harm	21 (34.4)	3 (5.3)	0 (0.0)	0.00
7	Personal injury or illness	5 (8.2)	4 (7)	6 (9.4)	0.89*
8	I worry about being alone and having to care for myself	34 (55.7)	38 (66.7)	12 (18.8)	0.00
9	I have tantrums or angry outburst	44 (72.1)	33 (57.9)	12 (18.8)	0.00
10	I've been the victim of unfair attacks on my character or reputation	42 (68.9)	35 (61.4)	5 (7.8)	0.00
11	I can't decide what kind of person I want to be	33 (54.1)	22 (38.6)	4 (6.3)	0.00
12	I think my spouse (or lover) may be unfaithful to me	21 (34.4)	6 (10.5)	4 (6.3)	0.00
13	I usually feel uncomfortable or helpless when I'm alone	23 (37.7)	28 (49.1)	4 (6.3)	0.00
14	I won't get involved with people until I'm certain they like me	27 (44.3)	27 (47.4)	7 (10.9)	0.00
15	I have little or no desire to have sex with anyone	25 (41.0)	23 (40.4)	13 (20.3)	0.02
16	I spend or charge more than I earn	20 (32.8)	14 (24.6)	9 (14.1)	0.04
17	People think I'm odd or eccentric	27 (44.3)	19 (33.3)	8 (12.5)	0.00
18	I go to extremes to try to keep people from leaving me	28 (45.9)	12 (21.1)	1 (1.6)	0.00
19	My feelings are like the weather, they are always changing	32 (52.5)	31 (54.4)	3 (4.7)	0.00
20	I act on impulse	47 (77)	31 (54.4)	13 (20.3)	0.00
22	I am self-controlled	28 (45.9)	32 (56.1)	59 (92.2)	0.00
23	I plan trips well ahead of time	34 (55.7)	35 (61.4)	52 (81.3)	0.01
24	I plan for job security	45 (73.8)	44 (77.2)	55 (85.9)	0.22*
25	People have a high opinion on me	39 (63.9)	42 (73.7)	60 (93.8)	0.00
26	I usually get fun and enjoyment out of life	29 (47.5)	33 (57.9)	60 (93.8)	0.00
27	Armed aggression to others	0 (0.0)	0 (0.0)	0 (0.0)	1.00*

* Statistically non-significant relationships are starred.

Stepwise backward item selection

A stepwise backward regression analysis was conducted to select the most suitable items to be included in the S-PLE scale. The result of this analysis was the selection of the following six items: ''I often feel empty inside'', ''Adult self harm'', ''I have tantrums or angry outbursts'', ''I have been the victim of unfair attacks on my character or reputation'', ''I can't decide what kind of person I want to be'', and ''I think my spouse (or lover) may be unfaithful to me''.

Receiver-operating characteristic (ROC) analysis

Once the relevant items were identified, a ROC analysis evaluated its capacity to correctly classify the individuals. The corresponding results are shown in Fig. 1. This figure shows the average ROCs for the 100 simulations mentioned above, together with their average area under the curve (AUC).

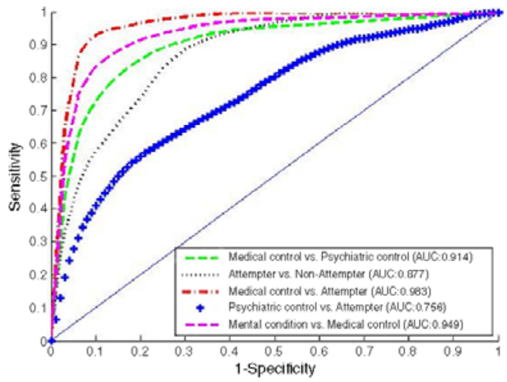


Figure 1 ROC curves of the S-PLE scale to discriminate between different pairwise comparisons.

Please cite this article in press as: Artieda-Urrutia P, et al. A Short Personality and Life Event (S-PLE) scale for detection of suicide attempters. Rev Psiquiatr Salud Ment (Barc.). 2015. http://dx.doi.org/10.1016/j.rpsm.2015.02.004

Table 4 Confusion matrix.[a]

	Real		
	Medical control ($n = 32$)	Psychiatric control ($n = 28$)	Suicide attempter ($n = 30$)
Predicted			
Medical control	27.79 (2.04)	4.97 (2.89)	1.35 (0.92)
Psychiatric control	4.19 (2.05)	13.64 (3.69)	8.17 (3.07)
Suicide attempter	0.02 (0.20)	9.39 (2.55)	20.48 (2.98)

Mean and standard deviation of the number of cases in each cell over 100 simulations.
[a] The performance of a given classification model can be evaluated by a confusion matrix. The confusion matrix displays the accuracy of the predictions made by the model. Thus, the diagonal displays the number of correct classifications made for each class (i.e. suicide attempters, psychiatric controls, or medical controls), and the off diagonal shows the errors made. For instance, of the 32 test medical controls, nearly 28 (27.79) were correctly classified as medical controls, and about four (4.19) and 0 (0.02) individuals were incorrectly classified as psychiatric controls and suicide attempters, respectively.

The proposed scale accurately discriminated medical controls from psychiatric controls (dashed line), achieving a sensitivity of nearly 90% with a specificity of 75%. Thus, the S-PLE scale could be used as a good instrument for screening mental diseases in primary care settings.

Depending on the type of Service (emergency department, walk-in clinics, and so on) the predictive parameters (i.e. sensitivity, specificity) could be assigned different values. For instance, a medical centre that fixes thresholds of 90% sensitivity in discriminating between medical and psychiatric controls, and of 70% specificity in discriminating between psychiatric controls and suicide attempters, will obtain the Confusion matrix shown in Table 4.

Psychometrical properties

A factor analysis with varimax rotation, applied on the selected items, extracted only one factor. This reflects the one-dimensionality of the S-PLE scale. The extracted factor (with eigenvalue higher than one) explained 36.9% of the variance. The first columns of Table 5 show the factor loadings after rotation. The contribution of the selected items to this factor is very similar.

Table 5 Factor analysis and weights of the S-PLE scale.

	Item	Factor loadings (after rotation)	Weights
1	I often feel empty inside	0.64	0.66
2	Adult self harm	0.70	0.46
3	I have been the victim of unfair attacks on my character or reputation	0.60	0.40
4	I think my spouse (or lover) may be unfaithful to me	0.87	0.36
5	I can't decide what kind of person I want to be	0.67	0.24
6	I have tantrums or angry outbursts	0.57	0.19
	Constant		1.15

Regarding internal consistency, the Cronbach's α was 0.65, whereas the test-retest reliability was 0.55 (calculated based on data from 85 individuals who participated in the retest), and the intraclass correlation coefficient was 0.53.

Proposed scale

Finally, the weights of the proposed scale were calculated using the full data set. These weights are shown in the last column of Table 5. The total score on the S-PLE scale for a given individual is obtained by summing the weights of the items that he/she answered affirmatively plus the constant. For instance, an individual who feels ''often empty inside'', and has a history of ''adult self-harm'' would score 2.27 [0.66 (emptiness) + 0.46 (adult self-harm) + 1.15 (constant)] (see Table 5 for weights).

Two thresholds are suggested to differentiate between the three types of individuals. Individuals with a score lower than 1.70 are considered mentally healthy; individuals with a score above 2.46 are considered at risk for attempting suicide; finally, individuals with scores between these two values are classified as individuals with a possible mental disorder. The values assigned to these thresholds are those that maximize the classification accuracy.

Discussion

We report on the development and testing of a Short Personality Event (S-PLE) scale. The S-PLE scale was demonstrated to be a reliable instrument for classifying individuals as past suicide attempters or not. Our results have also shown that the S-PLE scale was able to distinguish between individuals with mental conditions – either psychiatric controls or suicide attempters – and psychiatrically healthy controls. The performance of the S-PLE scale decreases when it comes to separating psychiatric controls from suicide attempters. This is not surprising given that psychiatric diagnoses are *per se* risk factors for suicidal behaviour. In any case, the AUC of the ROC curve comparing psychiatric controls and suicide attempters was still fairly acceptable (0.756). The study was completed with an assessment of the internal and external reliability. In this assessment the Cronbach's α took a value of 0.65, a value which is close to the benchmark suggested by Nunnaly in 1978.[26] Regarding the external reliability, the test–retest correlation was 0.55. This low correlation was expected, as some of the items included in the S-PLE scale (e.g. ''I have tantrums or angry outbursts'') measure the individual's state, not an individual's trait. Moreover, the fact that only six items compose this scale makes it particularly useful in settings where time is scarce.

Another major strength of the S-PLE scale is that it may indirectly evaluate the risk of suicidal behaviour. The items composing the S-PLE scale assess the way the examinee feels with respect to both himself and his social, work and emotional environment. This is important because most suicide attempters and completers do not display suicidal ideation in the previous appointments with their physicians,[12] thus hampering the prevention of suicidal behaviour. Fear of stigma, hospitalization or truncation of their plans could explain that lack of communication.[12,13] In this context, the S-PLE might be particularly interesting, as with the exception of adult self-harm, the S-PLE has no direct questions about suicide attempt. Accordingly, the S-PLE could be particularly useful for patients who might want to hide this information. In other words, this tool allows the clinician to identify those at risk circumventing reliance on patient reports of suicidal ideation.

Six items related to personality dysfunction, and a history of previous self-harm compose the S-PLE scale. The most discriminative items were ''I often feel empty inside'' and ''Adult self-harm''. These results are in keeping with the literature. For instance, in a recent review, we found that emptiness might have more predictive power than impulsiveness in the prediction of suicidal behaviour, and could be one of the clinical factors most closely associated with suicide attempt repetition,[27–29] and the addiction to suicidal behaviour.[29,30] Furthermore, a history of self-harm and affective instability are well-recognized risk factors for suicidal behaviour.[31,32]

A limitation of the present study is that the resulting set of 6 items has yet to be compared to specific scales that assess risk of suicidal behaviour. As well, the S-PLE has not yet been tested prospectively. Nonetheless, the S-PLE ability to identify those likely to have a psychiatric condition and those likely to have a past suicide attempt in about one minute has utility.

Conclusions

The S-PLE scale has shown acceptable psychometric properties, and is a valuable screening instrument in the assessment of risk of suicidal behaviour. The S-PLE scale brings together two important characteristics: it indirectly measures suicide risk, and can be easily answered within a very short time period. Thus, the S-PLE scale might be an interesting tool in primary or emergency care settings.

Funding

This article received support from the CIBERSAM (Banco de Instrumentos del CIBERSAM; http://www.cibersam.es/cibersam) to develop a scale capable of predicting suicidal behaviours (the Personality and Life Events Scale, PLE). The S-PLE will be included into the ''Banco de Instrumentos'' of the CIBERSAM. Paula Artieda-Urrutia has obtained competitive funding from the IDIPHIM (http://www.investigacionpuertadehierro.com/).

Conflict of interest

In the last three years, Dr. Hilario Blasco-Fontecilla has received lecture fees from Eli Lilly, AB-Biotics, Janssen, Rovi, and Shire. Maria A. Oquendo received royalties for the commercial use of the C-SSRS and her family owns stock in Bristol Myers Squibb. The remaining authors report no conflict of interest.

Acknowledgements

Monica Fernández-Rodriguez, Isabel González-Villalobos, Cesar Rodriguez-Lomas, Maria Martin-Garcia, and Rocio Blanco-Fernández helped us in recruiting some individuals for this study. Dr. Iruela-Cuadrado allowed us to carry out this study at Puerta de Hierro University Hospital.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.rpsm.2015.02.004.

References

1. Nock MK, Borges G, Bromet EJ, Alonso J, Angermeyer M, Beautrais A, et al. Cross-national prevalence and risk factors for suicidal ideation, plans and attempts. Br J Psychiatry. 2008;192:98–105.
2. Saiz PA, Bobes J. Suicide prevention in Spain: an uncovered clinical need. Rev Psiquiatr Salud Ment. 2014;7:1–4.
3. Giner L, Guija JA. Number of sucides in Spain: differences between data from the Spanish Statistical Office and the Institutes of Legal Medicine. Rev Psiquiatr Salud Ment. 2014;7:139–46.
4. WHO. World report on violence and health. Geneva, Switzerland: World Health Organization; 2002.
5. WHO. Figures and facts about suicide. Geneva, Switzerland: World Health Organization; 1999.
6. Coreil J, Bryant CA, Henderson JN. Social and behavioral foundations of public health. Sage Publications; 2001.
7. van der Feltz-Cornelis CM, Sarchiapone M, Postuvan V, Volker D, Roskar S, Grum AT, et al. Best practice elements of multilevel suicide prevention strategies: a review of systematic reviews. Crisis. 2011;32:319–33.
8. Jamison KR. Suicide and bipolar disorder. J Clin Psychiatry. 2000;61:47–51.
9. Isacsson G. Suicide prevention – a medical breakthrough? Acta Psychiatr Scand. 2000;102:113–7.
10. Hampton T. Depression care effort brings dramatic drop in large HMO population's suicide rate. JAMA. 2010;303:1903–5.
11. Davis AT, Schrueder C. The prediction of suicide. Med J Aust. 1990;153:552–4.
12. Smith EG, Kim HM, Ganoczy D, Stano C, Pfeiffer PN, Valenstein M. Suicide risk assessment received prior to suicide death by Veterans Health Administration patients with a history of depression. J Clin Psychiatry. 2013;74:226–32.

Please cite this article in press as: Artieda-Urrutia P, et al. A Short Personality and Life Event (S-PLE) scale for detection of suicide attempters. Rev Psiquiatr Salud Ment (Barc.). 2015. http://dx.doi.org/10.1016/j.rpsm.2015.02.004

13. Isometsa ET, Heikkinen ME, Marttunen MJ, Henriksson MM, Aro HM, Lonnqvist JK. The last appointment before suicide: is suicide intent communicated? Am J Psychiatry. 1995;152:919–22.
14. Mann JJ, Waternaux C, Haas GL, Malone KM. Toward a clinical model of suicidal behavior in psychiatric patients. Am J Psychiatry. 1999;156:181–9.
15. Oquendo MA, Currier D, Mann JJ. Prospective studies of suicidal behavior in major depressive and bipolar disorders: what is the evidence for predictive risk factors? Acta Psychiatr Scand. 2006;114:151–8.
16. Kolves K, Varnik A, Schneider B, Fritze J, Allik J. Recent life events and suicide: a case–control study in Tallinn and Frankfurt. Soc Sci Med. 2006;62:2887–96.
17. Hendin H, Al Jurdi RK, Houck PR, Hughes S, Turner JB. Evidence for significant improvement in prediction of acute risk for suicidal behavior. J Nerv Ment Dis. 2010;198:604–5.
18. Pokorny A. Prediction of suicide in psychiatric patients: report of a prospective study. Arch Gen Psychiatry. 1983;40:249–57.
19. Bolton JM, Spiwak R, Sareen J. Predicting suicide attempts with the SAD PERSONS scale: a longitudinal analysis. J Clin Psychiatry. 2012;73:e735–41.
20. Harriss L, Hawton K. Suicidal intent in deliberate self-harm and the risk of suicide: the predictive power of the Suicide Intent Scale. J Affect Disord. 2005;86:225–33.
21. Blasco-Fontecilla H, Delgado-Gomez D, Ruiz-Hernandez D, Aguado D, Baca-Garcia E, Lopez-Castroman J. Combining scales to assess suicide risk. J Psychiatr Res. 2012;46:1272–7.
22. Delgado-Gomez D, Blasco-Fontecilla H, Sukno F, Ramos-Plasencia MS, Baca-Garcia E. Suicide attempters classification: toward predictive models of suicidal behavior. Neurocomputing. 2012;92:3–8.
23. Myin-Germeys I, Oorschot M, Collip D, Lataster J, Delespaul P, van Os J. Experience sampling research in psychopathology: opening the black box of daily life. Psychol Med. 2009;39:1533–47.
24. Eysenbach G. The law of attrition. J Med Internet Res. 2005;7:e11.
25. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry. 1998;59 Suppl. 20:22–33.
26. Nunnally JC. Psychometric theory. New York: McGraw-Hill; 1978.
27. Blasco-Fontecilla H, de Leon-Martinez V, Delgado-Gomez D, Giner L, Guillaume S, Courtet P. Emptiness and suicidal behavior: an exploratory review. Suicidol Online. 2013;4:21–32.
28. Blasco-Fontecilla H, Jaussent I, Olie E, et al. A cross-sectional study of major repeaters: a distinct phenotype of suicidal behavior. Prim Care Companion CNS Disord. 2014;16.
29. Blasco-Fontecilla H, Baca-Garcia E, Courtet P, Garcia-Nieto R, de Leon J. Horror vacui: emptiness may be a core pathway in major suicide repeaters. A pilot study. Psychother Psychosom. 2015;84:117–9.
30. Blasco-Fontecilla H, Artieda-Urrutia P, Berenguer-Elias N, Garcia-Vega JM, Fernandez-Rodriguez M, Rodriguez-Lomas C, et al. Are major repeater patients addicted to suicidal behavior? Adicciones. 2014;26:321–33.
31. Palmier-Claus JE, Taylor PJ, Varese F, Pratt D. Does unstable mood increase risk of suicide? Theory, research and practice. J Affect Disord. 2012;143:5–15.
32. McGirr A, Alda M, Seguin M, Cabot S, Lesage A, Turecki G. Familial aggregation of suicide explained by cluster B traits: a three-group family study of suicide controlling for major depressive disorder. Am J Psychiatry. 2009;166:1124–34.

Please cite this article in press as: Artieda-Urrutia P, et al. A Short Personality and Life Event (S-PLE) scale for detection of suicide attempters. Rev Psiquiatr Salud Ment (Barc.). 2015. http://dx.doi.org/10.1016/j.rpsm.2015.02.004

2 Mejora de la predicción del desempeño futuro de los candidatos a partir de los datos de los procesos de selección de personal

En esta sección de la tesis, presentamos los resultados obtenidos en un estudio similar a los realizados anteriormente pero en un área diferente al de la salud. En este caso, el objetivo es determinar cuáles son las características que identifican a los buenos vendedores en un proceso de selección. Este es un problema de interés ya que, por ejemplo, Kotler ha indicado que más de la mitad de las ventas de una compañía son conseguidas únicamente por el 27% de los vendedores disponibles (Kotler, 1994). Greenberg y Greenberg aportan datos más extremos: el 20% de los vendedores de la empresa son los responsables del 80% de las ventas (Greenberg & Greenberg, 1980).

Este estudio fue realizado en colaboración con una compañía de seguros con implantación en todo el territorio nacional. Se disponía de datos sobre el rendimiento de 138 candidatos. De este total de candidatos, únicamente el 28% de ellos alcanzó un rendimiento satisfactorio en los tres meses del periodo de prueba. Este porcentaje era similar al mostrado en la literatura. Con anterioridad a este periodo de prueba, se habían obtenido diferentes variables de estos candidatos entre las que se encontraban el factor G, la necesidad económica, la extraversión o la ambición profesional.

Estas variables fueron analizadas por diferentes técnicas de clasificación. Se observó que los mejores resultados fueron obtenidos mediante la utilización de Máquinas de Vectores Soporte con selección de características. Esta técnica obtuvo una clasificación correcta de un 83.6 % utilizando únicamente tres características: sociabilidad, simpatía, y apertura a la experiencia. Además, todos los clasificadores seleccionados eligieron sociabilidad como una de las variables a tener en cuenta. La tasa de clasificación obtenida superaba en más de un 5% a las tasas que se estaban reportando en la literatura. Se calculó que esta mejora en la tasa de clasificación correcta permitiría reducir en 750 euros los costes por candidato seleccionado. Además de reducirse estos costes directos, la compañía también aumentaría las posibilidades de expansión a largo plazo (Kangis & Lago, 1997). El trabajo realizado se muestra a continuación en su totalidad (Delgado-Gomez, Aguado, Lopez-Castroman, Santacruz, & Artes-Rodriguez, 2011).

Expert Systems with Applications 38 (2011) 5129–5132

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Improving sale performance prediction using support vector machines

David Delgado-Gómez [a,*], David Aguado [d], Jorge Lopez-Castroman [c], Carlos Santacruz [d], Antonio Artés-Rodriguez [b]

[a] Department of Statistics, Carlos III University, Madrid, Spain
[b] Department of Signal and Communications Theory, Carlos III University, Madrid, Spain
[c] Department of Psychiatry, Fundacion Jimenez Diaz, Madrid, Spain
[d] Institute of Knowledge Engineering, Autonomous University, Madrid, Spain

ARTICLE INFO

Keywords:
Support vector machines
Sale performance prediction
Recruitment process

ABSTRACT

In this article, an expert system based on support vector machines is developed to predict the sale performance of some insurance company candidates. The system predicts the performance of these candidates based on some scores, which are measurements of cognitive characteristics, personality, selling skills and biodata. An experiment is conducted to compare the accuracy of the proposed system with respect to previously reported systems which use discriminant functions or decision trees. Results show that the proposed system is able to improve the accuracy of a baseline linear discriminant based system by more than 10% and that also exceeds the state of the art systems by almost 5%. The proposed approach can help to reduce considerably the direct and indirect expenses of the companies.

1. Introduction

The expansion and survival of many companies strongly depends on the accomplishment of their sales objectives. Therefore, it is of great importance to select the right salesmen. As shown by Kangis and Lago (1997), a poor recruitment process may conduct the company to lose a leading place and even to its disappearance. Moreover, it has been observed (Kotler, 1994) that the sales obtained by the best salesmen (27% of the total) constitute the 52% of a company's total number of sales. Greenberg and Greenberg (1980) reported even more striking values, finding that the 20% of the salesmen was responsible for the 80% of the company's sales.

These findings indicate that the human resources departments are unable to avoid selecting some low performance salesmen. One reason for the errors in the selection process is the difficult recognition of suitable candidates, able to overcome the obstacles inherent to this particular job. For instance, it has been observed that this job requires a high degree of autonomy and self-organization with almost no supervision (Churchill, Ford, Hartley, & Walker, 1985). Many salesmen feel isolated and give up their jobs for this reason. This produces frequent rotations that affect negatively to the company (Cho & Ngai, 2003). Another problem is the reduced number of candidates. In the analysis provided by O*NET, it is established that more than five million salesmen will be needed during 2004–2014 period.

In order to help the human resource department in the task of recognizing the best candidates, several research projects have proposed some predictors to measure the future performance of a candidate (Cravens, Ingram, Laforge, & Young, 1995). A comparison of the predictive capacity of these features can be found in Vinchur, Schippmann, Switzer, and Roth (1998). Recently, some authors have faced the problem from a different point of view. Instead of determining which are the best features that best predict performance, they analyzed the performance obtained by different classifiers. For instance, Cho and Ngai (2003) used discriminative linear classifiers and decision trees to obtain a correct classification rate close to 70%. In this article, the performance obtained by these classifiers proposed by Cho, and some related ones, is compared with respect to the relatively new support vector machines approach. This technique is considered one of the main advances for data classification in the XX century (Friedman, 2006) and it is widely used in several fields (Furey et al., 2000; Guyon, Weston, Barnhill, & Vapnik, 2002).

The remaining of the article is structured as follows. Section 2 provides a review of some of the most relevant features to predict sale performance that have appeared in the literature. These features will be used afterwards in the experiments. Section 3 describes the different classifier techniques, paying special attention to the support vector machine method. Section 4 displays the results obtained when assessing the accuracy of these classifiers in predicting the sale performance. The article concludes in Section 5 with a discussion of the obtained results.

* Corresponding author. Tel.: +34 664802317.
E-mail address: ddelgado@est-econ.uc3m.es (D. Delgado-Gómez).

0957-4174/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2010.10.003

2. Features

Among the different features proposed in the literature to discriminate high performance salesmen from low performance salesmen, ten features have been selected. These features can be classified in four groups: cognitive and general abilities (features 1 and 2), personality traits (features 3 to 6), selling skills (features 7 to 9) and biodata (feature 10). A brief description of these features follows:

- *Feature 1: G-Factor.* The G-Factor has been widely proposed in the literature as a measure of intelligence. It has been shown that the correlation of this feature with respect to the salesmen's performance ranges from 0.14 (Churchill et al., 1985) to 0.61 (Hunter & Hunter, 1984). In this article, the G-Factor was calculated using a scale composed of 20-item scale similar to the logical series developed by Cattell and Cattell (1994). Each item presented a series of three figures and the examinee had to decide, among four possible options, the one that completed the sequence.
- *Feature 2: Verbal reasoning.* Vinchur et al. (1998) reported a correlation of 0.08 between the verbal reasoning with respect to the salesman's performance. In this article, this feature was estimated using 31 items about synonyms, analogies and differences.
- *Feature 3: Extraversion.* Hough (1992) found that the extraversion of a salesman and the achieved performance were correlated by 0.25. Here, a 11-item scale with four possible answers ranging from agreement to disagreement was used to estimate this feature. An example of one of the items used in this category was "I am a person capable of precisely understand the information and communications in the social contexts".
- *Feature 4: Responsibility.* This feature is considered to be one of the best predictors of sale performance. Mount and Barrick (1995) reported that the correlation between responsibility and sales performance was in the a range of 0.13 to 0.52. In our case, a nine-item scale with four possible answers ranging from agreement to disagreement was used to estimate this feature. An example of one of the items used in this category was "If I have finalized a task, I do not spend time in carefully checking".
- *Feature 5: Agreeableness.* This feature has a slight positive correlation of 0.03 with respect to the sales performance (Vinchur et al., 1998). In our study, it was estimated using a 16-item scale with four possible answers ranging from agreement to disagreement. An example of one of the items used in this category was "I quickly realize when other person needs my help".
- *Feature 6: Openness to the experience.* Unlike previous features, openness to the experience has not been reported to be a good predictor of the salesmen performance. However, this feature can be decomposed in two factors (inner and outer experiences) and it has been indicated that only the outer factor would be correlated with the sales performance (Griffin & Hesketh, 2004). A 10-item scale with four possible answers ranging from agreement to disagreement was used in our sample to estimate this feature. An instance of one of the items was "I usually get enthusiastic when I have to face a new task and I compromise in doing it".
- *Feature 7: Auto-Efficacy* It was estimated using a 11-item scale. One of them was "I usually overcome the difficulties that appear at work".
- *Feature 8: Sociability.* It was estimated using a 11-item scale with four possible answers ranging from agreement to disagreement. One of the items was "I am a person able to correctly interpret the communications".
- *Feature 9: Professional ambition.* The correlation of this feature with the sales performance was estimated to be between 0.30 and 0.62 (Vinchur et al., 1998). In our study, it was estimated using a nine-item scale with four possible answers ranging from agreement to disagreement. An example of one of the items used in this category was "In the professional career, the most important is to achieve the best possible position".
- *Feature 10: Biodata (Economical need).* Biodata has shown a correlation between 0.18 and 0.40 with respect to the salesman performance. In this article, economical need is used as a biodata. It was estimated through five items, one of them being "How many people depend economically on you?".

3. Classification techniques

In this section, a brief review of some of the most commonly used techniques for predicting sales performance are described. Among these techniques, linear, quadratic or Fisher discriminant analysis and logistic regression can be found. The proposed support vector machine technique and decision trees (one of the state of the art techniques) are also described.

- *Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA).* These techniques emerged as a possible solution to the problem of classifying an individual represented by a multivariate vector $\vec{x}$ into one of the two classes π_1 and π_2 with probability density functions $f_1(\vec{x})$ and $f_2(\vec{x})$ respectively. The Bayes rule determines that

$$\vec{x} \in \pi_1 \quad if \quad \frac{f_1(x)}{f_2(x)} \geqslant \frac{L(2,1)p_2}{L(1,2)p_1}$$

where p_i is the prior probability of belonging to the ith class and $L(i,j)$ is the cost of classifying $\vec{x}$ as class j when it belongs to class i. When the density functions are Gaussian with the same variance–covariance matrix defined by the pooled covariance Σ, the analysis receives the name of LDA and the decision boundary is given by

$$x'\Sigma^{-1}(\mu_1 - \mu_2) - \frac{1}{2}\mu_1'\Sigma^{-1}\mu_1 + \frac{1}{2}\mu_2'\Sigma^{-1}\mu_2 = ln\left(\frac{L(2,1)p_2}{L(1,2)p_1}\right)$$

QDA is similar but uses covariance estimates stratified by group instead of the pooled covariance. In this article, equal costs and both priors and no-priors are considered.
- *Fisher Linear discriminant analysis.* A standard two-class Fisher Linear discriminant analysis projects the different data samples into an one-dimensional subspace so that, in a sense, it maximizes the ratio of the between-class scatter to the within-class scatter. Formally, let $\bar{x}$ represent the data, d_1 be the number of data elements corresponding to the first class and let d_2 be the number of elements corresponding to the second class. Let $\bar{x}_1$ and let $\bar{x}_2$ be the class mean vectors of these two classes, $\bar{x}$ be the total mean vector and x_i^j be the jth sample in the ith class, $i = 1, 2$. Then the between matrix is defined by:

$$S_B = d_1(\bar{x}_1 - \bar{x})(\bar{x}_1 - \bar{x})^t + d_2(\bar{x}_2 - \bar{x})(\bar{x}_2 - \bar{x})^t$$

and the within matrix is defined by:

$$S_W = \sum_{i=1}^{2}\sum_{j=1}^{d_i}\left(x_i^j - \bar{x}_i\right)\left(x_i^j - \bar{x}_i\right)^t$$

The projection α that best discriminates the two populations is given by the direction of the eigenvector associated to the maximum eigenvalue of $S_W^{-1}S_B$.
- *Logistic regression.* Let X be a vector containing some explanatory variables X_i, $i = 1,\ldots,n$, let Y be a binary response variable denoting the class which X belongs to (values 0 and 1) and let

$\pi(X)$ denotes the probability that the vector X corresponds to the class one. The logistic regression model has linear form for the logit of this probability given by

$$log\left(\frac{\pi(x)}{1-\pi(x)}\right) = \alpha + \sum_{i=1}^{n} X_i$$

- *Decision trees.* A decision tree is a non-linear classification algorithm. It sequentially divides the data in subgroups so that all or almost all the elements in a subgroup belong to the same class. The algorithm receives its name because the partition has the form of an inverted tree. Different criteria can be used for the data partitioning. In this article, the twoing rule was used. A more detailed description of decision trees can be found in Breiman, Friedman, Olshen, and Stone (1984).
- *Support vector machines.* Support vector machines (Shawe-Taylor & Cristianini, 2000) are non-linear classifiers that have captured the attention of the scientific community in recent years due to their excellent performance. If the M m-dimensional data x_i, $i = 1,\ldots,M$, which belong to the class $y_i(y_i \in \{-1,1\})$ are assumed to be linearly separable, they find the decision function $D(x) = w^t x + b$ where w is a m-dimensional vector, b is a bias term, and for $i = 1,\ldots,M$

$$w^t x_i + b = \begin{cases} \geqslant 1 & \text{for} \quad y_i = 1 \\ \leqslant -1 & \text{for} \quad y_i = -1 \end{cases}$$

However, the assumption that the data are linearly separable is rarely fulfilled. When this assumption is not granted, the support vector machine optimization is extended. In this case, non-negative slack variables ξ_i, $i = 1,\ldots,M$ are introduced. In this case, the optimization problem consists in minimizing

$$\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{M}\xi_i$$

subject to the constrains:

$$y_i(w^t x_i + b) \geqslant 1 - \xi_i \quad \text{for} \quad i = 1,\ldots,M$$

In SVM, to order to obtain non-linear separability, the original data are mapped implicitly into a high dimensional space. In this extended case, the optimization consist on maximizing:

$$\sum_{i=1}^{M}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{M}\alpha_i\alpha_j y_i y_j K(x_i,x_j)$$

subject to the constrains:

$$\sum_{i=1}^{M} y_i\alpha_i = 0$$

where $K(x_i,x_j)$ is called a Kernel.
In this article, the Gaussian Kernel is used. This kernel is defined by

$$k(x_i,x) = e^{-\|x_i-x\|^2/(2\sigma^2)}$$

4. Experimental results

This section aims at objectively assessing what is the effect of the type of classifier in the accuracy of the sale performance prediction. In order to achieve this, an experiment is conducted using a sample of 138 candidates (48.2% males, 51.8% females, mean age: 32 years old) that were selected in the recruitment process of a well-known insurance Spanish company. The company determined their performance based on the number of insurance policies that they sold during a three-months period. The number of policies sold is a standard indicator in the insurance area to measure salesmen performance (Brashear, Bellenger, Ingram, & Barksdale, 1984). The insurance company considers that a salesman obtains a high performance when in the three-months evaluation period is able to sell at least two policies per month. Otherwise, it considers that the salesman achieves a low performance. After the three-months evaluation period, 99 candidates showed a low performance while the remaining 39 achieved a high performance. During the recruitment process, the ten previously described features were obtained for each of the candidates.[1] The means and standard deviations of these ten features for the low and high performance salesmen groups are displayed in Table 1. The correlation of each of these features with respect to the salesmen's performance is also included. Interestingly, there is just a slight difference between the groups. Moreover, the group of salesmen that exhibited a better performance has higher scores in all the features except in economical need (which is a reverse feature). Furthermore, the low values of the correlation indicate a low linear relation between these features and the salesmen's performance.

Table 1
Columns one and two: means and standard deviations of each feature for the low and high performance salesmen groups. Column 3: correlations between the features with respect to the salesmen's performance.

Feature	Low perf. mean (SD)	High perf. mean (SD)	Correlation
G-Factor	6.22(2.54)	6.25(2.53)	0.006
Verbal reasoning	21.45(6.12)	22.67(4.9)	0.094
Extraversion	21.13(3.53)	22.61(4.07)	0.179
Responsibility	29.24(3.19)	29.64(3.74)	0.053
Cordiality	45.48(4.41)	47.07(4.89)	0.156
Experience openness	35.52(3.52)	37.05(3.47)	0.193
Efficacy	31.43(3.58)	32.89(4.96)	0.162
Sociability	31.80(3.52)	32.97(4.48)	0.137
Professional ambition	9.22(1.56)	9.64(1.28)	0.126
Economical need	11.26(4.29)	8.33(3.76)	−0.299

A standard leave one-out cross-validation procedure was used in order to measure their performance. Before showing the obtained results, indicate that the experiment faces two caveats. Firstly, the number of variables is relatively large for the number of available examples. Therefore, the classifiers are exposed to over-fitting. To avoid this problem, feature selection is usually applied. Instead of using all the ten features, only a suitable subset is used. Several techniques can be used to conduct feature selection such as forward, backward, stepwise or mutual information feature selection (Feature Extraction, Foundations & Applications, 2006). However, because data contains only ten variables, an exhaustive analysis was carried out by evaluating the 1023 possible feature subsets. Thus, each classifier chose the subset of features with which a higher accuracy was attained. Secondly, the data is quite unbalanced. Only 28% of the candidates achieved the goals established by the company. Therefore, classifiers are biased toward the majority class. A naive classifier classifying any candidate as low performance salesmen would have an accuracy of 72%.

Table 2 summarizes the obtained results. The most accurate classification results were obtained by the support vector machine and the decision trees techniques. This fact indicates again the non-linearity of the data. Moreover, it is also observed that the linear and quadratic discriminant analysis obtained a poor performance when the priors are not used. When the priors were used, their accuracy was improved by 7%. It can also be observed that the support vector machine approach was able to improve the accuracy of a standard linear discriminant analysis in more than

[1] The human resources department did not have access to these scales.

Table 2
Classification results.

Classified	Correct		Classified	Correct	
	Low	High		Low	High
Linear Discriminant Analysis without priors			*Quadratic Discriminant Analysis without priors*		
Low	68	31	Low	79	21
High	14	25	High	20	18
Total accuracy: 67.39%			Total accuracy: 70.29%		
Linear Discriminant Analysis with priors			*Quadratic Discriminant Analysis with priors*		
Low	93	29	Low	90	24
High	6	10	High	9	15
Total accuracy: 74.64%			Total accuracy: 76.09%		
Fisher Discriminant Analysis			*Logistic regression*		
Low	95	28	Low	94	30
High	4	11	High	5	9
Total accuracy: 76.81%			Total accuracy:74.64%		
Decision trees			*Support vector machines*		
Low	87	18	Low	96	21
High	12	21	High	3	18
Total accuracy: 78.26%			Total accuracy:82.61%		

a 10%. Another interesting fact was that the support vector machine approach used only three out of the ten features: agreeableness, openness to experience and sociability. The sociability variable was chosen by almost all the classifiers and the decision trees approach selected these three features together with the responsibility and professional ambition.

5. Conclusion

This article has analyzed the possibility of improving the accuracy of the current expert system for salesmen selection using support vector machines. Results have shown that the proposed system achieves 5% higher accuracy than the state of the art systems. This conveys a significant cost reduction to the company. For instance, insurance companies spend an average of 600 euros during the three first months for each selected candidate. If 150 candidates were selected, the company would invest 90000 euros to prepare them. Usual procedures, as shown in the introduction, select only 27% of the candidates with suitable sales performance (40 individuals in the example). After the three-months evaluation period, only these most efficient salesmen would remain in the company, at an efficient cost of 90000/40 = 2250 euros per salesman. However, if the proposed system (82% accuracy) was used, the cost per selected salesman would be reduced to 750 euros. The total reduction cost per selected individual would amount 1500 euros. Besides, increasing the rate of efficient salesmen, the company would also reduce the risk of loosing its position in the market. The inconvenience of a second recruitment process and the subsequent delay would be also unnecessary. Finally, the proposed system needs only three features. This allows to obtain the answers to the corresponding three questionnaires in a single interview and determine the future performance of the candidate in the same day.

References

Brashear, T. G., Bellenger, D. N., Ingram, T., & Barksdale, H. C. (1984). Salesperson behavior: Antecedents and links to performance. *Journal of business and industrial marketing, 48*, 9–21.

Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and Regression Trees*. Wadsworth.

Cattell, R. B., & Cattell, A. K. S. (1994). *Factor G Test*. Madrid: Tea Ediciones.

Cho, V., & Ngai, E. W. T. (2003). Data mining for selection of insurance sales agents. *Expert Systems, 20*(3), 123–132.

Churchill, G. A., Ford, N. M., Hartley, S. W., & Walker, O. C. (1985). The determinants of salesforce performance: A meta-analysis. *Journal of Marketing Research, 22*, 103–118.

Cravens, D. W., Ingram, T. N., Laforge, R. W., & Young, C. E. (1995). Behaviour based and outcome based salesforce control system. *Journal of Marketing, 54*, 68–81.

Feature Extraction, Foundations and Applications. (2006). I. Guyon, S. Gunn, M. Nikravesh, & L. A. Zadeh (Eds.), *Series studies in fuzziness and soft computing*. Physica-Verlag, Springer.

Friedman, J. H. (2006). Recent advances in predictive (machine) learning. *Journal of classification, 23*(2), 175–197.

Furey, T. S., Cristianini, N., Duffy, N., Bednarski, D. W., Schummer, M., & Haussler, D. (2000). Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics, 16*(10), 906–914.

Greenberg, H. M., & Greenberg, J. (1980). Job matching for better sales performance. *Harvard Business Review*(September), 128–133.

Griffin, B., & Hesketh, B. (2004). Why openness to experience is not a good predictor of job performance. *International Journal of selection and assessment, 12*(3), 243–251.

Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine Learning, 46*, 389–422.

Hough, L. M. (1992). The big five personality variable-construct confusion: Description versus prediction. *Human Performance, 5*, 139–155.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96*, 72–98.

Kangis, P., & Lago, H. (1997). Using Caliper to predict performance of salespeople. *International Journal of Manpower, 18*(7), 565–575.

Kotler, P. (1994). *Marketing Management*. Englewood Cliffs, NJ: Prentice-Hall.

Mount, M. K., & Barrick, M. R. (1995). The big five personality dimensions: Implications for the research and practice in human resource management. *Research in Personnel and Human Resource Management, 13*, 153–200.

Shawe Taylor, J., & Cristianini, N. (2000). *Support Vector Machines and other kernel based learning methods*. Cambridge University Press.

Vinchur, A. J., Schippmann, J. S., Switzer, F. S., & Roth, P. L. (1998). A meta-analytic review of predictors of job performance for salespeople. *Journal of applied psychology, 83*, 586–597.

3. Aplicación de técnicas no supervisadas de reconocimiento de patrones para mejorar el diagnóstico clínico y la identificación en los procesos de selección de personal

Los clasificadores utilizados en los trabajos expuestos anteriormente eran técnicas supervisadas ya que hacían uso de la etiqueta asociada a cada dato para ajustar sus parámetros. En concreto, estas etiquetas informaban si el participante había realizado un intento de suicidio recientemente o si había mostrado un alto rendimiento. Sin embargo, en ocasiones, no se dispone de estas etiquetas (o conviene abordar el problema sin ellas) y nos enfrentamos a un problema de clasificación no supervisada.

En este capítulo de la tesis, se abordan problemas similares a los precedentes pero solucionándolos a través de técnicas no supervisadas. Un referente claro son los modelos psicométricos que se utilizan desde la teoría de respuesta a los ítems en el contexto de los Test Adaptativos Informatizados. En el área de la salud, se han realizado dos trabajos que han sido motivados por la reciente aparición de varios artículos que proponen el uso de Test Adaptativos Informatizados como herramienta rápida de apoyo al diagnóstico clínico. En particular, se ha propuesto el uso de Test Adaptativos Informatizados como apoyo al diagnóstico de depresión (Fliege, Becker, Walter, Bjorner, Klapp, & Rose, 2005), ansiedad (Gibbons, Weiss, Pilkonis, Moore, Kim, & Kupfer, 2014), desórdenes de personalidad (Simms, Goldberg, Roberts, Watson, Welte, & Rotterman, 2011) o recientemente comportamiento suicida (De Beurs, de Vries, de Groot, de Keijser, & Kerkhof, 2014).

En el primero de los trabajos, que se expone a continuación, se utilizará el modelo de respuesta graduada de Samejima (Samejima, 1969) para analizar las propiedades del cuestionario de salud general de Goldberg en su versión de 28 ítems (General Health Questionnaire-28, GHQ-28) cuando es administrado a una muestra de víctimas del terrorismo. El cuestionario GHQ-28 se aplicó a 162 víctimas directas de ataques terroristas y a 729 familiares de las víctimas. Se estudia la fiabilidad de las puntuaciones y se aportan evidencias de validez mediante estudios factoriales y de relación con otras variables. El estudio factorial indicó que tres factores explican la varianza compartida entre los ítems del GHQ-28. A los ítems que definen cada uno de ellos se les aplicó el modelo de Samejima para realizar un estudio detallado de sus propiedades. Se compararon los resultados con los obtenidos en otras muestras y se observó la preeminencia de síntomas de ansiedad en ambos colectivos (víctimas y familiares) analizados.

Por otro lado, en el área de las ciencias del comportamiento, el segundo trabajo que se muestra en este capítulo analiza si el modelo de respuesta graduada de Samejima (1969) puede ser utilizado como una herramienta biométrica que permita identificar a un individuo a través de su personalidad (Delgado-Gómez, Sukno, Santarcruz, Aguado, & Artés-Rodriguez, 2010) lo que facilitaría el desarrollo de sistemas biométricos para verificar si la persona que se está entrevistando en un proceso selectivo es la misma que anteriormente ha realizado diversas pruebas online. La cuestión entronca directamente en la problemática actual relacionada con la administración de test a través de internet en entornos no controlados en lo que se denomina habitualmente Unproctored Internet Testing (UIT). En esencia, el trabajo explora, que sepamos por vez primera, las posibilidades de la evaluación psicométrica para la identificación de personas, a partir de Tests Adaptativos Informatizados que evalúan los cinco grandes rasgos de la personalidad. Los resultados fueron positivos. Con la aplicación de solo 45 ítems tipo Likert, que requiere no más de 3 minutos, se consiguió una tasa de identificación similar a la obtenida con los procedimientos biométricos tradicionales.

A continuación se muestran estos dos trabajos.

3.1 Evaluación de víctimas del terrorismo mediante el GHQ-28

Psychological Assessment
2013, Vol. 25, No. 1, 279–287

Psychometrical Assessment and Item Analysis of the General Health Questionnaire in Victims of Terrorism

David Delgado-Gomez
Universidad Carlos III

Jorge Lopez-Castroman
Inserm, U1061 and IIS-Fundacion Jimenez Diaz

Victoria de Leon-Martinez
IIS-Fundacion Jimenez Diaz

Enrique Baca-Garcia
IIS-Fundacion Jimenez Diaz, New York State Psychiatric Institute, and Columbia University

Maria Luisa Cabanas-Arrate
and Antonio Sanchez-Gonzalez
Fundación Instituto de Victimología

David Aguado
Instituto Ingeniería del Conocimiento

There is a need to assess the psychiatric morbidity that appears as a consequence of terrorist attacks. The General Health Questionnaire (GHQ) has been used to this end, but its psychometric properties have never been evaluated in a population affected by terrorism. A sample of 891 participants included 162 direct victims of terrorist attacks and 729 relatives of the victims. All participants were evaluated using the 28-item version of the GHQ (GHQ-28). We examined the reliability and external validity of scores on the scale using Cronbach's alpha and Pearson correlation with the State-Trait Anxiety Inventory (STAI), respectively. The factor structure of the scale was analyzed with varimax rotation. Samejima's (1969) graded response model was used to explore the item properties. The GHQ-28 scores showed good reliability and item-scale correlations. The factor analysis identified 3 factors: anxious-somatic symptoms, social dysfunction, and depression symptoms. All factors showed good correlation with the STAI. Before rotation, the first, second, and third factor explained 44.0%, 6.4%, and 5.0% of the variance, respectively. Varimax rotation redistributed the percentages of variance accounted for to 28.4%, 13.8%, and 13.2%, respectively. Items with the highest loadings in the first factor measured anxiety symptoms, whereas items with the highest loadings in the third factor measured suicide ideation. Samejima's model found that high scores in suicide-related items were associated with severe depression. The factor structure of the GHQ-28 found in this study underscores the preeminence of anxiety symptoms among victims of terrorism and their relatives. Item response analysis identified the most difficult and significant items for each factor.

Keywords: factor analysis, item response theory, terrorism victims, GHQ-28

In 2010, 11,604 terrorist attacks killed 13,186 and injured 30,665 persons globally (National Counter Terrorism Center, 2011). The indirect consequences of these attacks on mental health may be more frequent (10–100 psychiatric casualties for every physical injury), and they have more significance, cost and long-term effects wise, than previously thought (Ursano & Friedman, 2006). The families of the victims, independently of their own exposure, may also present a high risk for the development of psychopathology (Hoven et al., 2009; Stoddard et al., 2011). The most frequent mental disorders in the aftermath of a terrorist attack are posttraumatic stress disorder (PTSD), depression, and other anxiety disorders (Henriksen, Bolton, & Sareen, 2010; Salguero, Cano-Vindel, Iruarrizaga, Fernandez-Berrocal, & Galea, 2011). The aftereffects of the trauma exposure can affect victims and their

This article was published Online First December 3, 2012.

David Delgado-Gomez, Department of Statistics, Universidad Carlos III, Madrid, Spain; Jorge Lopez-Castroman, Inserm, U1061, Montpellier, France, and IIS-Fundacion Jimenez Diaz, Department of Psychiatry, CIBERSAM, Madrid, Spain; Victoria de Leon-Martinez, IIS-Fundacion Jimenez Diaz, Department of Psychiatry, CIBERSAM; Enrique Baca-Garcia, IIS-Fundacion Jimenez Diaz, Department of Psychiatry, CIBERSAM, and New York State Psychiatric Institute, and Department of Psychiatry, Columbia University; Maria Luisa Cabanas-Arrate and Antonio Sanchez-Gonzalez, Instituto de Investigación en Psiquiatría y Psicoterapia, Fundación Instituto de Victimología, Madrid, Spain; David Aguado, Instituto Ingeniería del Conocimiento, Madrid, Spain.

David Delgado-Gomez and Jorge Lopez-Castroman have contributed equally to this work. The work by Jorge Lopez-Castroman was supported by a FondaMental Foundation grant.

Correspondence concerning this article should be addressed to Jorge Lopez-Castroman, IIS-Fundacion Jimenez Diaz, Department of Psychiatry, Avenida Reyes Catolicos 2, 28040 Madrid, Spain. E-mail: jlopez@fjd.es

relatives for years (Baca, Cabanas, & Baca-García, 2002). PTSD in particular often follows a chronic course in the general population (Breslau et al., 1998; Kessler, Sonnega, Bromet, Hughes, & Nelson, 1995) and after exposure to terrorist attacks (DiGrande, Neria, Brackbill, Pulliam, & Galea, 2011; Kawana, Ishimatu, & Kanda, 2001; North, Pfefferbaum, Kawasaki, Lee, & Spitznagel, 2011).

Although early detection of mental disorders in exposed populations may prevent future disabilities (Stoddard et al., 2011), the performance of screening instruments after a terrorist attack may present limitations. However, they have seldom been studied (Brewin, Fuchkan, Huntley, & Scragg, 2010). For instance, the diagnostic performance of screening instruments, such as the Trauma Screening Questionnaire, in an affected population was improved when the interval from the terrorist attacks was longer (Brewin et al., 2010). Moreover, there is a specific need for outreach to detect individuals with mental disorders after terrorist attacks (Brewin et al., 2008; Pfefferbaum, North, Flynn, Norris, & DeMartino, 2002).

One instrument that has been successfully used to assess the psychopathology of terrorism victims and their relatives is the General Health Questionnaire (GHQ; Baca et al., 2002; Boscarino, Figley, & Adams, 2004). The GHQ, in all its versions, is a self-administered Likert-type questionnaire frequently used to detect nonpsychotic mental disorders. The results of the GHQ might depend on the characteristics of the sample. For instance, the GHQ provided greater sensitivity and specificity in a clinical sample of 382 adult females than in a comparable sample of 154 males (Hobbs, Ballinger, Greenwood, Martin, & McClure, 1984). Differences were also observed among 260 primary care respondents: respondents younger than 40 years of age had 3 times more risk of being misclassified with the GHQ than respondents older than 40 years of age (de Jesus Mari & Williams, 1986). Moreover, Tarnopolsky et al. indicated that validity coefficients could differ from one population to another. Therefore, it is necessary to analyze the reliability and external validity of the scores of the GHQ and its factor structure in any population under study (Tarnopolsky, Hand, McLean, Roberts, & Wiggins, 1979). To date, the GHQ has been validated in several populations varying in age (Baksheev, Robinson, Cosgrave, Baker, & Yung, 2011; Costa et al., 2006), sex (Romans-Clarkson, Walton, Herbison, & Mullen, 1989; Tran, Tran, & Fisher, 2012), nationality (Politi, Piccinelli, & Wilkinson, 1994; Quek, Low, Razack, &Loh, 2001), and work status (Banks et al., 1980), however not in terrorism victims. Moreover, Tarnopolsky et al. supported the validation of the GHQ on any particular population because validity coefficients could differ from one setting to another (Tarnopolsky et al., 1979).

The reliability, factor structure, and external validity represent the classical test theory in psychological assessment. This traditional assessment can be complemented with the item response theory (IRT), which is not commonly used to evaluate clinical instruments (Thomas, 2011) and, to our knowledge, has never been applied to the GHQ. The IRT analyses provide indices of discrimination and difficulty for each item, and therefore a deeper understanding of the questionnaire (P. J. Ferrando, 2001; Ploubidis & Frangou, 2011).

In this article, the psychometric properties of the GHQ as a screening questionnaire in a population of terrorism victims and their relatives were evaluated. Although the questionnaire has been used before to assess terrorism victims and their relatives (Baca et al., 2002), its psychometric properties have never been analyzed in this population. Thus, the first goal of this study was to analyze the reliability and external validity of scores of the 28-item version of the GHQ (GHQ-28) and its factor structure in this population. The second goal was to estimate the parameters of each item of the GHQ-28 using the IRT. The GHQ-28 items were then sorted according to the IRT. The item characteristic curves (ICCs) provided a visual understanding of the item responses. In conclusion, the IRT allowed a more robust estimate of the latent variable, given that the item parameters are independent of the sample.

Method

Participants

The sample was composed of 891 participants: 162 participants had experienced a terrorist attack, and 729 participants were relatives of the victims. The sample was obtained from interviews with victims of terrorist attacks in Spain and their immediate families (first-degree relatives). The average time from the terrorist attacks to the interviews was 13.5 years ($SD = 6.6$). This sample is part of the Phoenix Project (Baca et al., 2002), which was funded by the Association of Victims of Terrorism and designed to study the long-term consequences of terrorist attacks. The types of the terrorist attacks were mostly bombings (58.7%) or shootings (36.2%). All participants effectively completed the GHQ-28 and the State-Trait Anxiety Inventory (STAI; Spielberger, Gorsuch, & Lushene, 1970).

The mean age of the sample was 38.7 ± 14.8 years. Although there was a majority of females in the total sample (60.6%), the victims were composed of mostly males (69.7% males). The victims and their relatives were mostly married (50.3%) or single (34.4%). Regarding the level of education in the sample, 28.2% reported primary studies or less, 46.0% reported secondary studies, and 25.8% reported university studies. A majority of the victims (74.9%) were policemen or army soldiers.

Measures

The GHQ-28. The GHQ-28 is a Likert-type scale composed of 28 items selected from the 60-item GHQ version by a principal component analysis (Goldberg & Hillier, 1979). Each item receives a discrete score ranging from 0 to 3. The 28 items were then divided into four groups. Items in each group measure a different trait. These four traits were somatic, anxiety and depression symptoms, and social dysfunction (see Table 2). Three different approaches have appeared in the literature to evaluate the questionnaire. They differ in the way the score of each item is calculated. In the classical GHQ approach, values 0 and 1 score a point value of 0, whereas values 2 and 3 score a point value of 1. In the Likert approach, values are not transformed. There also exists a third uncommon approach, called the c-GHQ, where the score of each item relies on whether the item is formulated positively or negatively (Richard, Lussier, Gagnon, & Lamarche, 2004). In this article, the Likert approach was used. It has been observed that the Likert method is the preferable method as it produces less skewed distributions and is more suitable for correlation analysis and intergroup comparisons (Banks et al., 1980). The Spanish version of the questionnaire was handed to the participants (Lobo & Muñoz, 1996).

The STAI. The STAI (Spielberger et al., 1970) is a Likert-type self-report questionnaire composed of 40 items. The first 20 items measure anxiety as a state, whereas the last 20 items measure anxiety as a trait with a single score for each. Previous literature has demonstrated a high reliability and validity for STAI scores in different samples (Quek, Low, Razack, Loh, & Chua, 2004; Zhang & Gao, 2012).

Procedure

Reliability, factor analysis, and external validity. The reliability of the GHQ-28 was estimated using Cronbach's alpha. The adjusted correlations of each item with the GHQ-28 were also calculated. An exploratory factor analysis using varimax rotation was conducted to discover the structure of the GHQ-28. A maximum likelihood estimator was used for the exploratory factor analysis. Coefficients obtained with this method were similar to those obtained with other estimation methods, such as principal components and generalized least squares, except for an inversion of the second and third factor. However, the second and third factor presented very similar eigenvalues using any of the aforementioned methods (data not shown). In order to determine the factors to be retained, a parallel analysis was conducted following Hayton, Allen, and Scarpello (2004). The Bartlett method was used to calculate the factor scores. Finally, the external validity was analyzed in comparison with a questionnaire measuring anxiety (STAI) in agreement with Sánchez-López and Dresch (2008). We estimated the correlation between the STAI (trait and state scores) and the GHQ-28 (global and factor scores).

The graded response model. The IRT is based on the assumption that the answer given by an individual to a specific item depends on an individual's latent variable and the properties of the item. Changes to the latent variable will modify the response probability. Moreover, it is assumed that the latent variable under study follows a standard Gaussian distribution. This assumption allows the determination of the percentile at which an individual can be placed in a given population. Specifically, let i denote an item of an item bank, and let n_i be the number of possible choices for the ith item. Let k be the number of examinees. The probability that the kth examinee with a given latent variable value θ^k chooses the answer g $\in\{0, \ldots, n_i - 1\}$ on an item i is represented by $P_{i(g)}(\theta k)$. Several models have been developed in order to define this probability (van der Linden & Hambleton, 1997).

Among the different item response models, we used Samejima's (1969) graded response for the two following reasons: (a) Responses are ordered, and (bi) specific parameters for discrimination and difficulty are obtained for each item, allowing us to differentiate participants with the same score on two different items. This model calculates the probabilities of answering each of the item choices given the trait level θ^k of the examinee. The probability that the kth individual will choose the category g or higher on the ith item is expressed by,

$$P^*_{i(g)} = \frac{e^{a_i\left(\theta^k - b_{i(g)}\right)}}{1 + e^{a_i\left(\theta^k - b_{i(g)}\right)}},$$

where a_i is the discrimination parameter, $b_{i(g)}$ indicates the difficulty specified for the answer g, and the difficulty parameters satisfy that $b_{i(g)} < b_{i(g+1)}$. The probability $P_{i(g)}$ is computed as a difference of the cumulative probabilities for adjacent answers:

$$P_{i(g)} = P^*_{i(g)} - P^*_{i(g+1)}.$$

This formula requires the probabilities when $g = 0$ and $g = n_i$ for all the items:

$$P^*_{i(0)}\left(\theta^k\right) = 1, P^*_{i(n_i)}\left(\theta^k\right) = 0.$$

The IRT term *difficulty* cannot be directly applied to psychological assessment with the GHQ-28 as there are no correct or incorrect answers to the test and is substituted by the term *location* henceforward. The discrimination and location parameters that characterize each item of a questionnaire are estimated by the marginal maximum likelihood technique using the answers of a sample of examinees. The item location identifies the value of the latent variable where the probability of choosing or not choosing a specific answer to an item is the same. The higher the discrimi-

Table 1
Adjusted Item-Scale Correlation and Cronbach's Alphas After Removing Each Item in the GHQ-28

Item	Adjusted item-scale correlation	Cronbach's α if the item is eliminated
1	.603	0.954
2	.647	0.954
3	.766	0.953
4	.684	0.953
5	.589	0.954
6	.649	0.954
7	.615	0.954
8	.673	0.954
9	.657	0.954
10	.774	0.953
11	.769	0.953
12	.695	0.953
13	.773	0.953
14	.762	0.953
15	.288	0.956
16	.565	0.955
17	.548	0.955
18	.513	0.955
19	.514	0.955
20	.453	0.955
21	.584	0.954
22	.665	0.954
23	.729	0.953
24	.702	0.953
25	.602	0.954
26	.782	0.953
27	.689	0.954
28	.623	0.954
	Internal consistency	
Total sample	α = 0.955	Std α = 0.954
Victims	α = 0.962	Std α = 0.961
Relatives	α = 0.951	Std α = 0.950

Note. Internal consistency of the GHQ-28 as measured with Cronbach's alphas (Standardized) for the total sample (victims and their relatives). GHQ-28 = 28-item version of the General Health Questionnaire; Std = Standardized.

Table 2

Eigenvalues, Percentage of Explained Variance for Each Factor, and Maximum Likelihood Estimate of Factor Loadings for the GHQ-28

Factors	1	2	3
Eigenvalues and percentage of explained variance before rotation			
Eigenvalue	12.32	1.79	1.40
% Variance	43.99	6.39	5.01
% Cumulative variance	43.99	50.39	55.40
Eigenvalues and percentage of explained variance after rotation			
Eigenvalue	7.95	3.85	3.70
% Variance	28.40	13.75	13.24
% Cumulative variance	28.40	42.16	55.40
Factor loadings (Item: Have you recently . . .)			
Somatic			
1. been feeling perfectly well and in good health?	**0.53**	0.34	0.09
2. been feeling in need of a good tonic?	**0.61**	0.25	0.16
3. been feeling run down and out of sorts?	**0.70**	0.32	0.20
4. felt that you are ill?	**0.62**	0.28	0.18
5. been getting any pains in your head?	**0.61**	0.13	0.13
6. been getting a feeling of tightness or pressure in your head?	**0.66**	0.15	0.17
7. been having hot or cold spells?	**0.61**	0.16	0.15
Anxiety			
8. lost much sleep over worry?	**0.71**	0.15	0.16
9. had difficulty in staying asleep once you are off?	**0.67**	0.19	0.14
10. felt constantly under strain?	**0.79**	0.20	0.22
11. been getting edgy and bad-tempered?	**0.79**	0.19	0.23
12. been getting scared or panicky for no good reason?	**0.64**	0.20	0.26
13. been satistifed with the way you've carried out your task?	**0.71**	0.25	0.29
14. felt constantly under strain?	**0.76**	0.21	0.24
Social Dysfunction			
15. been managing to keep yourself busy and occupied?	0.14	**0.40**	0.01
16. been taking longer than usual to do things?	0.34	**0.52**	0.14
17. felt on the whole you were doing things well?	0.22	**0.71**	0.15
18. been satisfied with the way you've carried out your tasks?	0.18	**0.72**	0.13
19. felt that you are playing a useful part in things?	0.16	**0.60**	0.29
20. felt capable of making decisions about things?	0.09	**0.58**	0.27
21. been able to enjoy your normal day-to-day activities?	0.36	**0.47**	0.22
Depression			
22. been thinking of yourself as a worthless person?	0.35	**0.45**	**0.45**
23. felt that life is entirely hopeless?	**0.43**	**0.43**	**0.46**
24. felt that life isn't worth living?	0.38	0.33	**0.60**
25. thought of the possibility that you might do away with yourself?	0.23	0.16	**0.85**
26. found at times you couldn't do anything because your nerves were too bad?	**0.65**	0.28	0.38
27. found yourself wishing you were dead and away from it all?	0.33	0.30	**0.72**
28. found that the idea of taking your own life kept coming into your mind?	0.25	0.19	**0.83**

Note. GHQ-28 = 28-item version of the General Health Questionnaire. Bold type identifies .40 or larger loads.

nation value, the easier the discrimination is between the distinct levels of the latent variable. These parameters are graphically represented in the ICCs. The ICCs represent the probability of each item response conditioned on the specific values of the latent variable. The slope of the curve of the location values determines the discrimination of an item. An example of ICC with three possible answers, and a discriminative parameter (a_i) of 2.5, and location parameters (b_i) of −1 and 1 are displayed in Figure 2. In this article, the ICCs have been summarized in an expected average score for each item, in accordance with:

$$\sum_{k=0}^{n_i} k \int_{-\infty}^{\infty} P_{i(k)}(\theta) f(\theta)\, \partial\, \theta,$$

where $f(\theta)$ is the density function of a standard Gaussian distribution. The average score allows us to sort the items by their inherent location for a given latent variable. For instance, a person with a higher latent variable would obtain higher scores in low-average score items than a person with a smaller latent variable.

All procedures were performed using the ltm R statistical package.

Results

Reliability

Table 1 shows the Cronbach's alphas (and the standardized alpha) for the total sample, the victims and their relatives together with the adjusted item-scale correlation, and the Cronbach's alphas after removing each item. The calculated Cronbach's alphas were superior to 0.95 in all cases. The adjusted item-scale correlation ranged from .45 to .78 for all the items, except Item 15 in which the item-scale correlation was .29.

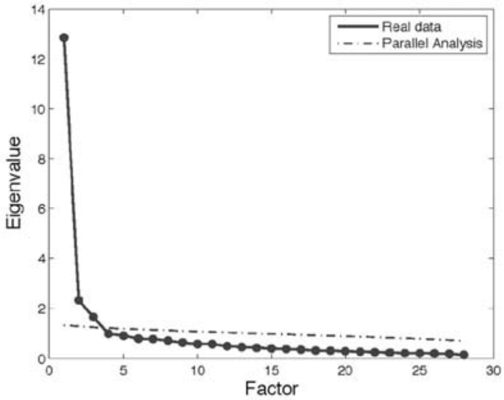


Figure 1. Plot of actual versus randomly generated eigenvalues.

Factor Analysis

Three factors were retained by the parallel analysis. They also showed an associated eigenvalue higher than 1 (see Figure 1). The eigenvalues of these factors together with the corresponding percentage of explained variance and the factor loadings are displayed in Table 2. The highest loadings of the first factor were found in the first 14 items (somatic and anxiety symptoms) and Item 26 (depression symptoms). The highest loadings of the second factor were located mainly in the items measuring social dysfunction symptoms. Finally, with the exception of Item 26, the highest loadings of the third factor were found in the items indicating depression symptoms. The loading of Item 22 was very similar for the first and third factor, but it was classified within the third factor in agreement with previous reports. When comparing the victims of terrorism and their relatives with regards to the factor scores, we found significant differences for the first (p = .046), second (p < .001), and third factor (p = .028). Factor structure and internal validity of samples separated by gender or class (victims or family members) showed no differences.

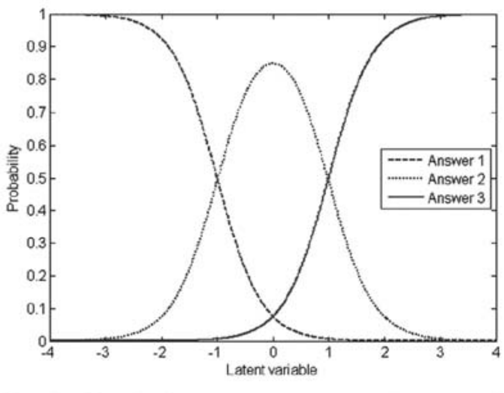


Figure 2. Example of an item characteristic curve with three possible answers: The discriminative parameter (a) is 2.5, and the location parameters (b_i) are −1 and 1.

Table 3

Correlation of the GHQ Global Score and the Obtained Factors With the State and Trait Global Scores of Anxiety, and Cronbach's Alphas for STAI Internal Validity

GHQ	STAI	
	State score r (p)	Trait score r (p)
Global score	.80 (<.0001)	.79 (<.0001)
Factor I	.67 (<.0001)	.65 (<.0001)
Factor II	.37 (<.0001)	.33 (<.0001)
Factor III	.35 (<.0001)	.39 (<.0001)
STAI Internal validity	State α	Trait α
Victims of terrorism	0.9596	0.9337
Relatives	0.9449	0.9189
Total	0.9492	0.9229

Note. GHQ = General Health Questionnaire; STAI = State-Trait Anxiety Inventory; r = Pearson's coefficient.

External validity. Table 3 shows the correlation of the GHQ-28 global score and the obtained factors with the state and trait global scores of anxiety according to the STAI. The first factor had the greatest correlation with the STAI resulting in a state score of 0.67 and a trait STAI score of 0.65.

IRT-based analyses. The factor analysis indicates that the GHQ-28 measures three dimensions or latent variables. We conducted an IRT-based analysis for each. On the basis of the factor analysis, items numbering from 1 to 14 and Item 26 were associated with the first latent variable, items numbering from 15 to 21 were associated with the second latent variable, and the remaining items were associated with the third latent variable. Table 4 shows the discrimination and location parameters for each item. The ICCs of the items with the highest and lowest expected average value for each dimension are displayed in Figure 2. We compared the average latent variables for each factor according to the IRT between the victims of terrorism and their relatives using t tests. We found significant differences (p < .001) between the groups in the three factors.

Discussion

In this study, we analyzed the psychometric properties, including an item analysis using IRT, of the GHQ-28 in a sample of terrorism victims and their relatives. The reliability of the scores of GHQ-28 was high, as demonstrated by an alpha value of over 0.955 in the total sample (Bland & Altman, 1997). Moreover, with the exception of Item 15 ("Have you recently felt that you are playing a useful part in things?"), all items exhibited a reasonably good adjusted item-scale correlation with values ranging from .45 to .78. The factor structure and internal validity did not differ when separating the sample by gender or class (victims or family members). Previous literature supports a continuous increase of risk for mental disorders depending on the degree to which participants are affected by a terrorist act, independently of being directly affected or through a familiar bond (Baca, Cabanas, Perez-Rodriguez, &

Table 4
Discrimination and Location Parameters for Each Item Sorted by Its Expected Average Score

	Item (Have you recently . . .)	A	B1	B2	B3	Av. score
	Factor 1					
12	been getting scared or panicky for no good reason?	2.11	0.15	1.34	2.29	1.48
7	been having hot or cold spells?	1.66	0.01	1.31	2.71	1.57
26	found at times you couldn't do anything because your nerves were too bad?	2.42	−0.11	1.04	2.17	1.61
6	been getting a feeling of tightness or pressure in your head?	1.86	−0.19	0.95	2.19	1.62
4	felt that you are ill?	1.88	−0.29	1.00	2.25	1.66
13	been satistifed with the way you've carried out your task?	2.70	−0.33	0.72	1.85	1.70
14	felt constantly under strain?	3.03	−0.41	0.68	1.81	1.73
2	been feeling in need of a good tonic?	1.78	−0.48	0.96	2.49	1.76
9	had difficulty in staying asleep once you are off?	1.90	−0.82	0.55	1.83	1.80
3	been feeling run down and out of sorts?	2.54	−0.71	0.67	1.94	1.84
5	been getting any pains in your head?	1.56	−0.93	0.85	2.43	1.85
11	been getting edgy and bad-tempered?	3.35	−0.81	0.41	1.56	1.86
8	lost much sleep over worry?	2.10	−0.98	0.51	1.81	1.87
10	felt constantly under strain?	3.33	−0.70	0.44	1.75	1.89
1	been feeling perfectly well and in good health?	1.52	−2.72	0.78	2.68	2.13
	Factor 2					
15	been managing to keep yourself busy and occupied?	0.94	−1.91	2.27	3.89	1.85
20	felt capable of making decisions about things?	1.73	−1.43	1.31	2.76	1.93
19	felt that you are playing a useful part in things?	1.79	−1.66	1.24	2.49	1.95
18	been satisfied with the way you've carried out your tasks?	3.22	−1.11	0.94	2.18	1.95
17	felt on the whole you were doing things well?	3.52	−1.40	1.17	2.35	1.98
21	been able to enjoy your normal day-to-day activities?	1.56	−1.98	0.90	2.47	2.02
16	been taking longer than usual to do things?	1.65	−2.54	0.88	2.50	2.09
	Factor 3					
27	found yourself wishing you were dead and away from it all?	3.50	0.84	1.59	2.62	1.27
28	found that the idea of taking your own life kept coming into your mind?	2.46	0.85	1.62	2.70	1.28
25	thought of the possibility that you might do away with yourself?	2.55	0.73	1.51	2.44	1.31
24	felt that life isn't worth living?	3.21	0.47	1.32	2.11	1.36
22	been thinking of yourself as a worthless person?	2.15	0.51	1.31	2.19	1.36
23	felt that life is entirely hopeless?	2.45	0.41	1.27	2.10	1.38

Note. A = discrimination parameter; B1, B2, and B3 = location parameters of the three possible responses to each item; Av. score = Average score of the item.

Baca-Garcia, 2004). These reasons lead us to analyze victims of terrorism and their family members together.

The factor analysis identified three factors. The highest loadings of the first factor appeared in the group of anxiety and somatic symptoms. The highest loadings of the second and third factor were associated with the social dysfunction and depression items, respectively. Previous reports identified similar factors (Aderibigbe, Riley, Lewin, & Gureje, 1996; Gibbons, Flores de Arévalo, & Mónico, 2004; Vallejo, Jordan, Diaz, Comeche, & Ortega, 2007; Werneke, Goldberg, Yalcin, & Ustun, 2000). Notably, the highest loadings of the third factor correspond to symptoms of severe depression associated with suicidal behavior (see Items 24, 25, 27, and 28 in Table 2). This finding go in line with the increased risk of suicide that has been reported in a sample of victims after the 11 March terrorist attacks in Madrid (Conejo-Galindo et al., 2008). Finally, the highly significant correlation between the STAI scores and the three factors of the GHQ-28 indicates a good external validity.

After varimax rotation, the first factor exhibited more variance (28.4%) than the other two jointly (13.8% and 13.2% for the second and third factor, respectively). Consequently, the GHQ-28 could be interpreted as a one-dimensional measure of psychiatric disorders (Banks et al., 1980) or, on the contrary, as a multidimensional measure in which all factors are needed to establish an individual's profile (Huppert, Walters, Day, & Elliott, 1989). A closer analysis of the first factor may favor a one-dimensional view in this sample. In support of this view, the highest loadings of the first factor were found in the anxiety items. The predominance of anxiety symptoms seems reasonable in a population who is particularly at risk for anxiety disorders due to their traumatic experiences (Stoddard et al., 2011). Moreover, the highly significant correlation between both the trait and state scores of the STAI and the three factors of the GHQ-28 supports the central role of anxiety in this population. Somatic symptoms, which are frequently associated with underlying anxiety (Romera et al., 2010; Zhu et al., 2012), also presented high loadings in the first factor.

The IRT analysis demonstrates the importance of items within each identified factor. Interestingly, we found a certain gradient in the first factor that could reflect the severity of anxiety symptoms. We found that the five most discriminative items were Items 12 and 26, both related with subjective feelings of anxiety, and Items 7, 6, and 4, related with somatic symptoms of anxiety. On the contrary, the least discriminative items are those questions that address anxiety symptoms indirectly. For instance, Item 1 ("good health"), Item 10 ("under strain"), or Item 8 ("lost sleep"). Within the second factor (social dysfunction), the most discriminative items appeared to be closely related with the capacity to commit

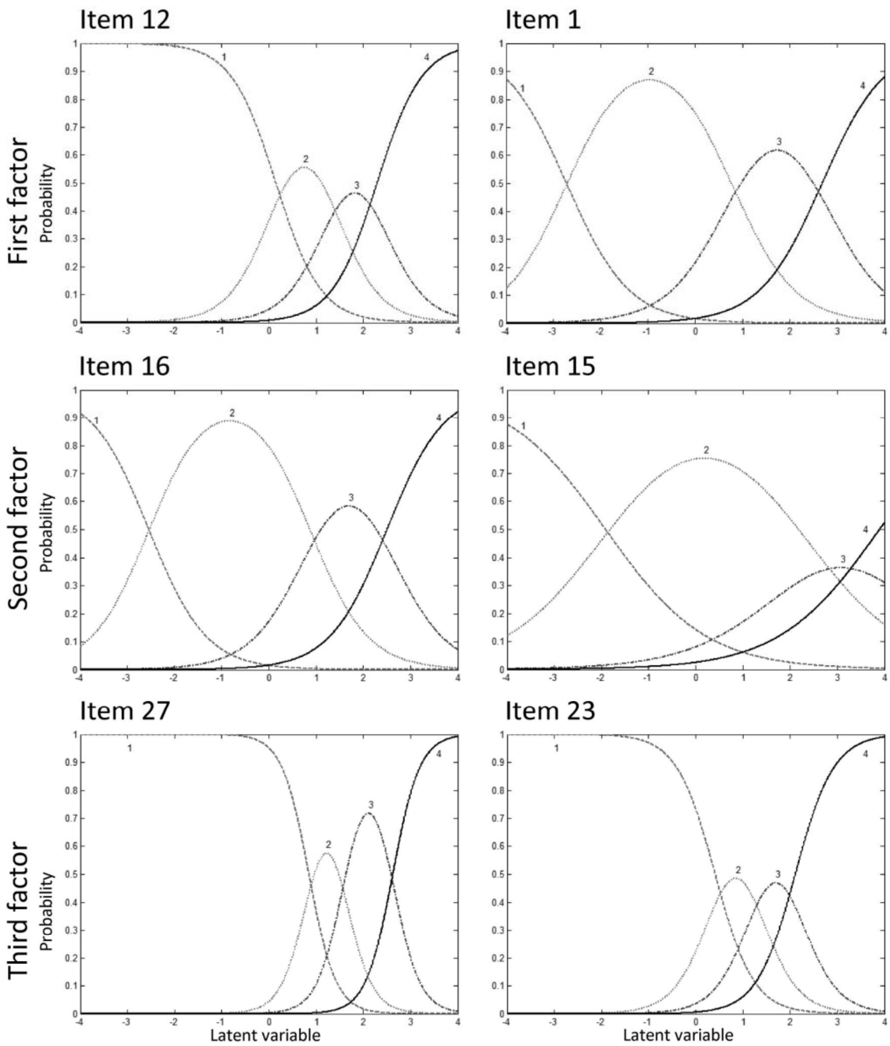


Figure 3. Item characteristic curves of the items with the lowest and highest expected average score of each factor.

oneself to any task (Item 15) and to making decisions (Item 20). Finally, the most discriminative items among the depression symptoms seem to be the items regarding passive or active ideas of suicide (Items 25, 27, and 28). On the contrary, low self-esteem and hopelessness (Items 22 and 23), usual descriptors of depression, appeared to be the least discriminative.

The ICCs (see Figure 3) show that items included in the same factor may have very different discrimination powers. For instance, it can be observed that the latent variable in Item 1 has a wider range of values of location for each answer that can be chosen. This implies that small differences in the latent variable will not result in different response options. Thus, Item 1 is very discriminative and can provide accurate information on the actual value of the latent variable (the anxiety factor). Item 12 is the least discriminative item in this factor. Small changes in the value of the latent variable may change the participant's response and therefore provides less valuable information. In the third factor, the small differences between the ICCs of the items indicate a similar capacity to discriminate the changes in the latent variable (depression).

This study presents several strengths. We examined a representative sample of victims of terrorism and their relatives. The traditional analyses of GHQ-28 were enhanced with a politomic model of item response. However, the validity of our results could have been confirmed with other sources, such as a clinical assessment or other screening scales. We had no information of the diagnoses assigned to the participants in the sample; therefore, gender and age effects could not be analyzed regarding the diagnoses of mental disorders. The time gap between the terrorist attacks and the GHQ-28 assessment is another limitation. However, several studies have demonstrated the long-term effects of terrorist attacks on direct victims and their relatives (Baca et al., 2002; DiGrande et al., 2011; L. Ferrando et al., 2011; Kawana et al., 2001; North et al., 2011). It appears symptoms may be more stable 1 year after the traumatic event rather than in the immediate aftermath (Brewin, 2005).

The analysis of the GHQ-28 applied to this sample of victims of terrorism and their relatives demonstrated a high reliability and a good external validity, as well as a three-factor structure. The first factor had the highest loadings in anxiety symptoms, suggesting that anxiety could be the most frequent psychopathological finding in this population. A selection of the most discriminative items, based on the IRT, could be used to improve the screening for detection of mental disorders.

References

Aderibigbe, Y. A., Riley, W., Lewin, T., & Gureje, O. (1996). Factor structure of the 28-item general health questionnaire in a sample of antenatal women. *International Journal of Psychiatry in Medicine, 26,* 263–269. doi:10.2190/3XAV-M1BC-DA2B-DCMF

Baca, E., Cabanas, M. L., & Baca-Garcia, E. (2002). Terrorist attacks and short-long time psychiatric morbidity. *Actas Espanolas de Psiquiatria, 30,* 85–90.

Baca, E., Cabanas, M. L., Perez-Rodriguez, M. M., & Baca-Garcia, E. (2004). Mental disorders in victims of terrorism and their families. *Medicina Clinica, 122,* 681–685.

Baksheev, G. N., Robinson, J., Cosgrave, E. M., Baker, K., & Yung, A. R. (2011). Validity of the 12-item General Health Questionnaire (GHQ-12) in detecting depressive and anxiety disorders among high school students. *Psychiatry Research, 187,* 291–296. doi:10.1016/j.psychres.2010.10.010

Banks, M. H., Clegg, C. W., Jackson, P. R., Kemp, N. J., Stafford, E. M., & Wall, T. D. (1980). The use of the general health questionnaire as indicator of mental health in occupational studies. *Journal of Occupational Psychology, 53,* 187–194. doi:10.1111/j.2044-8325.1980.tb00024.x

Bland, J. M., & Altman, D. G. (1997). Statistics notes: Cronbach's alpha. *British Medical Journal, 314,* 572. doi:10.1136/bmj.314.7080.572

Boscarino, J. A., Figley, C. R., & Adams, R. E. (2004). Compassion fatigue following the September 11 terrorist attacks: A study of secondary trauma among New York City social workers. *International Journal of Emergency Mental Health, 6,* 57–66.

Breslau, N., Kessler, R. C., Chilcoat, H. D., Schultz, L. R., Davis, G. C., & Andreski, P. (1998). Trauma and posttraumatic stress disorder in the community: The 1996 Detroit area survey of trauma. *Archives of General Psychiatry, 55,* 626–632. doi:10.1001/archpsyc.55.7.626

Brewin, C. R. (2005). Systematic review of screening instruments for adults at risk of PTSD. *Journal of Traumatic Stress, 18,* 53–62. doi:10.1002/jts.20007

Brewin, C. R., Fuchkan, N., Huntley, Z., & Scragg, P. (2010). Diagnostic accuracy of the Trauma Screening Questionnaire after the 2005 London bombings. *Journal of Traumatic Stress, 23,* 393–398.

Brewin, C. R., Scragg, P., Robertson, M., Thompson, M., d'Ardenne, P., & Ehlers, A. (2008). Promoting mental health following the London bombings: A screen and treat approach. *Journal of Traumatic Stress, 21,* 3–8. doi:10.1002/jts.20310

Conejo-Galindo, J., Medina, Ó., Fraguas, D., Terán, S., Sainz-Cortón, E., & Arango, C. (2008). Psychopathological sequelae of the 11 March terrorist attacks in Madrid: An epidemiological study of victims treated in a hospital. *European Archives of Psychiatry and Clinical Neuroscience, 258,* 28–34. doi:10.1007/s00406-007-0758-7

Costa, E., Barreto, S. M., Uchoa, E., Firmo, J. O., Lima-Costa, M. F., & Prince, M. (2006). Is the GDS-30 better than the GHQ-12 for screening depression in elderly people in the community? The Bambui Health Aging Study (BHAS). *International Psychogeriatrics, 18,* 493–503. doi:10.1017/S1041610205002954

de Jesus Mari, J., & Williams, P. (1986). Misclassification by psychiatric screening questionnaires. *Journal of Chronic Diseases, 39,* 371–378. doi:10.1016/0021-9681(86)90123-2

DiGrande, L., Neria, Y., Brackbill, R. M., Pulliam, P., & Galea, S. (2011). Long-term posttraumatic stress symptoms among 3,271 civilian survivors of the September 11, 2001, terrorist attacks on the World Trade Center. *American Journal of Epidemiology, 173,* 271–281. doi:10.1093/aje/kwq372

Ferrando, L., Galea, S., Sainz Cortón, E., Mingote, C., García Camba, E., Fernandez Líria, A., & Gabriel, R. (2011). Long-term psychopathology changes among the injured and members of the community after a massive terrorist attack. *European Psychiatry, 26,* 513–517. doi:10.1016/j.eurpsy.2010.07.009

Ferrando, P. J. (2001). The measurement of neuroticism using MMQ, MPI, EPI and EPQ items: A psychometrical analysis based on item response theory. *Personality and Individual Differences, 30,* 641–656. doi:10.1016/S0191-8869(00)00062-3

Gibbons, P. F., Flores de Arévalo, H., & Mónico, M. (2004). Assessment of the factor structure and reliability of the 28 item version of the General Health Questionnaire (GHQ-28) in El Salvador. *International Journal of Clinical and Health Psychology, 4,* 389–398.

Goldberg, D. P., & Hillier, V. F. (1979). A scaled version of the General Health Questionnaire. *Psychological Medicine, 9,* 139–145. doi:10.1017/S0033291700021644

Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods, 7,* 191–205. doi:10.1177/1094428104263675

Henriksen, C. A., Bolton, J. M., & Sareen, J. (2010). The psychological impact of terrorist attacks: Examining a dose-response relationship between exposure to 9/11 and Axis I mental disorders. *Depression and Anxiety, 27,* 993–1000. doi:10.1002/da.20742

Hobbs, P., Ballinger, C. B., Greenwood, C., Martin, B., & McClure, A. (1984). Factor analysis and validation of the general health questionnaire in men: A general practice survey. *British Journal of Psychiatry, 144,* 270–275. doi:10.1192/bjp.144.3.270

Hoven, C. W., Duarte, C. S., Wu, P., Doan, T., Singh, N., Mandell, D. J., . . . Cohen, P. (2009). Parental exposure to mass violence and child mental health: The First Responder and WTC Evacuee Study. *Clinical Child and Family Psychology Review, 12,* 95–112. doi:10.1007/s10567-009-0047-2

Huppert, F. A., Walters, D. E., Day, N. E., & Elliott, B. J. (1989). The factor structure of the General Health Questionnaire (GHQ-30). A reliability study on 6317 community residents. *British Journal of Psychiatry, 155,* 178–185. doi:10.1192/bjp.155.2.178

Kawana, N., Ishimatsu, S., & Kanda, K. (2001). Psycho-physiological effects of the terrorist sarin attack on the Tokyo subway system. *Military Medicine, 166,* 23–26.

Kessler, R. C., Sonnega, A., Bromet, E., Hughes, M., & Nelson, C. B. (1995). Posttraumatic stress disorder in the National Comorbidity Survey. *Archives of General Psychiatry, 52,* 1048–1060. doi:10.1001/archpsyc.1995.03950240066012

Lobo, A., & Muñoz, P. E. (1996). *Versiones en lengua española validadas. Cuestionario de Salud General GHQ (General Health Questionnaire). Guía para el usuario de las distintas versiones* [Validated versions in Spanish language. General Health Questionnaire. A user's guide for the different versions]. Barcelona, Spain: Masson.

National Counter Terrorism Center. (2011). *2010 Report on Terrorism.* Retrieved from www.nctc.gov

North, C. S., Pfefferbaum, B., Kawasaki, A., Lee, S., & Spitznagel, E. L. (2011). Psychosocial adjustment of directly exposed survivors 7 years after the Oklahoma City bombing. *Comprehensive Psychiatry, 52,* 1–8. doi:10.1016/j.comppsych.2010.04.003

Pfefferbaum, B., North, C. S., Flynn, B. W., Norris, F. H., & DeMartino, R. (2002). Disaster mental health services following the 1995 Oklahoma City bombing: Modifying approaches to address terrorism. *CNS Spectrums, 7,* 575–579.

Ploubidis, G. B., & Frangou, S. (2011). Neuroticism and psychological distress: To what extent is their association due to person-environment correlation? *European Psychiatry, 26,* 1–5. doi:10.1016/j.eurpsy.2009.11.003

Politi, P. L., Piccinelli, M., & Wilkinson, G. (1994). Reliability, validity and factor structure of the 12-item General Health Questionnaire among young males in Italy. *Acta Psychiatrica Scandinavica, 90,* 432–437. doi:10.1111/j.1600-0447.1994.tb01620.x

Quek, K. F., Low, W. Y., Razack, A. H., & Loh, C. S. (2001). Reliability and validity of the General Health Questionnaire (GHQ-12) among urological patients: A Malaysian study. *Psychiatry and Clinical Neurosciences, 55,* 509–513. doi:10.1046/j.1440-1819.2001.00897.x

Quek, K. F., Low, W. Y., Razack, A. H., Loh, C. S., & Chua, C. B. (2004). Reliability and validity of the Spielberger State-Trait Anxiety Inventory (STAI) among urological patients: A Malaysian study. *Medical Journal of Malaysia, 59,* 258–267.

Richard, C., Lussier, M. T., Gagnon, R., & Lamarche, L. (2004). GHQ-28 and cGHQ-28: Implications of two scoring methods for the GHQ in a primary care setting. *Social Psychiatry and Psychiatric Epidemiology, 39,* 235–243. doi:10.1007/s00127-004-0710-3

Romans-Clarkson, S. E., Walton, V. A., Herbison, G. P., & Mullen, P. E. (1989). Validity of the GHQ-28 in New Zealand women. *Australian and New Zealand Journal of Psychiatry, 23,* 187–196. doi:10.3109/00048678909062135

Romera, I., Fernandez-Perez, S., Montejo, A. L., Caballero, F., Caballero, L., Arbesu, J. A., . . . Gilaberte, I. (2010). Generalized anxiety disorder, with or without co-morbid major depressive disorder, in primary care: Prevalence of painful somatic symptoms, functioning and health status. *Journal of Affective Disorders, 127,* 160–168. doi:10.1016/j.jad.2010.05.009

Salguero, J. M., Cano-Vindel, A., Iruarrizaga, I., Fernandez-Berrocal, P., & Galea, S. (2011). Trajectory and predictors of depression in a 12-month prospective study after the Madrid March 11 terrorist attacks. *Journal of Psychiatric Research, 45,* 1395–1403. doi:10.1016/j.jpsychires.2011.05.012

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement, 34*(4, Pt. 2).

Sánchez-López, M. P., & Dresch, V. (2008). The 12-Item General Health Questionnaire (GHQ-12): Reliability, external validity and factor structure in the Spanish population. *Psicothema, 20,* 839–843.

Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. (1970). *Manual for the State-Trait Inventory.* Palo Alto, CA: Consulting Psychologists Press.

Stoddard, F. J., Jr., Gold, J., Henderson, S. W., Merlino, J. P., Norwood, A., Post, J. M., . . . Katz, C. L. (2011). Psychiatry and terrorism. *Journal of Nervous and Mental Disease, 199,* 537–543. doi:10.1097/NMD.0b013e318225ee90

Tarnopolsky, A., Hand, D. J., McLean, E. K., Roberts, H., & Wiggins, R. D. (1979). Validity and uses of a screening questionnaire (GHQ) in the community. *British Journal of Psychiatry, 134,* 508–515. doi:10.1192/bjp.134.5.508

Thomas, M. L. (2011). The value of item response theory in clinical assessment: A review. *Assessment, 18,* 291–307. doi:10.1177/1073191110374797

Tran, T. D., Tran, T., & Fisher, J. (2012). Validation of three psychometric instruments for screening for perinatal common mental disorders in men in the north of Vietnam. *Journal of Affective Disorders, 136,* 104–109. doi:10.1016/j.jad.2011.08.012

Ursano, R. J., & Friedman, M. J. (2006). Mental health and behavioral interventions for victims of disasters and mass violence systems, caring, planning, and needs. In E. C. Ritchie, M. J. Friedman, & P. J. Watson (Eds.), *Interventions following mass violence and disasters: Strategies for mental health practice* (pp. 405–414). New York, NY: Guilford Press.

Vallejo, M. A., Jordan, C. M., Diaz, M. I., Comeche, M. I., & Ortega, J. (2007). Psychological assessment via the Internet: A reliability and validity study of online (vs paper-and-pencil) versions of the General Health Questionnaire-28 (GHQ-28) and the Symptoms Check-List-90-Revised (SCL-90-R). *Journal of Medicine Internet Research, 9,* e2. doi:10.2196/jmir.9.1.e2

van der Linden, W. J., & Hambleton, R. K. (1997). *Handbook of modern item response theory.* New York, NY: Springer.

Werneke, U., Goldberg, D. P., Yalcin, I., & Ustun, B. T. (2000). The stability of the factor structure of the General Health Questionnaire. *Psychological Medicine, 30,* 823–829. doi:10.1017/S0033291799002287

Zhang, J., & Gao, Q. (2012). Validation of the trait anxiety scale for state-trait anxiety inventory in suicide victims and living controls of Chinese rural youths. *Archives of Suicide Research, 16,* 85–94. doi:10.1080/13811118.2012.641440

Zhu, C., Ou, L., Geng, Q., Zhang, M., Ye, R., Chen, J., & Jiang, W. (2012). Association of somatic symptoms with depression and anxiety in clinical patients of general hospitals in Guangzhou, China. *General Hospital Psychiatry, 34,* 113–120. doi:10.1016/j.genhosppsych.2011.09.005

Received March 23, 2012
Revision received September 14, 2012
Accepted September 17, 2012 ■

3.2 Identificación individual mediante rasgos de personalidad

Journal of Network and Computer Applications 33 (2010) 293–299

Contents lists available at ScienceDirect

Journal of Network and Computer Applications

journal homepage: www.elsevier.com/locate/jnca

Individual identification using personality traits

David Delgado-Gómez [a,*], Federico Sukno [c], David Aguado [b], Carlos Santacruz [b], Antonio Artés-Rodriguez [a]

[a] Department of Signal Theory and Communications, Carlos III University, Madrid, Spain
[b] Institute of Knowledge Engineering, Autonomous University, Madrid, Spain
[c] Computational Imaging Lab, Pompeu Fabra University, Barcelona, Spain

ARTICLE INFO

Article history:
Received 24 July 2009
Received in revised form
15 November 2009
Accepted 10 December 2009

Keywords:
Biometrics
Psychometrics
Personality traits
Samejima's model

ABSTRACT

In this article, a pioneer study is conducted to evaluate the possibility of identifying people through their personality traits. The study is conducted using the answers of a population of 734 individuals to a collection of 206 items. These items aim at measuring five common different personality traits usually called the big five. These five levels are neuroticism, extraversion, agreeableness, conscientiousness and openness. The traits are estimated using the widely used Samejima's model and then used to discriminate the individuals. Results point biometrics using personality traits as a new promising biometric modality.

1. Introduction

A biometric system is essentially a pattern recognition system that performs recognition based on some features derived from measurements of physiological or behavioral characteristics that an individual possesses (Prabhakar et al., 2003). Biometrics emerged initially from its extensive use in law enforcement and received further impulse due to the threatens to security derived from terrorist attacks in the last few years.[1] Nonetheless, biometrics is being increasingly used today to carry out person identification in a number of civilian applications, such as national ID documents, smart cards and so on (Wayman et al., 2005). The proliferation of online banks and credit card services has further increased the need for reliable identity management systems (Jain, 2007). According to a study from the International Biometric Group (The International Biometric Group, 2007), total revenues from biometrics were about $1200 millions in 2003, and projections suggest that by 2010 the total biometric market will have grown more than three times, reaching $3800 millions.

The most attractive feature of biometrics is that recognition is based on *who you are*, as opposed to the traditional recognition methods based on *what you know* (passwords) or *what you have* (tokens such as ID cards) (Jain, 2007). Several measurements have been proposed as inputs for biometric systems and they determine the biometric modality. Among the most traditional modalities we can find face (Kong et al., 2004; Dai and Yuen, 2007; Masip et al., 2009; Abate et al., 2007), fingerprint (Maltoni et al., 2003), signature (Plamondon and Srihari, 2000), speech (Reynolds et al., 2000), palm (Han et al., 2003) or iris (Daugman, 2006); while ear (Chang et al., 2003), vessel (Badawi, 2006), gait (Sarkar et al., 2005) or even teeth biometrics (Jain et al., 2003) are relatively more recent. Each modality has its strengths and weaknesses (depending on the aspect that is analyzed), and the choice depends on the requirements for each particular application (Veeramachaneni et al., 2005).

Biometric systems have shown good performance in certain scenarios and tasks such as controlling access to restricted areas at airports or database access and computer login. However, some research projects have shown several ways of breaking into these systems. As Ratha et al. (2003) exposes: "Overall, the most serious threat to biometric authentication systems is presenting, either physically or electronically, fake biometrics or previously acquired biometric templates". For instance, Matsumoto et al. (2002) demonstrated that it is feasible to create a fake fingerprint from silicon and gelatin, good enough to fool several commercial systems. Biometric templates security is an important issue because, unlike passwords and tokens, they are not easy to revoke. Although efforts have been done to design protection schemes (Teoh et al., 2006; Ratha et al., 2007), current mechanisms are yet not considered satisfactory enough (Jain, 2007).

In this article, a new biometric modality is considered: our personality. For decades, psychometrics has investigated models

* Corresponding author.
E-mail address: ddelgado@tsc.uc3m.es (D. Delgado-Gómez).

[1] The New York Times 24-08-2003: The Art and Craft of Security: Passports and Visas to Add High-Tech Identity Features.

1084-8045/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jnca.2009.12.009

that obtain precise measures of certain personality traits such as extraversion or agreeableness. In order to measure these traits, psychometric researchers assume that the answer of an individual to a given question depends on some item parameters and on the personality level of the individual. The difficulty of identifying an individual by his/her personality trait is the low discrimination that is obtained if a small number of questions are presented to the individual. Usually, these traits are measured using items which have between two and six possible different answers. This implies that there are only 216 different patterns, if for instance, three items with six different possibilities are presented to an individual. This number is too small if it is compared with the 2^{2048} possible pattern that an iris codification can take (Daugman, 2006). Based on these numbers, 2^{792} questions should be formulated to an individual in order to obtain a response pattern with a dimension comparable to an iris code. However, psychometric researchers have addressed this inconvenient by developing techniques that allow to determine which are the most informative items for a given ability level. This allows for a faster convergence of the estimated trait by adapting the questions that are presented to the real individual trait. On the other hand, if we were able to identify a person by some of his/her personality traits, it might be possible to develop a more secure biometric system. How could an impostor be capable of guessing what would be our answer to some previously unformulated questions?

In order to be viable, a biometric modality has to fulfill a number of requirements as it has been indicated by Ross et al. (2006). Following, the accomplishment of these requirements by the proposed modality is analyzed:

- *Universality.* Every person should have the (measured) characteristic. Clearly, every individual has several personality traits.
- *Uniqueness.* Two different individuals should not generate the same features. Each of the individual's personality traits is assumed to follow a standard Gaussian distribution. In this article, five personality traits are used. Therefore, each individual is described by a five dimensional vector obtained from a five dimensional Gaussian distribution. This suggests that each individual's personality representation should be sufficiently different across individuals.
- *Permanence.* The measured characteristic should be invariant with time. There are several studies which have provided evidence of the longitudinal robustness of personality traits (Block, 1981). For instance, Conley (1985) conducted a study which showed the stability of the traits over a 19-year period.
- *Measurability.* The characteristic should be (quantitatively) measurable. The proposed modality is based on the item response theory (IRT). Each individual is represented by a set of underlying traits. The trait estimations are independent of the administered test, which means that different tests provide similar traits. This is one of the advances of IRT with respect to the classical test theory in which different tests provided different scores. Moreover, the fact that the items are sequentially selected by adapting them to the examinee's trait provides a faster convergence to the real trait.
- *Performance.* This factor refers to the recognition accuracy and the resources required. The analysis of the accuracy of the proposed system is the objective of the manuscript and it is considered in the experimental section. The required computation can be achieved in real time.
- *Acceptability.* It refers to the fact that the population would be willing to present their biometric trait to the system. Compared to biometric modalities such as iris or fingerprint which require a close interaction between the individual and the system, clicking in a tactile screen the answers to some questions should not present an inconvenient.
- *Circumvention.* According to Ross et al. (2006), this factor refers the ease with which the trait of an individual can be imitated using artifacts. In order to get access into the system, the person has to answer some previously unformulated questions. The answers to these questions depend on two elements: the individual's trait level and the item parameters. So if an intruder wants to get access into the system by mimicking the individual's trait, the intruder would need to know both elements.

The set of possible scenarios where the proposed biometric technique could be applied is not restricted to security applications. For instance, nowadays the first stage in the candidate selection process for recruitment of many companies is handled through online tests (Bartram, 2002). These online tests allow to reduce the number of candidates to a subset of suitable candidates on which the company can spend more time using more traditional methods (e.g. interviews). By means of personality biometrics, companies could verify if these pre-selected candidates were actually the individuals who answered the online test.

The remainder of this article is structured as follows. Section 2 reviews some of the main elements of the item response theory. The Samejima model developed to measure personality traits and the maximum information item selection criterion are described. Section 3 shows the results obtained in an experiment which aims at testing the performance of personality traits as biometric measurements. The article concludes in Section 4 with a discussion of the obtained results.

2. Item response theory

Item response theory is based on the assumption that the answer given by an individual to a specific item depends on the individual's trait and on the properties of the item.[2] Specifically, let i denote an item of an item bank and let n_i be the number of possible choices for i-th item. Let K be the number of examinees. The probability that the k-th examinee with a given trait value θ^k chooses answer $x_i^k \in \{0,\ldots,n_i-1\}$ on item i is represented by $P(x_i^k|\theta^k)$. Different models have been developed in order to define this probability (van der Linden and Hambleton, 1996). In this work Samejima's graded response model (Samejima, 1969) is used.

2.1. The graded response model

Samejima (1969) developed a model to calculate the probabilities of answering each of the item choices given the trait level θ^k of the examinee. This model is specially suitable when the items measure to attitudes or personality traits tests with two or more discrete graded answers.The probability that the k-th individual will choose the answer x_i^k or higher on item i is expressed by

$$P^*\left(x_i^k|\theta^k\right) = \frac{\mathrm{Exp}\left\{D \cdot a_i^k \cdot \left[\theta^k - b_i^k(x_i^k)\right]\right\}}{1+\mathrm{Exp}\left\{D \cdot a_i^k \cdot \left[\theta^k - b_i^k(x_i^k)\right]\right\}} \tag{1}$$

[2] Moreover, the tests are built so that all the items measure the same underlying trait and so that the answers to the different items are independent.

where D is a scaling constant equal to 1.7, a_i^k and $b_i^k(x_i^k)$ are the i-th item response parameter satisfying $b_i^k(x_i^k+1) > b_i^k(x_i^k)$.

The probability $P(x_i^k|\theta^k)$ is therefore computed as

$$P(x_i^k|\theta^k) = P^*(x_i^k|\theta^k) - P^*(x_i^k+1|\theta^k) \quad (2)$$

This formula requires the probabilities when $x_i^k = 0$ and $x_i^k = n_i$ for all k:

$$P^*(x_i^k = 0|\theta^k) = 1, \quad P^*(x_i^k = n_i|\theta^k) = 0 \quad (3)$$

Therefore, this model has n_i parameters for item i and examinee k: n_i-1 of them are the $b_i(x_i^k)$ for $x_i^k = 1,\ldots,n_i-1$ and one is the parameter a_i^k.

2.2. Item parameter estimation

In order to estimate the parameters of each item, the response patterns of a sample population of K examinees is used. A k-th examinee response pattern is defined as a sequence of item responses, $V_m^k = (x_{i_1}^k, x_{i_2}^k, \ldots, x_{i_m}^k)$, where m is the total number of answered items and i_j represents the j-th item presented to the examinee k. The probability of obtaining that response pattern when the k-th examinee has a trait of θ^k is

$$P(V^k|\theta^k) = \prod_{j=1}^{m} P(x_{i_j}^k|\theta^k) \quad (4)$$

The probability of obtaining the set of all response patterns $\mathbf{V} = \{V^k|k=1,\ldots,K\}$ for all the examinees with the corresponding trait values $\Theta = (\theta^1,\ldots,\theta^K)$ is

$$P(\mathbf{V}|\mathbf{\Theta}) = \prod_{k=1}^{K} P(V^k|\theta^k) = \prod_{k=1}^{K}\prod_{j=1}^{m} P(x_{i_j}^k|\theta^k) \quad (5)$$

The probability $P(V^k)$ for a given response pattern V^k is determined from Eq. (4) as

$$P(V^k) = \int P(V^k|\theta^l)P(\theta^l)\partial\theta \quad (6)$$

where θ^l is assumed to be normally distributed.

The marginal probability of $\mathbf{V}$ is therefore given by

$$P(\mathbf{V}) = \prod_{k=1}^{K} P(V^k) = \prod_{k=1}^{K}\left(\int\left[P(\theta^l)\cdot\prod_{j=1}^{m} P(x_{i_j}^k|\theta^l)\right]\partial\theta\right) \quad (7)$$

which depends only on the set of all item parameters Λ:

$$\Lambda = \{a_{i_j}^k, b_{i_j}^k(1),\ldots,b_{i_j}^k(n_{i_j}-1)|k=1,\ldots,K, j=1,\ldots,m\} \quad (8)$$

Let us define a likelihood function as follows:

$$L_{\mathbf{V}}(\Lambda) = P(\mathbf{V}) \quad (9)$$

Bock and Aitkin (1981) developed an expectation maximization procedure for estimating Λ from (9). However, any general purpose optimization algorithm can be used to solve this problem. In our work, Lagarias's algorithm (Lagarias et al., 1998) for solving nonlinear problems with simple boundary constrains is used. It provides the solution Λ for the following problem:

$$\arg\min_{\Lambda}\{-\log[L_{\mathbf{V}}(\Lambda)]\} \quad (10)$$

subject to $b_{i_j}^k(s+1) > b_{i_j}^k(s)$ for $j=1,\ldots,m$, $k=1,\ldots,K$, and $s=1,\ldots,n_{i_j}-1$.

2.3. Trait estimation

Once the item parameters Λ are known, it is possible to estimate the trait level of each examinee based on his response pattern. Let $V_{m'}^k = (x_{i_1}^k, x_{i_2}^k,\ldots,x_{i_{m'}}^k)$ represent the response pattern of k-th examinee after answering m' items, where $m' \le m$. The k-th examinee's trait after presenting m' items, $\hat{\theta}_{m'}^k$, is estimated maximizing the probability of obtaining $V_{m'}^k$

$$\hat{\theta}_{m'}^k = \arg\max_{\theta}\{L(\theta|V_{m'}^k)\} \quad (11)$$

where

$$L(\theta|V_{m'}^k) = \prod_{j=1}^{m'} P(x_{i_j}^k|) \quad (12)$$

2.4. Maximum information

Once the trait $\hat{\theta}_{m'}^k$ has been estimated for the k-th examinee, the next item to be presented is selected by a maximum information based procedure among the unused items. The selected item maximizes the item information function defined by

$$I_i(\hat{\theta}_{m'}^k) = -\sum_{x=0}^{n_i-1}\left(\frac{\partial^2}{\partial\theta^2}\log P(x_i^k|\hat{\theta}_{m'}^k)\right)P(x_i^k|\hat{\theta}_{m'}^k) \quad (13)$$

It can be shown Muraki and Bock (1977) that the item information function for the graded response model can be expressed by

$$I_i(\theta^k) = \sum_{x=0}^{n_i-1}\frac{(a_i^k)^2}{P(x_i^k|\theta^k)}[P^*(x_i^k|\theta^k)[1-P^*(x_i^k|\theta^k)] - P^*(x_i^k+1|\theta^k)[1-P^*(x_i^k+1|\theta^k)]]^2 \quad (14)$$

where $P(x_i^k)$ represents the previously defined probability that the examinee k answers the category x in item i.

Observe that, for the k-th examinee

$$I(\theta^k) = \sum_{i=1}^{I} I_i(\theta^k) \quad (15)$$

which is inversely proportional to the error of the maximum likelihood estimator $\hat{\theta}^k$. That is, the asymptotic variance of $\hat{\theta}^k$ is the reciprocal of $I(\theta^k)$ where θ^k is the true value of $\hat{\theta}^k$.

3. Experimental results

In this section, the performance of the proposed biometric modality in both identification and verification experiments is evaluated. Identity verification (or "one-to-one") refers to a positive outcome of a specific identity. That is, during the verification process, the user requires providing a unique user ID (e.g. name) and the matching algorithm will compare his/her biometric template (in our case the personality trait) with the one it has on records (created during enrollment). The result is either "match" or "no-match".

Identification (or "one-to-many") refers to a positive outcome of identity within a predefined identities group (the *users*). Hence, the user is not required to provide a unique user ID, but to provide just a biometric sample. The matching algorithm will compare this template with each one of the enrolled users. The result is the identity of the best-match according to the template, sometimes allowing also for a negative outcome, which indicates that the person does not in the users group (closed- vs. open-set experiments).

Both experiments are conducted using the same database. This database was collected by the group of psychologists of the Institute of Knowledge Engineering.[3] It contains the answers of

[3] www.iic.uam.es.

734 individuals to a questionnaire of 206 items.[4] The items of the questionnaire aimed at measuring five different personality traits which, in the psychological field, are commonly known as the big five (Barrick and Mount, 1991). These traits are neuroticism (50 items), extraversion (50 items), agreeableness (49 items), conscientiousness (29 items) and openness (28 items). Following, these personality traits are briefly described:

- *Neuroticism.* This trait is frequently associated with impulsiveness, vulnerability, anxiety, depression and hostility (McCrae and Costa, 1999). One of the items used in the experiment to measure this trait was "I become exasperated when I listen a door squeaks".
- *Agreeableness.* According to McCrae and Costa (1999), this factor is defined as the degree in which people show themselves sociable, actives and talkative. Their model for measuring this trait includes items which aim at measuring gregariousness, cordiality, or positive emotions. One of the items related to this trait used in the experiment was: "I prefer team sports than individual sports".
- *Conscientiousness.* People with this trait high are usually considered collaborative, tolerant, empathetic and humble (Barrick and Mount, 1991). One of the items used in the experiment to measure this trait was: "I consider that I am not smarter than the other people".
- *Responsibility.* This personality trait measures how much effort a person puts in accomplishing the tasks he/she has committed to. Usually people who obtain a high score in this trait are considered well organized, competent and self-disciplined (Hogan and Ones, 1997). An example of an item used to measure this trait was: "I am an organized person".
- *Openness.* People who obtain a high score in this trait are usually described as imaginative, sensitive and open to new experiences and ideas (McCrae and John, 1992). An example of an item used for measuring this trait was: "I am not interested in art".

3.1. Experiment 1: identification

This first experiment is similar to the one conducted by Chang et al. (2003) in which a new biometric modality (ear) was presented. We constructed two sets: the first set (646 people) was used to estimate the item parameters. The second set (88 people) was used to measure the recognition performance when personality traits are used. The size of this second set was chosen to match the size of the probe set used by Chang et al. (2003) so that results can be compared. The answers and the corresponding items for the second set were randomly divided into gallery and probe sets. Each set included half the items and the corresponding answers for each of the five traits.[5] The items and answers of the gallery set were used to obtain the ability levels for each of the five personality traits (for each individual). The answers and items in the probe set were used to evaluate the accuracy of identifying these individuals by their answers to the second independent group of questions.

The first issue that was analyzed was the accuracy that can be achieved using personality traits, and its dependency upon the number of questions presented to the individuals. For all traits to have the same number of items, no more than 14 can be selected, corresponding to the items for conscientiousness and openness traits in the probe set. The cumulative match characteristic (CMC) curves (Chang et al., 2003) for the probe set depending on the number of items presented to the individuals are displayed in Fig. 1. The curves correspond to each of the five personality traits and their combination (by concatenating the traits). The comparisons were based on the Euclidean distance using the graded response model (Samejima, 1969; Hopfield, 1984), as each item allowed for six different answers gradually covering the whole range between antagonist answers.[6] The items were adaptively selected using the previously described maximum information criterion. From this figure, it is observed that all traits have a poor recognition performance when they are individually used.

However, when they are combined, the recognition performance improves considerably. For instance, if three items per trait are used (15 items in total), the combined CMC curve is better than any of the (individual) CMC curves when all the available items are considered. Another relevant aspect is the difference in the individual curves. According to the curves, the neuroticism trait is more suitable for biometric identification than the openness trait. Moreover, it is also observed that, after nine items are presented, only a slight improvement is achieved when more items are presented. This fact is displayed in the last plot of the figure.

It was also analyzed if an adaptive item selection improves the performance of a traditional test where all examinees receive the same questions. To this end, 30 different traditional tests were randomly created. These traditional tests will be referred to as fix tests from now on. They were built by randomly selecting items from the probe set until completing the chosen test length for each of the considered traits. The answers of the 88 examinees to these randomly selected items were used to estimate the individuals' trait levels. Fig. 2 displays the CMC curve for the adaptive test and the average CMC curve for the fix tests together with a 95% confidence interval. From this figure, we noticed that the improvement achieved when the questions are adaptively presented is statistically significant.

Finally, the obtained results are compared with the results presented by Chang et al. in Fig. 3. In spite of the lower number of features of personality traits when compared to face or ear, the three biometric modalities obtain similar CMC curves. Although the performance of personality traits is lower than face and ear for the first and second rank, it performs comparably for ranks 3–5 and it is the best of the tree afterwards.

3.2. Experiment 2: verification

The goal of this second experiment was to analyze the accuracy that can be achieved with the proposed modality in a verification experiment. To this end, individuals from the database were randomly selected to construct three sets: *users*, *validation impostors* and *test impostors*, containing 200, 70 and 25 people respectively. In this way, the group sizes match the well known Lausanne protocol for identity verification (Messer et al., 1999). The remaining 439 people were used to estimate item parameters (before the verification experiment itself).

The number of items used to estimate each of the five personality traits is displayed in Table 1. As the objective of this article was to determine if personality traits can be used for

[4] The data used in this article are available for research purposes by contacting the authors.

[5] The gallery set had one item more than the probe set for the agreeableness and conscientiousness traits.

[6] For example, the item: "I am an organized person" allows for six different answers, ranging from complete disagreement to total agreement. This allows for a calculating a "distance" between different answers.

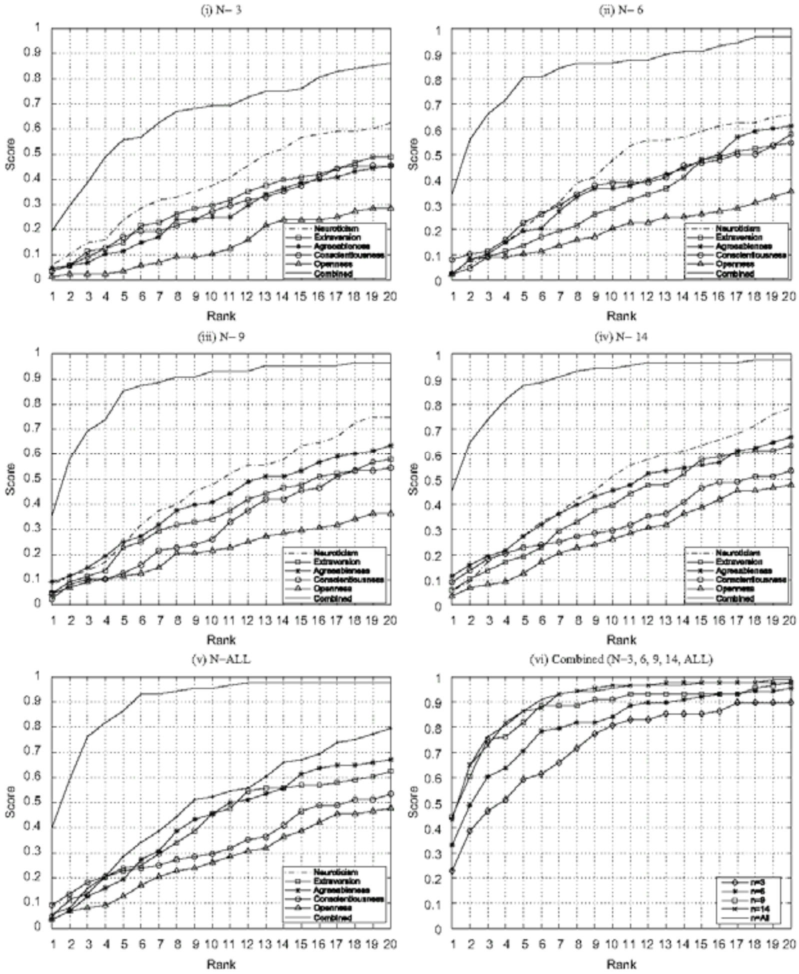


Fig. 1. (i)–(v): Cumulative match characteristic curves depending on the number N of presented items for the five considered traits and the vector composed of them. (vi) Cumulative match characteristic curves for the vector composed of the five considered traits depending on the number N of presented items. (i) $N=3$; (ii) $N=6$; (iii) $N=9$; (iv) $N=14$; (v) $N=$ ALL; (vi) Combined ($N=3$, 6, 9, 14, ALL).

verification tasks, only a reduced number of items were collected. This fact implies that the obtained accuracy is a lower-bound for the proposed modality. The reasons for a potential increase in accuracy are twofold. Firstly, in our experiment each examinee receives exactly the same items to estimate his/her trait in the training, validation and test sets. There is not adaptive item selection. This means that, although the examinee is answering n items, only m items ($m \leq n$) are informative for his/her trait. Therefore, the accuracy in the estimation is much lower than the one obtained if adaptive selection would have been used, as shown in Experiment 1 and in agreement with other results in the literature. Secondly, in a validation experiment, several scores are obtained in each of the three sets per individual to capture the variability of the measurements. However, due to the limited number of items available in our dataset, only one set of answers per individual and per task is used. Evidently, this has a negative effect in the performance of the classifiers, as there is no information about the variability of the measurements for a given individual.

Given a personality trait, the *user scores* for evaluation are defined as the absolute distances between the estimated trait of each evaluation user and the estimated trait of the corresponding training user. Similarly, the *impostor scores* for evaluation are defined as the absolute distances between the estimated trait of each evaluation impostor and each training user. By thresholding these scores to determine 'match" or "no-match" we obtain the false rejection rate (FRR) and false acceptance rate (FAR), respectively. The point where the FAR and FRR match is the equal error rate (EER) and determines the threshold to be used for the test set, where the FAR and FRR will (in general) differ and it is common practice to report their average, known as half total error rate (HTER).

The obtained EER and HTER for each individual trait are displayed in Table 2. This table also includes the results obtained

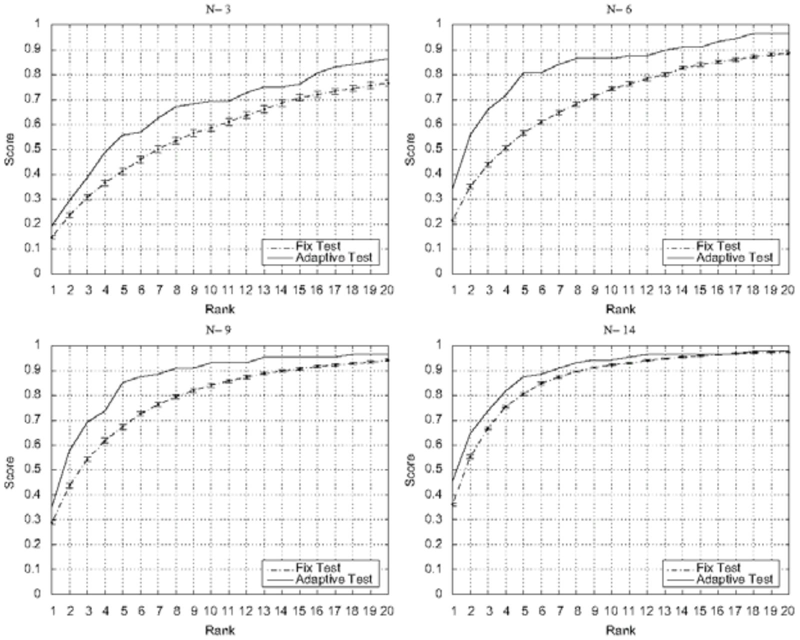


Fig. 2. Cumulative match characteristic curves for the vector composed of the five considered traits depending on the number N of presented items when an adaptive test and a traditional test (fix test) are conducted. The bars represent a 95% confident interval. $N=3$; $N=6$; $N=9$; $N=14$.

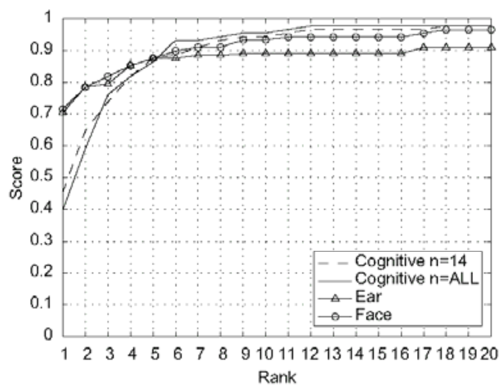


Fig. 3. Face, ear and personality performance recognition.

Table 1
Number of items used to estimate the individual's traits in the training, validation and test sets.

Personality	Training	Validation	Test
Neuroticism	20	15	15
Agreeableness	20	15	15
Conscientiousness	21	14	14
Responsibility	11	9	9
Openness	13	5	5

Table 2
Verification results.

Personality trait	Equal error rate (evaluation)	Half total error rate (test)
Neuroticism	26.5	27.02 ± 3.14
Agreeableness	32	35.8 ± 3.5
Conscientiousness	34	33.2 ± 3.33
Responsibility	34.5	36 ± 3.3
Openness	37.5	41.5 ± 3.5
All combined	20.5	22 ± 2.9
All combined with SVM	15.5	19.21 ± 2.72

when the five personality traits are jointly used by means of a direct concatenation of the feature vectors and when they are combined with a support vector machine classifier. The 95% confidence intervals are also included.

4. Conclusion

In this article, personality traits have been proposed as biometrical measurements. These traits can be estimated from the answers of an individual to certain questions developed for psychological tests. As a result, the proposed modality works with features of much lower dimensionality than other biometric modalities. For example, while iris deals with patterns of 2^{2048} elements, biometrics using personality traits have patterns of m^n elements, where m is the number of possible answers and n is the number of formulated questions. However, it has experimentally been shown that biometrics using personality traits can achieve results comparable to other modalities, such as face and ear.

Additionally, it has been shown that, if the questions are adaptively presented to the individuals, the discrimination is increased. The reason for this is that trait estimations in adaptive tests have a faster convergence to the individual's real trait level than in traditional tests. Adaptive tests avoid presenting too easy or too difficult questions for the individual, which would provide no information about his/her trait level.

An important issue for the practical implementation of personality trait based biometrics is the need to present around 45 questions to an individual. This process usually takes between 2 and 3 min in a computerized adaptive test. As a counterpart, this type of identification could be very promising against forgery: as the questions are not always the same, a potential impostor should learn how the authentic user *thinks* to be able to fake his/her answers.

Finally, the fact that there is a noticeable difference in the recognition performance of the individual cognitive traits indicates that it might be possible to reduce the number of questions if more discriminative traits are considered. For instance, general intelligence or risk aversion might be investigated. Classification models or item selection criteria are other lines for potential improvement.

References

Abate AF, Nappi M, Riccio D, Sabatino G. 2D and 3D face recognition: a survey. Pattern Recognition Letters 2007;28:1885–906.

Badawi AM. Hand vein biometric verification prototype: a testing performance and patterns similarity. In: Proceedings of the 2006 international conference on image processing, computer vision and pattern recognition: 2006. p. 3–9.

Barrick MR, Mount MK. The big five personality dimension and job performance: a meta-analysis. Personnel Psychology 1991;44:1–26.

Bartram D. Internet recruitment and selection: kissing frogs to find princes. International Journal of Selection and Assesment 2002;8:261–74.

Block J. Some enduring and consequential structure of personality. In: Further explorations in personality. New York: Wiley; 1981. p. 27–43.

Bock RD, Aitkin M. Marginal maximum likelihood estimation of item parameters: application of an EM algorithm. Psychometrika 1981;46:443–59.

Chang K, Bowyer KW, Sarkar S, Victor B. Comparison and combination of ear and face images in appearance-based biometrics. IEEE Transactions on Pattern Analysis and Machine Intelligence 2003;25:1160–5.

Conley JJ. Longitudinal stability of personality traits: a multitrait-multimethod-multioccasion analysis. Journal of Personality and Social Psychology 1985;49: 1266–82.

Dai DQ, Yuen PC. Face recognition by regularized discriminant analysis. IEEE Transactions on Systems, Man, and Cybernetics B, Cybernetics 2007;4(37): 1080–5.

Daugman J. Probing the uniqueness and randomness of IrisCodes: results from 200 billion iris pair comparisons. Proceedings of the IEEE 2006;94:1927–35.

Han CC, Cheng HL, Linb CL, Fanb KC. Personal authentication using palm-print features. Pattern Recognition 2003;36:371–81.

Hogan R, Ones J. Concientiousness and integrity at work. In: Handbook of personality psychology. San Diego: Academic Press; 1997. p. 849–70.

Hopfield JJ. Neurons with graded response have collective computational properties like those of two-state neurons. Proceedings of the National Academy of Sciences of the United States of America 1984;81:3088–92.

Jain AK. Biometric recognition: overview and recent advances. In: Lecture notes in computer science: progress in pattern recognition, image analysis and applications, vol. 4756. Berlin: Springer; 2007. p. 13–9.

Jain AK. Biometric recognition. Nature 2007;449:39–40.

Jain AK, Chen H, Minut S. Dental biometrics: human identification using dental radiographs. In: Lecture notes in computer science, vol. 2688. Berlin: Springer; 2003. p. 429–37.

Kong SG, Heo J, Abidi BR, Paik J, Abidi MA. Recent advances in visual and infrared face recognition – a review. Computer Vision and Image Understanding 2004;97:103–35.

Lagarias JC, Reeds JA, Wright MH, Wright PE. Convergence properties of the Nelder–Mead simplex method in low dimensions. SIAM Journal of Optimization 1998;9:112–47.

Maltoni D, Maio D, Jain KA, Prabhakar S. Handbook of fingerprint recognition. Berlin: Springer; 2003.

Masip D, Lapedriza A, Vitria J. Boosted online learning for face recognition. IEEE Transactions on Systems, Man, and Cybernetics B, Cybernetics 2009;39: 530–538.

Matsumoto T, Matsumoto H, Yamada K, Hoshino AK. Impact of artificial gummy fingers on fingerprint systems. Proceedings of the SPIE Optical Security and Counterfeit Deterrence Techniques IV 2002;4677:275–89.

McCrae RR, John OP. The five factor model: issues and applications. Journal of Personality 1992;60:175–215.

McCrae RR, Costa PT. A five factor theory of personality. Handbook of personality theory and research; 1999. p. 139–53.

Messer K, Matas J, Kittler J, Luettin J, Maitre G. XM2VTS: the extended M2VTS database. In: Proceedings of the second international conference on audio and video-based biometric person authentication; 1999. p. 72–7.

Muraki E, Bock RD. Parscale. Item analysis and test scoring for rating-scale data. Chicago, IL: Scientific Software, Inc.; 1977.

Plamondon R, Srihari SN. Online and off-line handwritting recognition: a comprehensive survey. IEEE Transactions on Pattern Analysis and Machine Intelligence 2000;22:63–84.

Prabhakar S, Pankanti S, Jain AK. Biometric recognition: security and privacy concerns. IEEE Security & Privacy 2003;1:33–42.

Ratha NK, Connell JH, Bolle RM. Biometrics break ins and band aids. Pattern Recognition Letters 2003;24:2105–13.

Ratha NK, Chikkerur S, Connell JH, Bolle RM. Generating cancelable fingerprint templates. IEEE Transactions on Pattern Analysis and Machine Intelligence 2007;29:561–72.

Reynolds DA, Quatieri TF, Dunn RB. Speaker verification using adapted gaussian mixture models. Digital Signal Processing 2000;10:19–41.

Ross A, Nandakumar K, Jain AK. Handbook of multibiometrics. Berlin: Springer; 2006. [Section 1.6.].

Sarkar S, Phillips PJ, Liu Z, Vega R, Grother P, Bowyer KW. The humanID gait challenge problem: data sets, performance, and analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 2005;27:162–77.

Samejima F. Estimation of latent ability using a response pattern of graded scores. Psychometrika, Monograph Supplement 17; 1969.

The International Biometric Group. Biometric market and industry report 2006–2010; 2007. 〈www.biometricgroup.com〉.

Teoh ABJ, Goh A, Ngo DCL. Random multispace quantization as an analytic mechanism for biohashing of biometric and random identity inputs. IEEE Transactions on Pattern Analysis and Machine Intelligence 2006;28: 1892–1901.

van der Linden WJ, Hambleton RK. Handbook of modern item response theory. Berlin: Springer; 1996.

Veeramachaneni K, Osadciw LA, Varshney PK. An adaptive multimodal biometric management algorithm. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews 2005;35:344–56.

Wayman JL, Jain AK, Maltoni D, Maio. Biometric systems, technology, design and performance evaluation. Heidelberg: Springer; 2005.

4. Obtención de nuevos marcadores de rasgos de personalidad basados en expresiones faciales y corporales

Antes de entrar en detalle en este capítulo, se mostrará un ejemplo que muestra la importancia de poseer buenas variables predictivas. Este ejemplo ha sido obtenido de la presentación realizada por uno de los ponentes en una de las conferencias denominadas VisionDays, que se celebran anualmente en Lyngby (Dinamarca). Este ponente era uno de los evaluadores de las distintas propuestas de proyectos de investigación en Dinamarca. En esta charla, el ponente expuso sus experiencias en la evaluación de los distintos proyectos. Explicó que años antes, uno de los temas sobre los que más propuestas se recibieron consistía en identificar las distintas células en videos grabados durante el proceso de meiosis. Expuso que la mayoría de los trabajos proponían diversas aproximaciones algorítmicas complejas. Sin embargo, también comentó que se presentó un proyecto en el que se proponía el uso de un líquido que, al ser inyectado antes de obtener las imágenes, aumentaba el contraste con las células. Este líquido hacía innecesarias todas las otras soluciones algorítmicas propuestas. En este ejemplo, se observa que poseer buenos predictores es igual (o más) importante que elegir los clasificadores apropiados.

En Psicología, como se ha podido apreciar en los capítulos anteriores, las respuestas proporcionadas por los participantes a los ítems de los diferentes cuestionarios son sin duda los predictores mayoritariamente usados. Sin embargo, diversos trabajos han mostrado que los cuestionarios presentan diversas limitaciones. En primer lugar, la veracidad de las respuestas no está garantizada. Por ejemplo, se ha observado que el 22% de los adultos con trastorno por déficit de atención e hiperactividad tiende a proporcionar respuestas que sugieren una gravedad mayor que la existente para tener acceso a drogas estimulantes (Marshall, Schroeder, O'Brien, Fischer, Ries, & Blesi, 2010). Como otro ejemplo, podemos mencionar que algunos pacientes ingresados por riesgo suicida tienden a proporcionar respuestas que muestran una mejoría de su estado para ser dados de alta y consumar el suicidio (Simon & Gutheil, 2009). Un segundo inconveniente que ha sido indicado es que los cuestionarios no son adecuados para administrarlos varias veces al mismo examinado. (Chamberlain & Sahakian, 2007). Finalmente, otro inconveniente que ha sido apuntado es que los cuestionarios pueden estar influenciados por la cultura (Moeller, Barratt, Dougherty, Schmitz, & Swann, 2001).

En los últimos años, se han propuesto distintas alternativas a los test tradicionales (tanto en su versión de lápiz y papel como en su versión informatizada). Probablemente, la más utilizada son los test por ordenador en los que los participantes realizan pruebas simples y de duración relativamente corta en las que se obtienen distintos índices como son el número de aciertos/fallos o el tiempo de reacción. Algunos de estos test son los test de rendimiento continuo (denominados CPTs: Continuous Performance Tests) (Conners & Staff, 2000; Nosek & Banaji, 2001; Greenberg, Kindschi, Dupuy, & Hughes, 1999) , la tarea de ordenación de cartas de Wisconsin (Heaton, 1993) o el test Stroop en su versión informatizada (Assef, Capovilla, & Capovilla, 2007).

En este capítulo de la tesis, se propone mejorar estos test computarizados dotándolos con técnicas automáticas de reconocimiento de patrones corporales y faciales. Para ello, nos basaremos en los resultados obtenidos en la investigación realizada en el área de Psicología de la emoción. Aunque nuestro objetivo es evaluar rasgos de personalidad y no emociones, las herramientas desarrolladas en el campo de las emociones pueden adaptarse fácilmente a nuestros intereses.

En los últimos años, el área de Psicología de la emoción ha obtenido diversos avances en caracterizar cómo se expresan las distintas emociones, cómo las reconocemos o cuáles son sus bases neurobiológicas. Aunque podemos expresar nuestras emociones a través de diferentes canales o modalidades, en un artículo relativamente reciente, Gelder (2009) señaló que aproximadamente el 95% de los trabajos aparecidos en el área de emociones se centran en expresiones faciales. El 5% restante está principalmente relacionado con el análisis de la voz.

El sistema FACS ("Facial Action Coding System") es una de las razones por las que la mayor parte de la investigación en el área de Psicología de la emoción se ha centrado en las expresiones faciales. Este sistema de codificación, desarrollado por Ekman y Friedsen (Ekman & Friesen, 1978), permite caracterizar las distintas expresiones faciales en función de las denominadas unidades de acción (UA). Estas UAs son 44 variables binarias que identifican ciertas acciones en el rostro como consecuencia de movimientos de los músculos faciales. A modo de ejemplo, levantar la ceja derecha o bajar la esquina del labio izquierdo[3] son dos de las UAs que componen el sistema FACS. Estas unidades de acción pueden aparecer de forma aislada o en combinación. Se han encontrado más de 7000 combinaciones diferentes de unidades de acción (Scherer & Ekman, 1982).

El sistema FACS fue inicialmente pensado para que un observador entrenado pudiera anotar las diferentes UAs que encontraba en imágenes estáticas. Sin embargo, los avances en las tecnologías han creado nuevos retos como es el reconocimiento de UAs en secuencias de video. Esto ha dado lugar a la necesidad de automatizar el reconocimiento de las UAs. Para solventar esta necesidad, se han desarrollado diversos trabajos que permiten obtener automáticamente varias de las UAs (Valstar & Pantic, 2006; Tian, Kanade, & Cohn, 2001).

[3] Estas unidades de acción puede encontrarse en http://www.cs.cmu.edu/~face/facs.htm

Muchos de estos trabajos se han basado en los modelos activos de forma (Cootes, Taylor, Cooper, & Graham, 1995) y los modelos activos de apariencia (Cootes, Edwards, & Taylor, 2001). Estas técnicas estadísticas son complejas, de difícil elaboración y utilizadas principalmente por investigadores en el área de visión por ordenador. Requieren disponer de bases de datos con las que entrenar los modelos y las condiciones de captura durante los experimentos deben ser similares a las existentes en las bases de datos. Estos motivos han provocado que investigadores en áreas como Psicología de la personalidad o de la emoción hayan tenido limitado su uso.

En relación a las expresiones corporales, como se ha comentado previamente, su utilización en trabajos de investigación ha sido marginal hasta hace relativamente poco. Sin embargo, en los últimos años esta tendencia ha cambiado, y han empezado a aparecer un número considerable de trabajos que proponen el uso de expresiones corporales como predictores de emociones. Esto ha sido debido a diferentes motivos. Primeramente, se ha mostrado que las tasas de reconocimiento de emociones utilizando expresiones corporales es similar, y en ocasiones superior, a las obtenidas mediante expresiones faciales (Tuminello & Davidson, 2011). Además, se ha mostrado que cuando la expresión facial y la expresión corporal son incongruentes, la expresión corporal es el factor dominante en el reconocimiento de la emoción (Meeren, van Heijsbergen, & Gelder, 2005). Otro aspecto interesante mostrado por algunos investigadores es que las personas tienden a controlar más las expresiones faciales que las corporales cuando intentan ocultar sus emociones (Ekman & Friesen, 1974). Es importante indicar que este interés reciente por el reconocimiento de las expresiones corporales no reside únicamente en la posibilidad de utilizarlo como una alternativa a las expresiones faciales, sino en la posibilidad de desarrollar sistemas multimodales (Scherer & Ellgring, 2007).

Como resultado de los distintos trabajos de investigación desarrollados, se han propuesto diversos predictores corporales que pueden ser utilizados para identificar diversas emociones. Algunos de ellos son la inclinación de la persona, la apertura del cuerpo, la posición de la cabeza, velocidades en diferentes direcciones o cuantificación de grados entre articulaciones. Una buena revisión sobre los diversos predictores que han aparecido en la literatura es la desarrollada por Kleinsmith y Bianchi-Berthouze (Kleinsmith & Bianchi-Berthouze, 2013).

Al igual que con los predictores faciales, un aspecto importante a considerar, es como de accesibles y de precisos son los predictores corporales. Hasta hace unos años, estos predictores eran obtenidos de forma rudimentaria y poco precisa, o alternativamente, a través de tecnologías caras y complicadas. En el primer grupo, muchos trabajos se basaban en las descripciones proporcionadas por un grupo de observadores “expertos” (Coulson, 2004; Clavel, Plessier, Martin, Ach, & Morel, 2009). Realizar estos trabajos era complicado por el tiempo necesario en etiquetar las imágenes disponibles o las secuencias de video y los resultados dependían de la habilidad de los observadores. En el segundo grupo, los métodos de captura de movimiento eran mayoritariamente intrusivos[4] y muy costosos[5] (con precios cercanos a los 8000 euros). Entre los sistemas existentes se encontraban los mecánicos basados en goniómetros colocados a lo largo del cuerpo, los magnéticos, los inerciales basados en acelerómetros y los ópticos basados en la colocación de emisores de luz. También existían métodos no intrusivos basados en técnica de visión por ordenador que eran usados en menor medida al requerir una alta potencia computacional y ofrecer una menor precisión.

En los últimos años, estas barreras para la obtención de estas medidas faciales y corporales han desaparecido gracias a la aparición de la cámara Kinect. Este dispositivo de bajo coste (200 euros) desarrollado por Microsoft es un dispositivo que integra una cámara tradicional RGB, un conjunto de cuatro micrófonos, un proyector y un sensor de infrarrojos. Estos elementos en conjunción con unas librerías informáticas proporcionadas gratuitamente por Microsoft (el denominado Kinect SDK) permiten la identificación en tiempo real y de forma precisa de 25 puntos del cuerpo del participante. Estos puntos se muestran en la Figura 1.

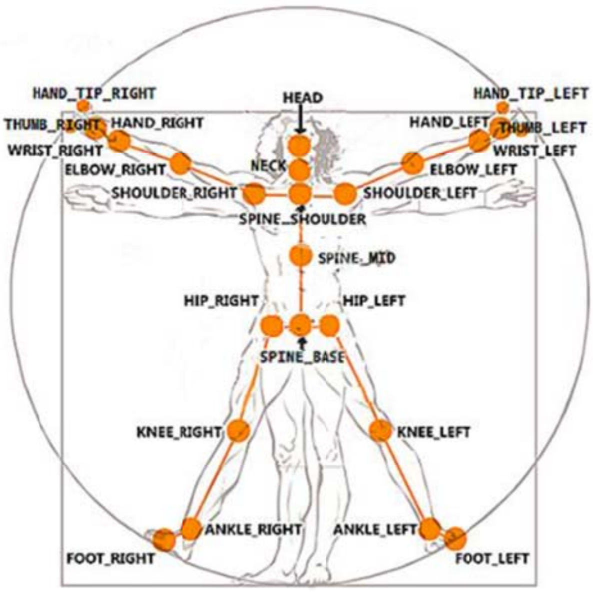


Figura 1: Distintos puntos corporales capturados por la cámara Kinect (Cook, 2014).

Por otro lado, Microsoft también proporciona, de forma gratuita, una segunda librería que permite obtener algunas de las UAs del sistema de codificación FACS. Además de estas UAs que Microsoft proporciona por defecto, también existe la posibilidad de calcular las restantes o incluso otro tipo de patrones implementando algoritmos en la malla de puntos

[4] http://www.xsens.com/en/company-pages/company/human-mocap

[5] Por ejemplo, el precio del exoesqueleto *Gypsy 7* es de 8000$: http://www.metamotion.com/gypsy/gypsy-motion-capture-system.htm.

faciales que es posible obtener en cada instante de la prueba. Uno de estos mallados de puntos obtenido en una de las pruebas realizadas se puede apreciar en la Figura 2.

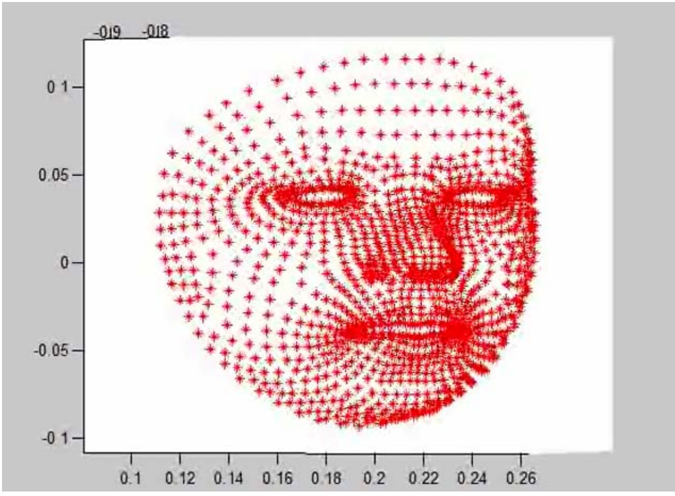


Figura 2: Puntos faciales obtenidos con la cámara Kinect.

De los párrafos anteriores, se puede intuir el gran potencial que tiene este dispositivo para las distintas áreas que componen Psicología. Por ello, y debido a que este dispositivo es desconocido para la mayoría de los investigadores en estas áreas se ha elaborado un documento acerca de la cámara Kinect: historia, componentes, librerías existentes y aplicaciones actuales que puede descargarse en la siguiente dirección: https://www.dropbox.com/s/ahb3an07p077680/Ap%C3%A9ndice%20KINECT.docx?dl=0.

En este capítulo de la tesis presentamos el primer trabajo realizado en esta nueva línea de investigación que se centra en el análisis de movimientos del participante durante la realización de una prueba computarizada. Los trabajos preliminares realizados en relación al análisis facial al igual que los trabajos que actualmente se están realizando y futuras líneas de investigación se expondrán con más detalle en el capítulo siguiente de conclusiones. El objetivo de este trabajo inicial consistía en mejorar la valoración de impulsividad proporcionada por el test de rendimiento continuo de Conners.

El test de rendimiento de Conners (Conners, 2000) es una de las pruebas computarizadas más usadas para evaluar niveles de atención e impulsividad. La prueba para adultos dura aproximadamente 15 minutos. Durante la prueba, se muestra una secuencia de estímulos consistente en 360 letras del alfabeto en la pantalla del ordenador. Esta secuencia de letras se compone de 18 bloques de 20 letras. Cada bloque se divide a su vez en dos sub-bloques idénticos tanto en composición y orden de letras como en tiempo entre estímulos. El tiempo entre estímulos dentro de un bloque puede ser de 1, 2 o 4 segundos. Cada vez que aparece una letra distinta de una prefijada (en nuestro caso la X), el participante tiene que pulsar lo más rápido posible la barra espaciadora. Cuando aparece el estímulo X, el participante debe inhibir la acción. Durante la realización del test se calcula el tiempo promedio de reacción y el número de omisiones para los estímulos que no son X y el número de comisiones (pulsar la barra cuando se debía haber inhibido la acción) para los estímulos X. El número de comisiones y el tiempo de reacción son utilizados como índices de la impulsividad de un individuo (Edman, Schalling, & Levander, 1983; Conners & Staff, 2000).

Durante unas pruebas iniciales, se observó que algunos participantes iniciaban el movimiento de presionar la barra espaciadora ante un estímulo inhibitorio X. Sin embargo, justo antes de presionar la barra, se daban cuenta del tipo de estímulo e inhibían la acción sin que este comportamiento, signo de impulsividad, quedase registrado. La hipótesis que planteamos es que detectar este patrón de comportamiento podría proporcionarnos una medida más precisa de impulsividad. Para facilitar la captura de movimiento con Kinect, solicitamos a los participantes que en lugar de presionar la barra espaciadora, levantasen su brazo dominante cada vez que apareciese un estímulo que no fuera la letra X.

Para evaluar la técnica propuesta, una muestra de 22 participantes realizó el CPT tanto en su versión tradicional como en su versión “kinectizada”. En ambos enfoques se obtuvo tanto el tiempo de reacción promedio como el número de comisiones para cada uno de los participantes. Además de estas dos medidas, cada participante rellenó el cuestionario de impulsividad de Barratt en su adaptación española (Oquendo, Baca-García, Graver, Morales, Montalvan, & Mann, 2001). El rendimiento de cada técnica fue medido a través de las correlaciones del tiempo de reacción promedio y el número de comisiones con respecto a la puntuación obtenida en el test de Barratt. Se observó que la correlación entre la impulsividad medida con la escala de Barratt y el número de comisiones obtenido con el test tradicional era de 0.217 mientras que la correlación cuando las comisiones eran obtenidas utilizando la Kinect era de 0.462. En relación a los tiempos de reacción, las correlaciones fueron de 0.371 cuando el test era realizado de la forma tradicional y 0.382 utilizando la Kinect. Es decir, como se había planteado, la aproximación propuesta mejoró las valoraciones de impulsividad cuando era caracterizada a través del número de comisiones, sin empeorar el índice de impulsividad basado en los tiempos de reacción.

Este artículo (Delgado-Gómez, y otros, 2015) se presenta a continuación. El material complementario puede encontrarse en el Apéndice A.

Behav Res
DOI 10.3758/s13428-015-0668-y

Improving impulsivity assessment using movement recognition: A pilot study

David Delgado-Gómez[1] · Carlos Carmona-Vázquez[2] · Sofia Bayona[3] · Juan Ardoy-Cuadros[4] · David Aguado[5] · Enrique Baca-García[6] · Jorge Lopez-Castroman[7]

Abstract The Continuous Performance Test (CPT) is a widely used computerized test to assess impulsivity. This article proposes the use of a CPT variant based on movement recognition to obtain more accurate measurements of impulsivity. In this pilot study, 22 volunteers participated in a CPT experiment responding to the stimuli by raising his or her dominant hand instead of pressing the space bar in a keyboard. Using this method, correlations of self-reported impulsivity with number of commission errors and average reaction time improved those obtained with standard CPT.

Keywords Continuous Performance Tests · Kinect · Impulsivity · Barratt's Impulsiveness Scale

Electronic supplementary material The online version of this article (doi:10.3758/s13428-015-0668-y) contains supplementary material, which is available to authorized users.

✉ David Delgado-Gómez
david.delgado@uc3m.es

1 Department of Statistics, Universidad Carlos III, c/Madrid, 126, 28903 Getafe, Spain

2 Faculty of Telecommunications, Universidad Carlos III, Leganés, Spain

3 Department of Computer Science, Universidad Rey Juan Carlos, Móstoles, Spain

4 Departamento de Medicina y Cirugía, Psicología, Medicina Preventiva y Salud Pública, Universidad Rey Juan Carlos, Alcorcón, Spain

5 Knowledge Engineering Institute, Autonomous University of Madrid, Madrid, Spain

6 Department of Psychiatry, Fundación Jimenez Diaz Hospital, Madrid, Spain

7 Department of Adult Psychiatry, CHRU Nimes and INSERM U1061, Nimes, France

Introduction

Although a consensual definition does not exist, the multifaceted construct of impulsivity can be understood as "a predisposition toward rapid, unplanned reactions to internal or external stimuli without regard to the negative consequences of these reactions to the impulsive individual or to others" (Moeller, Barratt, Dougherty, Schmitz, & Swann, 2001). Impulsivity appears in every major system of personality (Whiteside & Lynam, 2000), characterizes several mental disorders – such as drug use disorders or borderline personality disorder (Chamberlain & Sahakian, 2007), and has been directly associated with educational failure (Spinella & Miley, 2003). Of note, several dimensions of impulsivity, such as behavioral disinhibition, risky decision-making, or delay discounting, have been described (Avila, Cuenca, Félix, Parcet, & Miranda, 2004; Dom, De Wilde, Hulstijn, & Sabbe, 2006; Glicksohn, Leshem, & Aharoni, 2005). Thus, impulsivity traits can be defined and measured in different ways. In this study, we will use the term impulsivity to mean behavioral disinhibition, which refers to acting without thinking or failing to inhibit an initiated response (Iacono, Malone, & McGue, 2007).

The assessment of an individual's impulsivity is frequently carried out through questionnaires in clinical settings (Chamberlain & Sahakian, 2007). However, this approach presents some weaknesses. Firstly, the construction of a questionnaire is influenced by culture and the correspondence with the actual trait is difficult to validate (Möller, 2009). Secondly, the reliability of the individual answering the questionnaire is not guaranteed and might be affected by life events. For instance, psychiatric patients may modify their answers in order to get access to drugs or to be discharged (Marshall et al., 2010). Thirdly, questionnaires are unsuitable for repeated use because of learning biases (Chamberlain & Sahakian, 2007).

Published online: 20 October 2015

Thus, task-oriented computerized tests have been developed during recent decades to obtain more reliable impulsivity measures.

The Continuous Performance Test (CPT) is allegedly the most popular clinic-based measurement of response inhibition (for a review see Riccio, Reynolds, & Lowe, 2004). CPTs consist generally of a fast presentation of continuously changing stimuli, among which an infrequent "target" requires a particular action, such as pressing a space bar (Conners & Staff, 2000). This test allows quantifying two variables that are related to impulsivity: the average hit reaction time (HRT) and the number of commission errors. Additionally, the recent emergence of affordable motion-sensing devices into the market has opened promising lines of research that try to improve psychometric assessments using new behavioral indexes. Movement recognition using depth/color images can help to study behavioral patterns (Liu & Shao, 2013) or to estimate attention levels combining body posture and head orientation (Stanley, 2013).

This article aims at improving the assessment of impulsivity through the CPT task using a motion sensor device (Microsoft Kinect©), which is readily available for consumer entertainment. The movement recognition ability of this kind of device captures behavioral patterns that are unnoticed by usual CPTs, such as uncompleted actions in which the subject starts the movement but does not press the space bar. We hypothesize that capturing these behaviors will provide a more accurate evaluation of impulsivity.

Material and methods

Sample and set-up

In order to evaluate the proposed method, a sample of 22 university students conducted the CPT twice, once responding through a keyboard (standard CPT) and once raising their dominant hand in front of a Kinect sensor (Kinect CPT, set-up shown in Fig. 1S). Two screening questions were made to rule out psychiatric diagnoses among the participants: "Have you ever been diagnosed with a mental illness?" and "Have you ever received psychological or psychiatric treatment?" All participants responded negatively to both questions. They were also first-time users of Kinect and not aware of the study procedure. To prevent recall biases in the second test, two versions of CPT were randomly generated using the Psychology Experiment Building Language (Mueller & Piper, 2014). The assignation of these two tests was counterbalanced and alternated for each subsequent subject, half of the participants (n = 11) completed the keyboard version first and the other half completed the Kinect version first. The average age of the participants was 24.7 years, with a standard deviation of 4.05. Most participants were males (63 %).

Standard Continuous Performance Test (CPT)

In this study, the CPT presented 360 letters on a computer screen with an interval of 1, 2, or 4 s between them, altogether lasting approximately 14 min. The participants were asked to hit the space bar on the keyboard every time a letter, other than X, appeared on the screen. Whenever the X appeared, the participants were asked to refrain from hitting the space bar. For each subject, a collection of different measures was recorded: (1) the type of stimulus and its onset time, (2) whether the participant pressed the space bar or not; and (3) the elapsed time between the emergence of the stimulus and the instant when the space bar was hit. The HRTs were measured as the elapsed time between the moment when non-X letters appeared on the screen and the moment when the space bar was hit. The commission errors represented the number of times that participants hit the space bar to an X stimulus. To make tests with different numbers of X stimuli comparable, we used the proportion of commission errors with respect to the total number of X stimuli instead of using just the number of commission errors. Thus, HRT was calculated on non-X stimuli, while commission errors were quantified on X stimuli. A very fast HRT, when combined with high commission error rates, indicates impulsivity (Riccio et al., 2004).

Kinect CPT

A standard RGB color camera, an infrared projector, an infrared sensor, and a set of four microphones compose the Microsoft Kinect sensor. The Kinect software development kit (SDK), which can be freely downloaded from Microsoft's Kinect for Windows website, allows locating 20 different body parts of the individuals. Among them, the real-time location of the dominant hand was measured according to screen coordinates and used to establish a variant of the standard CPT. The authors can provide the software developed for the present study on request.

In this version of the CPT, the participants were placed in front of a screen and presented with non-X stimuli to which they should raise their hand and X stimuli to which they should inhibit the action. This method captures events that cannot be detected using standard CPT, particularly those in which the participant starts an action but stops it before completion. The following measures were registered to quantify HRT and commission errors: the onset time and type of stimulus, and the position of the dominant hand – in screen coordinates – every 1/35 s.

The HRT was calculated as the elapsed time between the appearance of the non-X stimuli and the moment when the participants started raising their hand. This point was determined by adjusting a piecewise function to the first part of the signal, using the gradient descent technique (see Fig. 2S). The time lapse between stimuli was maintained equal to the original values (i.e., 1, 2 and 4 s). Please see supporting information for details of the procedure.

To avoid an overestimation of response times, the HRTs for non-X stimuli was only calculated if the hand was at resting position before the next stimulus appeared in the screen (Fig. 3S). We adjusted a regression line using six sample points before and six sample points after the appearance of the stimulus. The choice of the number of sample points was based on a sensitivity analysis. The slope of this line would be close to zero if the participant started the movement from a resting position, but it would adopt a negative value if the hand were still moving down. Thus, if the slope of the regression line exceeded a certain threshold (T1), the movement started from the resting position. The value of T1 was determined using a supervised learning approach by checking if every instance of a small sample of reactions had started from the resting position. T1 threshold was the value that most accurately classified the previously annotated reactions. On the other hand, to avoid an underestimation of reaction times in premature movements, a second threshold (T2) was calculated using a similar procedure. We considered that the reaction started after the emergence of the stimulus when the value of the slope was smaller than T2. In summary, the valid reactions were the ones for which the estimated value of the slope was situated between T1 and T2. Another potential method to measure the reaction of each participant could have been based on the speed of the dominant hand (Studenka, Zelaznik & Balasubramaniam, 2012).

Finally, the following Motion-based Impulsivity Index (MBI) was used to determine commission errors:

$$\mathrm{MBI}(i) = (H_x - h_x)/(M - h_x)$$

where h_X is the height (of the dominant hand) at the instant when the i^{th} stimulus appears; H_X is the maximum height since the onset of a X stimulus until the next stimulus appears; and M is the maximum height attained in the interval that follows the previous or the posterior stimulus (see Fig. 4S). Notice that the numerator accounts for the displacement of the hand between the appearances of stimuli i and i+1; and the denominator represents the longest displacement of the hand after the previous or posterior stimulus. Consequently, this index will be close to 0 if the individual inhibits the action and close to 1 if the action is completed. Of note, the MBI avoids the limitations of using a spatial threshold, which should be adapted to the height of each participant, as well as their reaction patterns. Below, in the results section, we suggest an approach to transform the MBI into a binary value that indicates if the examinee has made a commission. The supplementary video illustrates how the Kinect CPT works.

Barratt impulsiveness scale

All participants were native Spanish speakers and completed the 11th version of the Barratt Impulsiveness Scale (BIS-11) in its Spanish version (Oquendo et al., 2001). The BIS-11 is a self-report scale, which includes 30 items such as "I plan tasks carefully" or "I am self controlled." Each item's score ranks from 1 (rarely/never) to 4 (almost always/always). A total score for BIS-11 is obtained by summing up all item scores. Participants were asked to complete the BIS-11 assessment once the CPTs were performed to avoid any bias.

We compared the performance of Kinect CPT and standard CPT calculating the correlation of BIS-11 scores with the number of commission errors and the average reaction time in both types of CPT.

Results

Figure 1 shows an example of the trajectory of the dominant hand in Kinect CPT. The X-axis indicates the time and the Y-axis represents the height of the hand in screen coordinates. The plot is displayed in dashed lines when the individual should not raise his or her hand (X stimuli); otherwise it is displayed in solid line. The onset time of each stimulus is represented with a vertical dashed line. Two longer examples and their corresponding indexes of commission errors can be observed in Figs. 5S and 6S, respectively.

The average HRT (SD) for Kinect CPT was 0.382 s (0.05) while the average HRT for standard CPT was 0.371 s (0.04). The average proportion of commission errors (SD) was 0.31 (0.16) for Kinect CPT and 0.39 (0.20) for standard CPT. No significant difference was found between average HRT ($t(42) = 0.80$; $p = 0.42$) or average proportion of commission errors ($t(42) = 1.46$; $p = 0.15$) using Kinect CPT and standard CPT. The average BIS-11 score was 66.13 (9.22). There was a high correlation in commission errors ($r = 0.568$) and HRT ($r = 0.768$) between Kinect CPT and standard CPT. Given the relatively small sample size, statistical tests do not detect significant differences in correlations when comparing standard CPT and Kinect CPT.

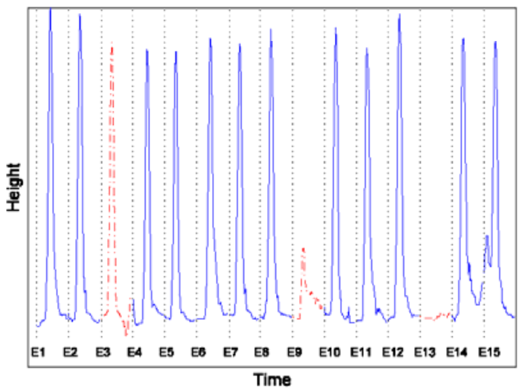


Fig. 1 Trajectory of the dominant hand for 15 stimuli. Solid lines represent the responses to non-X stimuli, while red dashed lines represent responses to X stimuli

In the standard CPT task, we found a correlation between BIS-11 and commission errors of 0.217. However, for Kinect CPT this correlation depends on the threshold that identifies movements as commission errors. Figure 2 shows the correlation between commission errors and BIS-11 as a function of the selected threshold. To facilitate the comparison, a horizontal line representing the correlation obtained according to standard CPT is depicted. The highest correlation for Kinect CPT is obtained with a 6 % threshold: 0.524. This value, which more than doubles the correlation with standard CPT, clearly indicates the improvement provided by Kinect CPT despite the lack of significant differences. Indeed, regardless of the threshold value, correlations between BIS-11 and Kinect CPT are largely superior to those obtained with standard CPT. The distribution of the indexes could also be fitted to a mixture of two Gaussian distributions (Taxt, Hjort, & Eikvil, 1990) to determine the value that best separates (statistically) both groups. Any record above this value would be considered a commission error. The intersection point of the two Gaussians defined a 13 % threshold with a correlation of 0.462.

On the other hand, the correlation between BIS-11 and HRT was −0.466 for standard CPT while this correlation was −0.534 for Kinect CPT. In both cases, the more impulsive the participant, the shorter the reaction time.

Discussion

In this article, an innovative implementation of the CPT employing the Microsoft Kinect motion sensor has been compared with a standard CPT to measure two indexes of impulsivity: commission errors and HRT. The correlation between the proportion of commission errors and the BIS-11 was

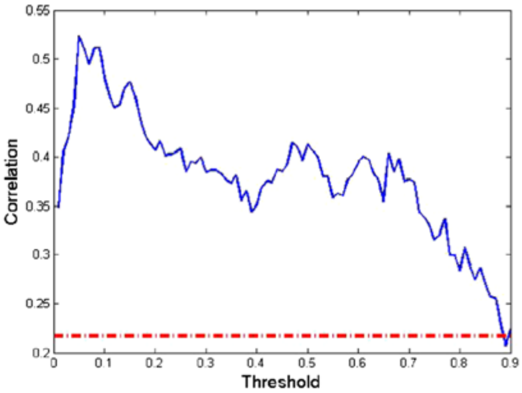


Fig. 2 Correlation between BIS-11 and the Kinect Continuous Performance Test (CPT) depending on the threshold applied to determine commission errors. The red dashed horizontal line at the bottom marks the correlation between BIS-11 and standard CPT (r=0.217)

positive for both methods, confirming that impulsive participants make more commission errors than non-impulsive ones (Conners & Staff, 2000). Likewise, the correlation between HRTs and BIS-11 were negative for both types of CPTs, reflecting the shorter reaction times found among impulsive individuals (Conners & Staff, 2000). The longer distance to "hit" in Kinect CPT compared to standard CPT agrees with our finding of higher reaction times and consequently less commission errors. Participants had more time to inhibit their actions in Kinect CPT.

As hypothesized, the Kinect CPT achieved higher correlations in absolute value between both indexes of impulsivity (commission errors and HRT) and the BIS-11, than the standard CPT. These findings suggest that capturing incomplete "hits" by Kinect CPT allows a more precise measurement of impulsivity when compared with standard CPT. The relationship of impulsivity with HRT and commission errors in CPT tests is well established (Edman, Schalling, Levander, 1983). Because of the central role played by impulsivity in many psychiatric syndromes, an easy but reliable way to assess impulsivity has become an important issue in the psychiatric field. Thus, a more informative version of CPT could be an accurate and culturally unbiased measurement of impulsivity, in turn facilitating early diagnosis and personalized treatments. The utilization of a Kinect CPT would convey other potential applications derived from movement recognition: (1) patterns of movement can be combined with personality measures or other psychometrical tests to explore their interaction; (2) movement recognition analyzes non-programmed actions and thus makes easier the evaluation of handicapped or aged persons (i.e., the recognition can be adapted to the movement); (3) movement disorders could be evaluated or controlled for when using a CPT test.

The results also open new avenues for research. For instance, movement analysis of other body parts or body posture could inform about sustained attention or vigilance (Stanley, 2013). Moreover, this article has been focused on the CPT, but the development of specific tests or the adaptation of old ones could maximize the psychometric utility of motion sensor devices. As already indicated, this kind of device is increasingly affordable while in parallel specific software and algorithms for data analyses are becoming available. Thus, the implementation of movement recognition and analysis for personality assessment is now feasible in many different areas.

Some limitations should be mentioned. First, although a 1-s lapse is enough in standard CPTs to press and release the space bar, some participants found difficulties in raising and returning the hand to a resting position in the Kinect CPT version within this time interval. Second, participants became tired during the test. Analysing a shorter and horizontal movement could solve these limitations. Finally, as this is a pilot study on a small sample, we were unable to find statistical differences between correlations in standard and Kinect CPTs.

However, the value of the correlation between BIS-11 and commission errors would be more than doubled by selecting the most performing threshold in Kinect CPT. New studies on larger and more representative samples, including clinical subjects, are necessary to confirm our findings, as well as to examine the evolution of the MBI over time.

In summary, standard and technology-enhanced measures of impulsivity on the CPT were comparable, and these measures were correlated to the participant's ratings on the BIS-11. This pilot study presents a novel method, based on movement recognition, that might increase the precision of impulsivity measures.

References

Avila, C., Cuenca, I., Félix, V., Parcet, M.-A., & Miranda, A. (2004). Measuring impulsivity in school-aged boys and examining its relationship with ADHD and ODD ratings. *Journal of Abnormal Child Psychology, 32*(3), 295–304.

Chamberlain, S. R., & Sahakian, B. J. (2007). The neuropsychiatry of impulsivity. *Current Opinion in Psychiatry.*

Conners, C. K., & Staff, M. (2000). Conners' Continuous Performance Test II (CPT II V. 5). *North Tonawanda.*

Dom, G., De Wilde, B., Hulstijn, W., & Sabbe, B. (2006). Dimensions of impulsive behaviour in abstinent alcoholics. *Personality and Individual Differences, 42*(3), 465–476. doi:10.1016/j.paid.2006.08.007

Edman, G., Schalling, D., & Levander, S. E. (1983). Impulsivity and speed and errors in a reaction time task: a contribution to the construct validity of the concept of impulsivity. *Acta psychologica, 53,* 1–8.

Glicksohn, J., Leshem, R., & Aharoni, R. (2005). Impulsivity and time estimation: Casting a net to catch a fish. *Personality and Individual Differences, 40*(2), 261–271. doi:10.1016/j.paid.2005.07.003

Iacono, W. G., Malone, S. M., & McGue, M. (2007). Behavioral disinhibition and the development of early-onset addiction: common and specific influences. *Clinical Psychology, 4,* 325–348. doi:10.1146/annurev.clinpsy.4.022007.141157

Liu, L., & Shao, L. (2013). *Learning discriminative representations from RGB-D video data* (pp. 1493–1500). AAAI Press.

Marshall, P., Schroeder, R., O'Brien, J., Fischer, R., Ries, A., Blesi, B., & Barker, J. (2010). Effectiveness of symptom validity measures in identifying cognitive and behavioral symptom exaggeration in adult attention deficit hyperactivity disorder. *The Clinical Neuropsychologist, 24*(7), 1204–1237. doi:10.1080/13854046.2010.514290

Moeller, F. G., Barratt, E. S., Dougherty, D. M., Schmitz, J. M., & Swann, A. C. (2001). Psychiatric Aspects of Impulsivity. *American Journal of Psychiatry, 158*(11), 1783–1793. doi:10.1176/appi.ajp.158.11.1783

Möller, H.-J. (2009). Standardised rating scales in psychiatry: methodological basis, their possibilities and limitations and descriptions of important rating scales. *The World Journal of Biological Psychiatry : the Official Journal of the World Federation of Societies of Biological Psychiatry, 10*(1), 6–26. doi:10.1080/15622970802264606

Mueller, S. T., & Piper, B. J. (2014). The Psychology Experiment Building Language (PEBL) and PEBL Test Battery. *Journal of Neuroscience Methods, 222,* 250–259. doi:10.1016/j.jneumeth.2013.10.024

Oquendo, M. A., Baca-Garcia, E., Graver, R., Morales, M., Montalvan, V., & Mann, J. J. (2001). Spanish adaptation of the Barratt Impulsiveness Scale (BIS-11). *European Journal of Psychiatry, 15*(3), 147–155.

Riccio, C. A., Reynolds, C. R., & Lowe, P. A. (2004). *Clinical Applications of Continuous Performance Tests.* Wiley.

Spinella, M., & Miley, W. M. (2003). Impulsivity and academic achievement in college students. *College Student Journal.*

Stanley, D. (2013). *Measuring attention using Microsoft Kinect.*

Studenka, B. E., Zelaznik, H. N., & Balasubramaniam, R. (2012). The distinction between tapping and circle drawing with and without tactile feedback: An examination of the sources of timing variance. *Quarterly Journal of Experimental Psychology, 65,* 1086–1100. doi:10.1080/17470218.2011.640404

Taxt, T., Hjort, N. L., & Eikvil, L. (1990). Statistical classification using a linear mixture of two multinormal probability densities. *Pattern Recognition Letters, 12*(12), 731–737. doi:10.1016/0167-8655(91)90070-3

Whiteside, S. P., & Lynam, D. R. (2000). The Five Factor Model and impulsivity: using a structural model of personality to understand impulsivity. *Personality and Individual Differences, 30*(4), 669–689. doi:10.1016/S0191-8869(00)00064-7

5. Discusión

En esta tesis se han investigado los beneficios que pueden aportar las técnicas avanzadas de reconocimiento de patrones en la resolución de diversos problemas actuales en las ciencias de la salud y del comportamiento. Se ha analizado su aportación tanto desde el punto de vista del análisis y tratamiento de los predictores cuando nos encontramos en un contexto de análisis supervisado al existir un criterio a predecir, como desde el punto de vista de análisis no supervisado en el que la aplicación de estos métodos avanzados se desarrolla en contextos que no tienen como objetivo final la predicción de un determinado criterio. Finalmente se ha considerado también la utilización de estas técnicas para la generación de nuevas evidencias estimulares que puedan ser consideradas en contextos evaluativos.

Respecto de la primera cuestión abordada, la elección y utilización de los clasificadores, se ha mostrado que la aplicación de estas técnicas para el análisis de las respuestas emitidas por las personas ante diferentes cuestionarios permite mejorar considerablemente la precisión con la que se clasifica a personas con determinadas características. Más específicamente, una mirada detallada a los diferentes resultados presentados nos muestra como las Máquinas de Soporte Vectorial son sumamente efectivas en ámbitos tan diferentes como la predicción de los intentos de suicidio y el pronóstico del nivel de ventas que alcanzará un comercial de seguros. En gran medida la ventaja de estos clasificadores descansa en dos cuestiones fundamentales: la aproximación no lineal a la predicción y la idea de aprendizaje automático. Es claro que las relaciones entre las variables no necesariamente están sujetas a una relación lineal y, por tanto, los métodos que se liberan de esta restricción pueden aportar un importante valor diferencial. Por otro lado la idea de aprendizaje automático nos lleva a que en la aplicación de estas técnicas se contempla una fuerte carga computacional en la que se itera sucesivamente sobre los conjuntos de datos hasta encontrar las mejores soluciones posibles. A pesar de estos prometedores resultados una cuestión fundamental queda sin resolver y, de hecho, supone una importante limitación para la explicación de las relaciones entre los predictores: el funcionamiento de estas técnicas como una caja negra en la que son conocidas las entradas, son conocidas las salidas, pero poco puede decirse acerca del modelo sustantivo que da cuenta de las relaciones entre los predictores para generar una buena clasificación. En cualquier caso, desde el punto de vista aplicado, en el campo de la prevención del suicidio es relevante señalar que la detección temprana del mismo a través de los cuestionarios y la mejora en las tasas de acierto en la clasificación son un primer objetivo de gran importancia aplicada y con un fuerte componente social. En este sentido el desarrollo de procedimientos alternativos que puedan mejorar la eficacia de las Máquinas de Vectores Soporte se constituye en un aspecto crucial a desarrollar en el futuro junto con la inclusión de nuevos predictores que pudieran ser incluidos en los modelos generados a partir de estas técnicas.

Por otro lado también son especialmente relevantes los resultados obtenidos al combinar ítems de diferentes escalas mediante el algoritmo Lars-en. Se ha mostrado que es posible crear una escala, conteniendo un número relativamente pequeño de ítems, con una precisión similar a la que se obtendría con un número mucho mayor. Sin duda estos resultados abren la puerta a interesantes desarrollos metodológicos desde el ámbito de la psicometría. Acostumbrados a trabajar desde la Teoría Clásica de Test sobre el concepto de escala, los desarrollos generados desde la Teoría de Respuesta a los Ítems en los que el objeto de análisis es el ítem han supuesto una verdadera revolución en el modo en el que se aborda actualmente la evaluación psicométrica. Sin embargo, a pesar de este análisis en el ítem, el concepto de escala sigue prevaleciendo. Probablemente por consideraciones más que fundadas en los constructos que se están evaluando y en la precisión con la que ha de apreciarse ese constructo. A pesar de ello, la utilización de estos algoritmos y de desarrollos similares en los que puedan utilizarse elementos de diferentes escalas para configurar test breves (desde el punto de vista de que sean los más eficientes para maximizar la clasificación de un determinado criterio) es una cuestión de sumo interés. La aportación en este sentido es plantear la evaluación como un problema de clasificación supervisado (ya que se conoce el criterio a predecir), lo que implica la utilización de información valiosa externa a la escala de medida (diagnóstico médico o el historial clínico del paciente) para la determinación de los criterios a utilizar para administrar o seleccionar los ítems más adecuados.

Arroja esta cuestión importantes retos relacionados con preservar la validez de contenido de las diferentes escalas y la adaptación de los elementos en función de las variaciones en el criterio a predecir. En los Test Adaptativos Informatizados existen procedimientos de control del contenido para asegurar que el test que se aplica evalúa el constructo en su totalidad y no aplica solo ítems de un cierto contenido porque sean los más eficientes. La incorporación de restricciones de este estilo redundaría en una mejora de la validez de contenido de los test generados mediante estos algoritmos y, en cierta medida, ayudaría a abrir aunque mínimamente esa caja negra. En cualquier caso la comparación entre estas estrategias con las más habituales en psicometría para conseguir versiones breves de los instrumentos nos permitiría tener una visión más integral del grado de valor añadido que tienen. Adicionalmente una posible futura línea de trabajo que podría mejorar los resultados presentados consistiría en seleccionar adaptativamente los ítems. La combinación en este sentido de los avances propuestos en la tesis junto con la asentada tecnología adaptativa generada desde el campo de conocimiento de los Test Adaptativos Informatizados resultaría en un interesante avance aplicado.

Respecto de la segunda cuestión trabajada en la tesis: la utilización de las técnicas no supervisadas, se muestra cómo es posible utilizar un modelo bien asentado en la literatura psicométrica como es el modelo de respuesta graduada de Samejima desde una visión del mismo como una técnica no supervisada de reconocimiento de patrones. En este sentido parece relevante señalar cómo el análisis realizado sobre el cuestionario GHQ-28 podría ser complementado con el desarrollo de estudios de invarianza para comprobar si la estructura factorial es o no la misma en las dos subpoblaciones (víctimas y familiares) y en la población general. Una extensión directa de este primer estudio es la consideración de modelos multidimensionales (el propio modelo multidimensional de Samejima) para dar cuenta de un modelado más efectivo de la estructura dimensional del cuestionario. La utilización de modelos multidimensionales con respuestas graduadas ha sido poco explorada en ámbitos aplicados y con objetivos de desarrollo de Test Adaptativos Informatizados. Ello se une a otra de las líneas de trabajo que parecen relevantes: la utilización de las técnicas de reconocimiento de patrones con el objetivo de mejorar la selección de ítems en los Test Adaptativos Informatizados. Actualmente, los criterios de máxima información son probablemente los más comúnmente utilizados. Sin embargo, se ha observado que el criterio de Máxima Información podría no ser adecuado en los primeros niveles del test cuando no se dispone de suficientes información del examinado. En este sentido la inclusión de técnicas de reconocimiento de patrones cuando existe una considerable incertidumbre sobre el verdadero nivel latente del examinado podría redundar en una mejora de la eficiencia de los Test Adaptativos Informatizados. Por otro lado, la utilización biométrica de las medidas psicométricas de la personalidad, sugiere sin duda interesantes desarrollos en el ámbito del control de la identidad en ámbitos de testing como la selección masiva de personal. En ellos el desarrollo de test de verificación en entornos controlados de aplicación se considera como uno de los mecanismos más efectivos para asegurar dicho control. La idea es verificar que las puntuaciones obtenidas por una persona en un entorno no controlado de evaluación son obtenidas por esa persona y no ha habido mecanismos de ayuda o engaño en la obtención de las mismas. En este sentido el procedimiento más habitual suele ser comparar la puntuación obtenida en la aplicación no controlada con la puntuación obtenida en un breve test de verificación. Diversos procedimientos se han desarrollado al respecto y la aproximación seguida en el trabajo presentado en esta tesis pudiera ser utilizada en combinación con algunos de ellos especialmente en aquellos basados en el *person-fit*. En cualquier caso una cuestión adicional que aparece es la profundización acerca de qué características han de tener los constructos medidos y, por tanto, cuáles de ellos y por qué son más útiles en términos de identificación biométrica. Por ejemplo, la estabilidad de los constructos medidos redundaría necesariamente en la eficiencia del procedimiento propuesto; también, la estabilidad de la persona que responde.

En relación a la creación de nuevos predictores, los trabajos presentados suponen una primera incursión en un campo nuevo en el que la conjunción de las aproximaciones psicométricas con las técnicas avanzadas de reconocimiento de patrones implementadas en el estudio presentado sugiere la generación de novedosos mecanismos de medición psicológica. El trabajo presentado abre la puerta al desarrollo de métodos eficientes para la exploración del significado psicológico de los patrones corporales adoptados por las personas. El análisis de estos patrones corporales al tiempo que una persona realiza una determinada prueba de evaluación psicológica permite introducir una nueva variable en la explicación del rasgo que intenta apreciar. Actualmente hemos comenzado a explorar esta línea de trabajo con un primer estudio en el que se reconocen los patrones corporales de un grupo de 22 participantes mientras realizan el test Stroop. Aunque los resultados son muy preliminares se ha observado que los participantes más extrovertidos (mayor puntuación en extroversión en el cuestionario NEO-FFI; McCrae & Costa, 2004) tienden a realizar un mayor número de movimientos, lo que podría ser interpretados en términos de un correlato corporal de la extroversión. No solo es posible obtener información acerca del movimiento sino que a partir de los distintos predictores propuestos desde los estudios de las emociones se incluirán mediciones de la inclinación del torso del individuo, los grados eulerianos de la cabeza, la apertura de las piernas o la posición de los brazos. En definitiva un contexto de investigación-acción que puede facilitar la generación precisa de nuevas medidas de la respuesta no verbal ni auto-informada de las personas y que, por tanto, puede derivar en una mejora sustantiva de la validez con la que apreciamos el nivel de rasgo de las personas. A su vez, este tipo de aproximación redundará sin duda en el planteamiento de nuevas estrategias y aproximaciones psicométricas que traten de dar cuenta de las garantías con que estos marcadores son obtenidos. Todo un reto.

APÉNDICES

A continuación se presentan los siguientes apéndices:

Apéndice A. Información complementaria del artículo: Improving impulsivity assessment using movement recognition: a pilot study.

Apéndice B. Listado de artículos presentados junto con su fecha de publicación.

Apéndice A

Supporting information

Procedure to calculate reaction time

We hereby describe the procedure to determine the instant when the examinee initiates the reaction to the stimulus. In order to obtain a better understanding of the procedure, Fig. 2S shows the screen coordinates of the examinee's dominant hand from the moment that the stimulus appears until the next stimulus is displayed. In this figure, *n* represents the number of samples taken by the Kinect sensor. For two stimuli separated by 2 s, Kinect provides between 60 and 80 samples depending on the current computer load, each containing 20 body points. In addition, *f(n)* represents the position of the dominant hand for each of the samples, *M* is the maximum value of *f(n)*, n_M denotes the value of *n* where *M* is attained, and *m* represents the minimum value of *f(n)*.

The hit reaction time for the stimulus being considered is calculated at the moment in which the next stimulus appears. To facilitate the procedure, the original signal was simplified as follows. Instead of considering the full signal, we reduced it by considering the segment *f(n'),* where *n'* takes the values from 1 to *T*; and *T* is given by the maximum value of *n* satisfying both, *f(n)* is less than *(M+m)/2*; and n is less than n_M. The reduced signal is fitted by a piecewise function, which is composed of two linear components and relies on three parameters. This function is defined by:

$$g(n') = \begin{cases} a & 1 \le n' \le t \\ m(x-t)+a & t < n' \le T \end{cases}$$

where *t* represents the intersection point of both linear functions; *a* characterizes the first, constant, function and *m* is the slope of the second function. These parameters are fixed to minimize the mean square error between the signal and the function, namely,

$$\min_{a,t,m} \sum_{i=1}^{n'} \left((f(i)-g(i)\right)^2$$

The value of the parameters can be obtained using any optimization library that provides a gradient descent algorithm. The fitted function is displayed in Fig. 2S.

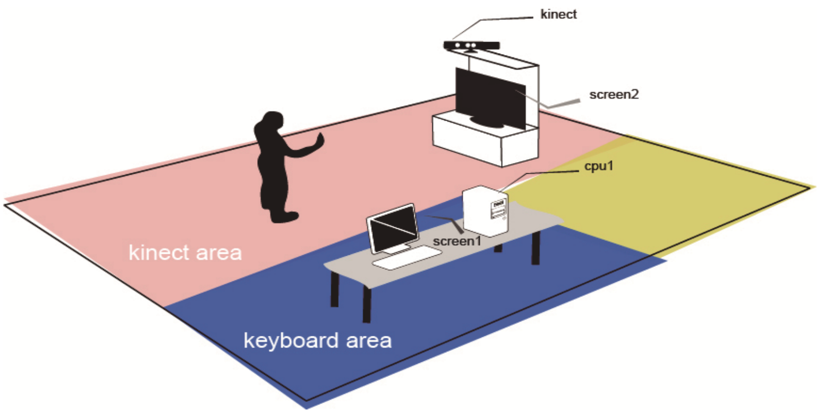


Fig. 1S Experimental set-up

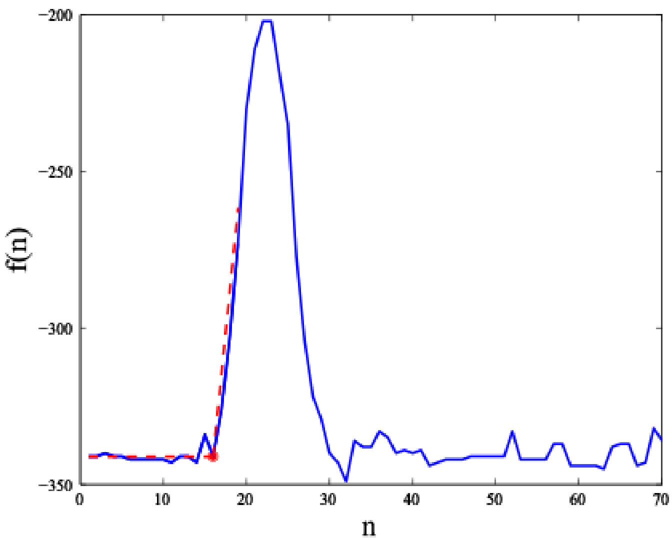


Fig. 2S Computation of the reaction time. *n* represents the number of samples taken by the Kinect sensor from the stimulus onset. *f(n)* represents the position of the dominant hand for each of the samples. The solid line represents the trajectory of the dominant hand; the red dot shows the estimated moment in which the reaction starts; and the dashed line is the fitted piecewise function. The horizontal distance between the onset of the stimulus and the red dot corresponds to the reaction time

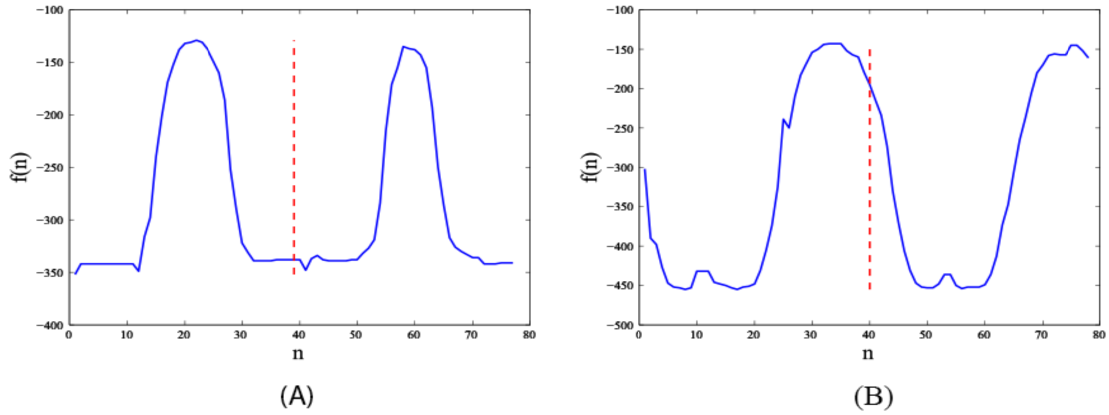


Fig. 3S Two examples of the trajectory of the dominant hand. The solid line represents the trajectory of the hand. The dashed line indicates the moment when the next stimulus appears. (A) The hand of the participant returns to a resting position before the next movement; (B) the hand has not yet reached a resting position before the appearance of the next stimulus. n represents the number of samples taken by the Kinect sensor. f(n) represents the position of the dominant hand for each of the samples

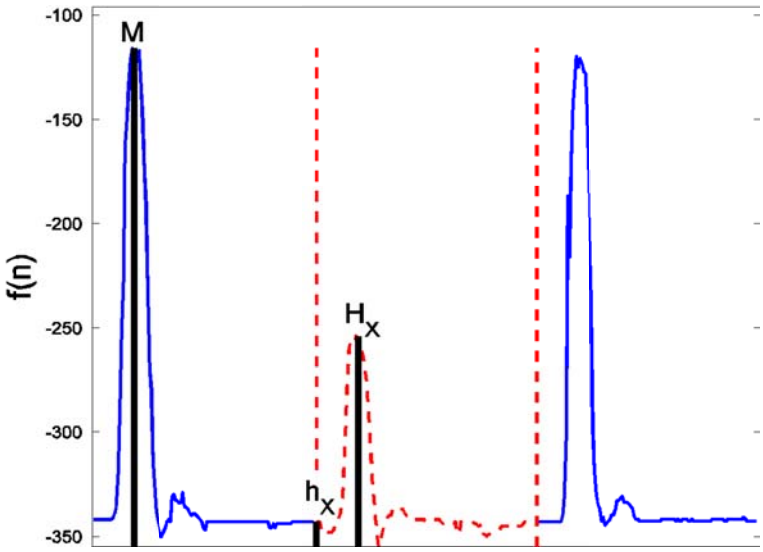
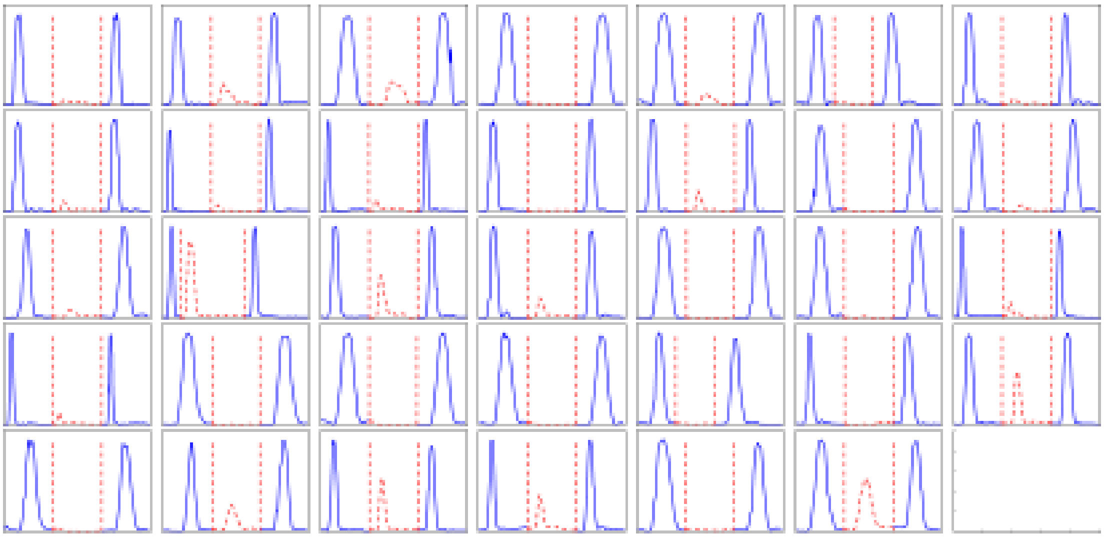
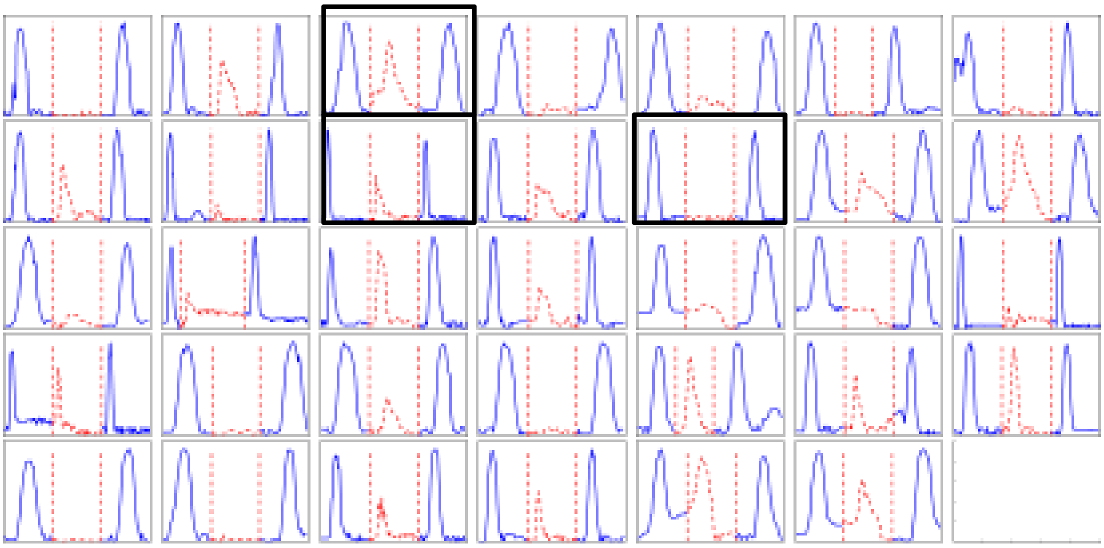


Fig. 4S Representation of the parameters used to calculate the Motion-based Impulsivity Index (MBI). The reaction of the participant to an X stimulus (dashed line) is preceded and followed by two non-X stimuli (solid lines). M is the maximum height attained in the interval that follows the previous or the posterior stimulus. In this example, the participant initiates the action but inhibits it almost immediately. The value of the MBI in this particular case was MBIx=0.35. *f(n)* represents the position of the dominant hand for each of the samples.



(A)



(B)

Fig. 5S Trajectories of the hands of two different participants responding to X stimuli. (**A**) Non-impulsive participant (low BIS-11 score); (**B**) impulsive participant (high BIS-11 score). Three different reactions to X stimuli can be appreciated for participant B: E3 (commission error), E10 (partial "hit") and E12 (correct inhibition)

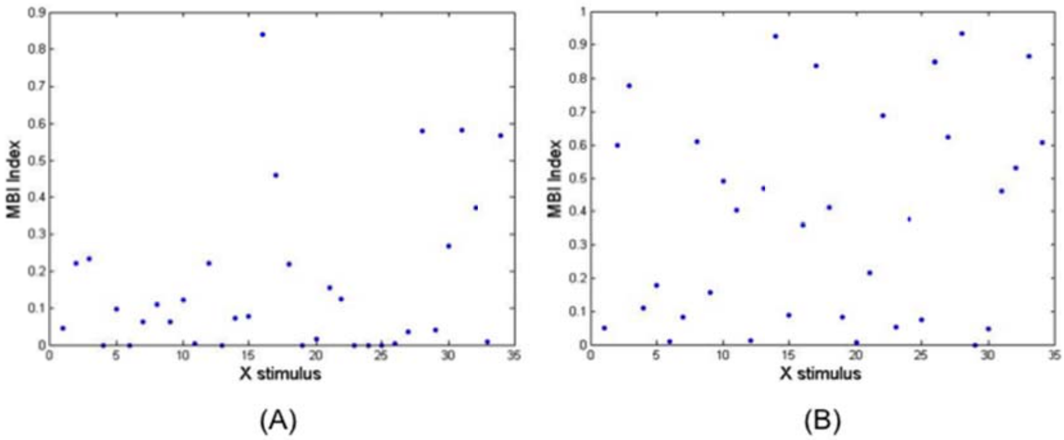


Fig. 6S Motion-based Impulsivity Indexes (MBIs) of two participants responding to the X stimuli (corresponding to registers shown in Fig. 5s). The participant (**B**) with a high BIS-11 score also shows high MBIs

Apéndice B

Listado de publicaciones presentadas en la Tesis con su fecha de publicación:

Delgado-Gomez D, Blasco-Fontecilla H, Alegria AA, Legido-Gil T, Artes-Rodriguez A, Baca-Garcia E. (2011) Improving the accuracy of suicide attempter classification. Artificial Intelligence in Medicine 52(3): 165-8. doi: 10.1016/j.artmed.2011.05.004.	Junio 2011
Delgado-Gomez D, Blasco-Fontecilla H, Sukno, F., Ramos-Plasencia, M.S., Baca-Garcia, E (2012). Suicide attempters classification: Toward predictive models of suicidal behavior. *Neurocomputing 92* (1): -8. doi:10.1016/j.neucom.2011.08.033	Septiembre, 2012
Blasco-Fontecilla, H., Delgado-Gómez, D., Ruiz-Hernández, D., Aguado, D., Baca-García, E. & López-Castroman, J. (2012). Combining scales to assess suicide risk. *Journal of Psychiatric Research 46* (10), 1272-1277.	Octubre, 2012
Artieda-Urrutia, Paula, Delgado-Gómez, David, ruiz-Hernández, Diego, García-Vega, Juan Manuel, Berenguer-Elías, Nuria, Oquendo, María A., Blasco-Fontecilla, Hilario (2015, in press). A short Personality and Life Event (S-PLE) scale for detection of suicide attempters*. Revista de Psiquiartría y Salud Mental.*	In press, 2015
Delgado-Gómez, D., Aguado, D., López-Castromán, J., Santacruz, C. y Artés-Rodríguez, A. (2011). Improving sale performance prediction using support vector machines. *Expert Systems with Applications, 38, (2011),* 5129-5132.	Septiembre, 2011
Delgado-Gómez, D., López-Castromán, J., De León, V., Baca-García, E., Cabanas-Arrate, M.L., Sánchez-González, A. & Aguado, D. (2013). Psychometrical assessment and item analysis of the General Health Questionnaire in victims of terrorism. *Psychological Assessment,* 25 (1), 279-287. Doi: 10.1037/a0030645.	Febrero, 2013
Delgado-Gómez, D., Sukno, F., Aguado, D., Santacruz, C. y Artés-Rodríguez, A. (2010). Individual Identification Using Personality Traits*. Journal of Network and computer applications,* 33, 293-299.	Junio, 2010
Delgado-Gómez, D., Carmona-Vázquez, C., Bayona, S., Ardoy-Cuadros, J., Aguado, D., Baca-García, E., Lopez-Castroman, E. (2015) Improving impulsivity assessment using movement recognition: A pilot study. *Behavioral Research Methods, 47* (3). DOI: 10.3758/s13428-015-0668-y	Septiembre, 2015

Referencias

(s.f.). Recuperado el 15 de 01 de 2014, de http://lucente.us/past/career/natural/dreamspace/index.html

(s.f.). Recuperado el 15 de 01 de 2014, de http://www.nintendo.com/wii/what-is-wii/#/controls

Artieda-Urrutia, P., Delgado-Gómez, P., Ruiz-Hernández, D., García-Vega, J., Berenguer-Elias, N., Oquendo, M., y otros. (2015). A short personality and life event scale for detection of suicide attempters. *Revista de psiquiatrñia y salud mental*.

Ashley, J. W. (2012). *Beginning Kinect Programming with the Microsoft Kinect SDK.* Apress.

Assef, E., Capovilla, A., & Capovilla, F. (2007). Computerized Stroop test to assess selective attention in childen with attention deficit hyperactivity disorder. *The Spanish Journal of Psychology, 10*(1), 33-40.

Baker, F., & Kim, S. (2004). *Item Response Theory.* CRC Press.

Becker, J., Fliege, H., Kocalevent, R., Bjorner, J., Rose, M., Walter, O., y otros. (2008). Functioning and validity of A Computerized Adaptive Test to measure anxiety (A-CAT). *Depress Anxiety, 25*(12), E182-194.

Blasco-Fontecilla, H., Alegria, A., Delgado-Gomez, D., Legido-Gil, T., Saiz-Ruiz, J., Oquendo, M., y otros. (2012). Age of first suicide attempt in men and women: and admixture analysis. *The Scientific World Journal*.

Blasco-Fontecilla, H., Bacar-Garcia, E., Courtet, P., Garcia-Nieto, R., & de Leon, J. (2015). Horror Vacui: emptiness might distinguish between major suicide repeaters and nonmajor suicide repeaters: a pilot study. *Psychotherapy and psychosomatics, 84*(2).

Blasco-Fontecilla, H., Delgado-Gómez, D., Legido-Gil, T., De Leon, J., Perez-Rodriguez, M., & Baca-García, E. (2012). Can the Holmes-Rahe Social Readjusment Rating Scale (SRRS) be used as a suicide risk? Ax exploratory study. *Archives of suicide research, 16*(1), 13-28.

Blasco-Fontecilla, H., Delgado-Gomez, D., Ruiz-Hernandez, D., Aguado, D., Baca-García, E., & Lopez-Castroman, J. (2012). Combining scales to assess suicide. *Journal of Psychiatric Research, 46*(10), 117-119.

Bo, L., Ren, X., & Fox, D. (2012). Unsupervised feature learning for rgb-d based object recognition. *ISER.*

Bolt, R. A. (1980). "Put-that-there: Voice and gesture at the graphics interface". *7th annual conference on computer graphics and interactive techniques (SIGGRAPH '80')*, *14*, págs. 262-270.

Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and Regression Trees.* Belmont: Wadsworth International Group.

Brodsky, B., Groves, S., Oquendo, M., Mann, J., & Stanley, B. (2006). Interpersonal precipitants and suicide attempts in borderline personality disorder. *Suicide Life Threatening Behaviour, 36*(3), 313-322.

Brown, G., Goodwin, F., Ballenger, J., Goyer, P., & Major, L. (1979). Aggression in humans correlates with cerebrospinal fluid amine metabolites. *Psychiatry Research, 1*(2), 131-139.

Callicot, J., Bertolino, A., Egan, M., Mattay, V., Langheim, F., & Weinberger, D. (2000). Selective relationship between prefrontal N-acetylaspartate measures and negative symptomps in schizophrenia. *American Journal of Psychiatry, 157*(10), 1646-1651.

Canetto, S. (1998). The gender paradox in suicide. *Suicide and life threatening behavior, 28*(1), 1-23.

Chalmers, R. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of statistica software, 48*(6), 1-29.

Chamberlain, S., & Sahakian, B. (2007). The neuropsychiatry of impulsivity. *Current opinion in Psychiatry, 20*(3), 255-261.

Chang, Y. J., Chen, S. F., & Huang, J. D. (2011). A Kinect-based system for physical rehabilitation: A pilot study for young adults with motor disabilities. *Research in developmental disabilities, 32*(6), 2566-2570.

Cheng, H., Zheng, N., & Qin, J. (2005). Pedestrian detection using sparse gabor filter and support vector machine. *IEEE Intelligent Vehicles Symposium.*

Cheville, A., Wang, C., Ni, P., Jette, A., & Basford, J. (2014). Age, sex, and symptom intensity influence test taking parameters on functional patient-reported outcomes. *American Journal of Physical Medicine & Rehabilitation, 16*(9).

Chye, C., & Nakajima, T. (2012). Game based approach to learn Martial Arts for beginners. *IEEE 18th International Conference on Embedded and Real-Time Computing Systems and Applications (RTCSA)*, (págs. 482-485).

Clavel, C., Plessier, J., Martin, J., Ach, L., & Morel, B. (2009). Combining facial and postural expressions of emotions in a virtual character. *Proc. Ninth Int'l Conf. Intelligent Virtual Agents*, (pp. 287-300).

Conners, C. K. (2000). Conners' Continuous Performance Test II (CPT II V. 5). .

Conners, C., & Staff, M. (2000). *Conners' Continuous Performance Test II (CPT II.V5).* North Tonawanda, NY: Multi-Health Systems Inc.

Cook, B. (24 de 4 de 2014). Recuperado el 13 de 9 de 2015, de http://www.bryancook.net/2014/03/drawing-kinect-v2-body-joints.html

Cootes, T., Edwards, G., & Taylor, C. (2001). Active apperance models. *IEEE transactions on pattern analysis and machine intelligence, 6*, 681-685.

Cootes, T., Taylor, C., Cooper, D., & Graham, J. (1995). Active shape models- their training and application. *Computer vision and image understanding, 61*(1), 38-59.

Coreil, J., Bryant, C., & Henderson, J. (2001). *Social and behavioural foundations of public health.* Sage.

Coulson, M. (2004). Attributing emotion to static body postures: recognition accuracy, confusions and viewpoint dependence. *J. Nonverbal Behavior, 28*, 117-139.

Cruz, L., Lucio, D., & Velho, L. (2012). Kinect and rgbd images: Challenges and applications. (págs. 36-49). In Graphics, Patterns and Images Tutorials, IEEE Conference on 25th SIBGRAPI .

De Beurs, D., de Vries, A., de Groot, M., de Keijser, J., & Kerkhof, A. (2014). Applying computer adaptive testing to optimize online assessment of suicidal behavior: a simulation study. *Journal medical internet research, 16*(9), e207.

Delgado-Gomez, D., Aguado, D., Lopez-Castroman, J., & Artes-Rodriguez, A. (2011). Improving sale performance prediction using support vector machines. *Expert systems with applications, 38*(5), 5129-5132.

Delgado-Gomez, D., Aguado, D., Lopez-Castroman, J., Santacruz, C., & Artes-Rodriguez, A. (2011). Improving sale performance prediction using support vector machines. *Expert systems with applications, 38*(5), 5129-5132.

Delgado-Gomez, D., Blasco-Fontecilla, H., Alegria, A., Legido-Gil, T., Artes-Rodriguez, A., & Baca-Garcia, E. (2011). Improving the accuracy of suicide attempter classification. *Artificial Intelligence In Medicine, 52*(3), 165-168.

Delgado-Gomez, D., Blasco-Fontecilla, H., Sukno, F., Ramos-Plasencia, M., & Baca-García, E. (2012). Suicide attempters classification: Toward predictive models of suicidal behavior. *Neurocomputing, 3*(8), 92.

Delgado-Gómez, D., Carmona-Vazquez, C., Bayona, S., Ardoy-Cuadros, J., Aguado, D., Baca-García, E., y otros. (2015). Improving impulsivity assessment using movement recognition: A pilot study. *Behavior Research Methods*.

Delgado-Gomez, D., Lopez-Castroman, J., de Leon-Martinez, V., Baca-Garcia, E., Cabanas-Arrate, M., Sanchez-Gonzalez, A., y otros. (2013). Psychometrical assessment and item analysis of the General Health Questionnaire in victims of terrorism. *Psychological Assessment, 25*(1).

Delgado-Gómez, D., Sukno, F., Santarcruz, C., Aguado, D., & Artés-Rodriguez, A. (2010). Individual identification using personality traits. *Journal of Network and Computer Applications, 33*(3), 293-299.

Edman, G., Schalling, D., & Levander, S. (1983). Impulsivity and speed and errors in a reaction time task: A contribution to the construct validity of the concept of impulsivity. *Acta psychologica, 53*(1), 1-8.

Ekman, P., & Friesen, W. (1974). Detecting decepction from the body or face. *J. Personality Social Psychol., 29*(3), 288-298.

Ekman, P., & Friesen, W. (1978). *The Facial Action Coding System: a technique for the measurement of facial movement.* San Francisco: Consulting Psychologists Press.

Fliege, H., Becker, J., Walter, O., Bjorner, J., Klapp, B., & Rose, M. (2005). Development of a computer-adaptive test for depression (D-CAT). *Quality life research, 14*(10), 2277-2291.

Fonseca-Pedero, E., Menendez, L., Paino, M., Lemos-Giraldez, S., & Muñiz, J. (2013). International Journal of Methods in Psychiatric Research. *Plos One, 8*(9), 73201-73209.

Gao, Y., & Mandryk, R. L. (2012). The Acute Cognitive Benefits of Casual Exergame Play. *Proceedings of the 30th international conference on Human factors in computing systems (CHI '12)*, (págs. 1863-1872). Austin, Texas, USA.

Gardner, W., Shear, K., Keheller, K., Pajer, K., Mammen, O., Buysse, D., y otros. (2004). Computerized adaptive measurement of depression: a simulation study. *BMC Psychiatry, 4*(1).

Gelder, B. (2009). Why bodies? Twelve reasons for including bodily expressions in affective neuroscience. *Philosophica Trans. Royal Soc., 364*(3), 3475-3484.

Gibbons, R., Weiss, D., Kupfer, D., Frank, E., Fagiolini, A., Grochocinski, V., y otros. (2008). Using computerized adaptive testing to reduce the burden of mental health assessment. *Psychiatry services, 59*(4), 361-368.

Gibbons, R., Weiss, D., Pilkonis, P., Frank, E., Moore, T., Kim, J., y otros. (2012). Development of a computerized adaptive test for depression. *Archives of general psychiatry, 69*(11), 1104-1112.

Gibbons, R., Weiss, D., Pilkonis, P., Moore, F., Kim, J., & Kupfer, D. (2014). Development of the CAT-ANX: A Computerized Adaptive Test for Anxiety. *American journal of psychiatry, 171*(2), 187-194.

Greenberg, H., & Greenberg, J. (1980). Job matching for better sales performance. *Harvard Business Review*, 128-133.

Greenberg, L., Kindschi, C., Dupuy, T., & Hughes, S. (1999). *Test Of Variable of Attention continuous performance test.* The TOVA company.

Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine learning, 46*(1), 389-422.

Gvion, Y., & Apter, A. (2011). Agression, impulsivity and suicide behaviour: a review of the literature. *Archives of suicide research, 15*(2), 93-112.

Hambleton, R., Swaminathan, H., & Rogers, H. (1991). *Fundamentals of item response theory.* Newbury: Park.

Hampton, T. (2010). Depression care effort brings dramatic drop in large HMO populations's suicide rate. *Journal of the American medical assication, 303*, 1903-1905.

Hayton, J., Allen, D., & Scarpello, V. (2004). Factor Retention Decisions in Exploratory Factor Analysis: a Tutorial on Parallel Analysis. *Organizational Research Methods, 7*(2), 191-205.

Heaton, R. (1993). *Wisconsin card sorting test: computer version 2.* Odessa: Psychological assessment resources.

Hilton, G., Osindero, S., & Teh, Y. (2006). A fast learning algorithm for deep belief nets. *Neural Computation, 18*(7), 1527-1554.

Holmes, E., Crane, C., Fennell, M., & Williams, J. (2007). Imaginery about suicide in depression - "flashforwards"? *Journal of behaviour therapy and experimental psichiatry, 38*, 423-434.

Holmes, T., & Rahe, R. (1967). The social readjustment rating scale. *Journal of Psychosomatic Research, 11*(2), 213-218.

Hsu, H. M. (s.f.). The potential of Kinect in education. *International Journal of Information and Education Technology, 1*(5), 365-370.

Isaacson, G. (2000). Suicide prevention - a medical breakthrough? *Acta psychiatrica scandinavica, 102*(2), 113-117.

Izadi, S., Newcombe, R., Kim, D., Hilliges, O., Molyneaux, D., Hodges, S., y otros. (2011). Kinectfusion: Real-time dynamic 3d surface reconstruction and interaction. *ACM SIGGRAPH.*

Jana, A. (2012). *Kinect for Windows SDK Programming Guide.* PACKT.

Kangis, P., & Lago, H. (1997). Using Caliper to predict performance of sales people. *International Journal of Manpower, 18*(7), 565-575.

Keilp, J., Gorlyn, M., Oquendo, M., Brodsky, B., Ellis, S., Stanley, B., y otros. (2006). Aggressiveness, not impulsiveness or hostility, distinguishes suicide attempters with major depression. *Psychological medicine, 36*, 1779-1788.

Khademi, M., Mousavi Hondori, H., Lopes, C. V., Dodakian, L., & Cramer, S. C. (2012). Haptic Augmented Reality to monitor human arm's stiffness in rehabilitation., (pág. IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES)).

Kleinsmith, A., & Bianchi-Berthouze, N. (2013). Affective body expression perception and recognition: a survey. *IEEE Transactions on affective computing, 4*(1), 15-33.

Kohut, H. (2013). *The analysis of the self: A systematic approach to the psychoanalityc treatment of narcissisti.* Chicago: University of Chicago Press.

Konrath, S., Meier, B., & Bushman, B. (2014). Development and validation of the Single Item Narcissism Scale (SINS). *Plos One, 9*(8).

Kotler, P. (1994). *Marketing management.* Englewood Cliffs: Prentice-Hall.

Lange, B., Chang, C. Y., Suma, E., Newman, B., Rizzo, A. S., & Bolas, M. (2011). Development and evaluation of low cost game-based balance rehabilitation tool using the Microsoft Kinect sensor. *Engineering in Medicine and Biology Society (EMBS)*, (págs. 1831-1834). Boston, Massachusetts USA.

Loranger, A. (1994). The International Personality Disorder Examination. *Archives of general Psyciatry, 51*(3).

Mann, J., Waternaux, C., Haas, G., & Malone, K. (1999). Toward a clinical model of suicidal behavior in psychiatric patientes. *American Journal of Psychiatry, 156*, 181-189.

Marshall, P., Schroeder, R., O'Brien, J., Fischer, R., Ries, A., & Blesi, B. (2010). Effectiveness of symptom validity measures in identifying cognitive and behavioral symptom exaggeration in adult attention deficit hyperactivity disorder. *The Clinical Neuropsychologist, 24*, 1204--1237.

Mccrae, R., & Costa, P. (2004). A contemplated revision of the NEO Five-Factor Inventory. *Personality and individual diferences, 36*(3), 587-596.

Meeren, H., van Heijsbergen, C., & Gelder, B. (2005). Rapid perceptual integration of facial expresions and emotional body language. *Proc. Nat'l Academy of Sciences USA, 102*(45), 16518-16523.

Miles, R. (2012). Start Here! Learn Microsoft Kinect API. Gravenstein Highway North Sebastopol.

Moeller, F., Barratt, E., Dougherty, D., Schmitz, J., & Swann, A. (2001). Psychiatric aspects of impulsivity. *American journal of Psychiatry, 158*(11), 1783-1793.

Mulcahey, M., Slavin, M., Ni, P., Vogel, L., Kozin, S., Haley, S., y otros. (2015). Computerized Adaptive Tests Detect Change Following Orthopaedic Surgery in Youth with Cerebral Palsy. *The journal of bone & surgery, 97*(18), 1482-1494.

Muraki, E., & Bock, R. (1997). *Parscale: IRT item analysis and test scoring for rating scale data.* Chicago: Scientific Software.

Nosek, B., & Banaji, M. (2001). The go/no-go association tast. *Social cognition, 19*(6), 625-666.

Oquendo, M., Baca-García, E., Graver, R., Morales, M., Montalvan, V., & Mann, J. (2001). Spanish adaptation of the Barratt impulsiveness scale. *European journal of psychiatry, 15*(3), 147-155.

Patton, J., Standord, M., & Barratt, E. (1995). Factor structure of the Barratt impulsiveness scale. *Journal of clinical psychology, 51*(6), 768-774.

Pavani, S., Delgado, D., & Frangi, A. (2010). Haar-like features with optimally weighted rectangles for rapid object detection. *Pattern Recognition, 43*(1), 160-172.

Petersen, M., Aaronson, N., Chie, W., Conroy, T., Constantini, A., Hammerlid, E., y otros. (2015). Development of an item bank for computerized adaptive test (CAT) measurement of pain. *Quality Life Research*.

Pillajo, C., & Sierra, J. E. (2013). Human machine interface HMI using Kinect sensor to control a SCARA robot. *2013 IEEE Colombian conference on communications and computing*, (págs. 1-5).

Qiu, J., & Helbig, R. (2012). Body posture as an indicator of workload in mental work. *Human Factors, 54*, 626-635.

Rector, K., Bennett, C. L., & Kientz, J. A. (2013). Eyes-free yoga: an exergame using depth cameras for blind & low vision exercise. *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility.*

Riley, B., Funk, R., Dennis, M., Lennox, R., & Finkelman, M. (2010). CAT Item selection and person fit. *IACAT:International Association for Computerized Adaptive Testing.* Arheim.

Rudd, M. (2000). The Suicidal Mode: A Cognitive-Behavioral Model of Suicidality. *Suicide life threatening behaviour, 30*(1), 18-33.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika*.

Scherer, K., & Ekman, P. (1982). *Handbook of methos is nonverbal behavior research.* Cambridge: Cambridge university Press.

Scherer, K., & Ellgring, H. (2007). Multimodal expression of emotion: affect programs or componentional appraisal patterns? *Emotion, 7*, 158-171.

Sebastian, A., Jacob, G., Lieb, K., & Tuscher, O. (2013). Impulsivity in borderline personality disorder: A matter of disturbed impulse control or a facet of emotional dysregulation? *Current Psychiatry Reports, 339*(8).

Shah, A. (2007). The relationship between suicide rates and age: an analysis of multinational data from the World Health Organization. *International psychogeriatrics, 19*(6), 1141-1152.

Shawe-Taylor, J., & Cristianini, N. (2004). *Kernel methods for pattern analysis.* Cambridge: Cambridge university press.

Shotton, J., Sharp, T., Kipman, A., Fitzgibbon, A., Finocchio, M., Blake, A., y otros. (2013). Real-time human pose recognition in parts from single depth images. *56*(1), 116-124.

Simms, L., Goldberg, L., Roberts, J., Watson, D., Welte, J., & Rotterman, J. (2011). Computerized adaptive assessment of personality disorder: introducing the CAT-PD project. *Journal of personality assessment, 93*(4), 380-389.

Simon, R., & Gutheil, G. (2009). Sudden improvement among high-risk suicidal pacients:should it be trusted? *Psychiatric services, 60*, 387-389.

Soloff, P., Lynch, K., Kelly, T., Malone, K., & Mann, J. (2014). Characteristics of suicide attempts of patientes with major depressive episode and borderline personality disorder: a comparative study. *American journal of psychiatry, 157*(4), 601-608.

Spinello, L., & Arras, K. (2011). People detection in rgb-d data. *International Conference on Intelligent Robots and Systems.*

Stanley, D. (2013). *Measuring attention using Microsoft Kinect.*

Stowers, J., Hayes, M., & Bainbridge-Smith, A. (2011). Altitude control of a quadrotor helicopter using depth map from Microsoft Kinect sensor. Altitude control of a quadrotor helicopter using depth map from Microsoft Kinect sensor. *2011 IEEE International Conference on Mechatronics (ICM)*, (págs. 358-362).

Suykens, J. A. (2001, May). Nonlinear modelling and support vector machines. In Instrumentation and Measurement Technology Conference, 2001. IMTC 2001. *Proceedings of the 18th IEEE* (Vol. 1, pp. 287-294). IEEE.

Tay, F., & Cao, L. (2001). Application of support vector machines in financial time series forecasting. *Omega, 29*(4), 309-317.

Tian, Y., Kanade, T., & Cohn, J. (2001). Recognizing action units for facial expression analysis. *IEEE Trans. Pattern Anal. Machine Intell., 23*, 97-115.

Tong, S., & Koller, D. (2002). Support vector machine active learning with applications to text classification. *The journal of Machine Learning Research, 2*, 45-46.

Tuminello, R., & Davidson, D. (2011). What the face and body reveal: in-group emotion effect and stereotyping of emotion in African.American and European-American children. *J. Experimental Child Psychology, 11*(3), 258-274.

Turvey, C., Conwell, Y., Jones, M., Phillips, C., Simonsick, E., Pearson, J., y otros. (2002). Risk factors for late-sife suicide: a prospective, community-based study. *The american journal of geriatric psychiatry, 10*(4), 398-406.

Ueno, M., & Songmuang, P. (2010). Computerized Adaptive Testing Based on Decision Tree. *IEEE 10th international conference on advanced learning technologies (ICALT).*

Valstar, M., & Pantic, M. (2006). Fullty automatic facial action unit detection and temporal analysis. *Computer Vision and Pattern Recognition Workshop.*

Van der Linden, W., & Hambleton, R. (2013). *Handbook of modern item response theory.* New York: Springer.

Van Gestel, T., Suykens, J. A., Baesens, B., Viaene, S., Vanthienen, J., Dedene, G., & Vandewalle, J. (2004). Benchmarking least squares support vector machine classifiers. *Machine Learning, 54*(1), 5-32.

Vera, L., Gimeno, J., Coma, I., & Fernández, M. (2011). Augmented mirror: interactive augmented reality system based on kinect. *Human-Computer Interaction–INTERACT* (págs. 483-486). Springer Berlin Heidelberg.

World report on violence and health. (2002). Geneva: World Health Organization.

Yu, X.; Wu, L.; Liu, Q.; Zhou, H.;. (2011). *Third Chinese conference on intelligent visual surveillance*, (págs. 49-52). Children tantrum behaviour analysis based on Kinect sensor.