

Tradução e Ferramenta de Validação de Conteúdo de Avaliação do Ensino pelos Alunos

Translation and Validation Content of Instrument for Students' Evaluation of Teaching

Joice Trindade Silveira *¹
João Batista Teixeira da Rocha ²

¹ Universidade Federal do Pampa

² Universidade Federal de Santa Maria

O Student Evaluation Educacional Quality (SEEQ) é um instrumento multidimensional de avaliação do ensino pelos alunos. O objetivo deste trabalho foi realizar o processo de tradução e validação de conteúdo do SEEQ para o Brasil. O trabalho foi desenvolvido em quatro etapas: 1. Tradução reversa (back translation); 2. Consolidação da revisão preliminar; 3. Validação de conteúdo da versão preliminar; 4. Teste piloto. Quatro tradutores participaram da primeira etapa, dois realizando a tradução do inglês para o português e dois do português para o inglês. Os itens foram analisados por um comitê para a obtenção de uma versão preliminar e, posteriormente, foram avaliados por especialistas quanto aos critérios de clareza de linguagem, pertinência prática, relevância teórica e dimensão. A versão preliminar apresentou muitos itens com coeficiente de validade de conteúdo (CVC) abaixo do ponto de corte, sendo, por isso, adaptado e reavaliado. Após as modificações, o instrumento apresentou CVC satisfatório (igual ou maior a 0,80) nos três critérios independentes e no instrumento como um todo. O coeficiente Kappa do critério dimensão foi moderado (0,544). Após o teste piloto, o instrumento está apto à aplicação em pesquisas. Futuramente, para sua aplicação prática, é necessário que sejam investigadas as características psicométricas da escala.

Palavras-chave: Avaliação de cursos, Métodos de avaliação, Avaliação institucional, Avaliação da atividade docente pelos estudantes.

Student Evaluation of Educational Quality (SEEQ) is a multidimensional instrument to evaluate the teaching quality. The aim of this study was the translation and content validation process of SEEQ for Brazil. The study was done in four steps: 1. Back translation; 2. Consolidation of the preliminary version; 3. Content validation of preliminary version; 4. Pilot test. Four translators participated in the first stage, two performing the translation from English to Portuguese language and two from Portuguese to English. Items were reviewed by a committee to obtain the preliminary version of the instrument and then they were evaluated by experts in clarity of language, practical relevance, theoretical relevance and dimension criteria. The preliminary version presented many items with content validity coefficient (CVC) below the cut-off value, and therefore it was adapted and re-evaluated. After the modifications, the instrument presented satisfactory CVC (≥ 0.80) at both three criteria and the whole instrument. The Kappa coefficient dimension criteria was moderate (0.544). After the pilot test, the instrument is able to use in research. For practical application, however, it is necessary that the psychometric characteristics of the scale are still investigated.

Keywords: Course evaluation, Evaluation criteria, Evaluation method, Institutional evaluation, Student evaluation of teacher performance.

*Contacto: joicetsilveira@gmail.com

issn: 1989-0397

www.rinace.net/riee/

<https://revistas.uam.es/riee>

Recibido: 26 de octubre de 2016

1ª Evaluación: 2 de diciembre de 2016

Aceptado: 22 de diciembre de 2016

1. Introdução

No Brasil, apesar de termos um Sistema Nacional de Avaliação do Ensino Superior (SINAES) já implementado há alguns anos (BRASIL, 2004; SINAES, 2003), ainda estamos distantes de uma avaliação do ensino pelos alunos realmente universal e eficiente. De acordo com o SINAES, esta atividade é de responsabilidade dos cursos de graduação, que são autônomos para definir seus objetivos e formatos de avaliação. No entanto, definir critérios para avaliação nunca é um processo fácil, pois, ao tomarmos uma determinada corrente pedagógica, deixamos de atender a outra, e um consenso passa a ser distante. Além disso, por vezes há receio dos próprios professores sobre essas avaliações, por acreditarem que elas possam ser utilizadas de forma equivocada pelos alunos. Nesses embates, o tema acaba sendo negligenciado (Chonko, Tanner e Davis, 2002).

Internacionalmente, há diversos instrumentos de avaliação do ensino pelos alunos e que são frutos de longas pesquisas sobre ensino e atividade docente. Os SETs (*Students Evaluate Teaching*), como são conhecidos, são, na sua maioria multidimensionais e podem ser utilizados total ou parcialmente dentro das instituições de forma a contribuir com o processo avaliativo. Um desses instrumentos é o SEEQ (*Student Evaluation Educacional Quality*) (Marsh, 1984, 1990), objeto da investigação deste trabalho. Devido às dificuldades relativas ao processo de avaliação do ensino, bem como a escassez de instrumentos com essa finalidade disponíveis na literatura científica em língua portuguesa, o objetivo deste trabalho foi realizar o processo de tradução e validação de conteúdo do SEEQ para o Brasil.

2. Referencial teórico

No âmbito da graduação, o ensino pode ser considerado como a atividade principal da universidade. Sua avaliação, portanto, deveria preceder os planejamentos de disciplinas, aulas e atividades e ser amplamente discutida entre alunos e professores. Uma avaliação adequada pode retroalimentar o curso e a instituição sobre a efetividade do seu corpo docente, auxiliar estudantes na escolha de disciplinas, estimular atividades de pesquisa em educação e quantificar a efetividade docente com vistas ao seu desenvolvimento e promoção (Marsh, 1984).

Quando questionados, os alunos reconhecem que há algumas características que influenciam positivamente a qualidade do ensino, como a competência do professor, a sua facilidade de comunicação com os alunos, a organização, a empatia e as exigências da disciplina (Cohen, 1980; Feldman, 2007; Marsh, 1984). Logo, embora haja fatores que estão fora do alcance dos professores, há, também, critérios de qualidade que podem fazer parte dessas avaliações e contribuir para melhorar a qualidade do ensino.

A avaliação do ensino é um tema negligenciado nas universidades brasileiras por diversos motivos. Primeiramente, há dificuldade na definição de critérios e modelos para tais avaliações, visto que, tanto os modelos formativos quanto os somativos, possuem suas limitações. Os primeiros em virtude da dificuldade de atingir um número expressivo de alunos e da análise relativista, e os segundos devido à definição de quais critérios e escalas seriam utilizados para tais avaliações e à análise puramente objetivista num

campo tão complexo como a educação (Fernández, Sánchez e García, 1996). O SINAES, apesar de considerar a avaliação do ensino pelos alunos como um importante indicador de qualidade, verifica somente se a autoavaliação dos cursos está implementada e se os resultados subsidiam melhorias (BRASIL, 2004; SINAES, 2003). Ao não propor critérios, garante a autonomia das instituições quanto às autoavaliações, mas, também, deixa em aberto a possibilidade destas realizarem avaliações superficiais, que dão pouco retorno ao professor sobre a sua prática. Outro motivo é a resistência dos docentes, já que muitos consideram que estas avaliações poderiam ser utilizadas de forma equivocada pelos alunos, permitindo atitudes como penalização ou vingança do professor por motivos vários (Chonko, Tanner e Davis, 2002). Nesse contexto, a discussão em torno de critérios e meios acaba retardando o processo avaliativo.

Internacionalmente os instrumentos mais utilizados para avaliações do ensino pelos alunos são os SETs (Students Evaluate Teaching). Normalmente eles possuem uma parte quantitativa, que consiste em questionários fechados com respostas medidas através de escalas, e outra qualitativa, com questões abertas para que os estudantes respondam livremente. Um desses questionários é o SEEQ (*Student Evaluation Educacional Quality*), utilizado em mais de 500 universidades no mundo. Ele foi desenvolvido pelo psicólogo educacional especialista em psicometria Herbert Marsh, e é utilizado para avaliar a qualidade do ensino de cursos individuais ou disciplinas. Os estudantes respondem a 34 questões objetivas (fechadas), usando uma escala likert de 5 pontos, sobre 9 dimensões relacionadas à efetividade do ensino e 3 questões subjetivas (abertas) (Marsh, 1984, 1990). Dessa forma, o SEEQ pode ser usado tanto para avaliações somativas quanto formativas, visto que as perguntas objetivas podem dar um *feedback* aos professores sobre características do seu ensino que poderiam ser melhoradas, e as subjetivas fornecem uma avaliação geral sobre esse ensino (Keane e Labhrainn, 2005).

Para chegar às nove dimensões incluídas no SEEQ foram realizados estudos preliminares na Universidade da Califórnia-Los Angeles (UCLA), a partir da necessidade do desenvolvimento de uma avaliação institucional. Em um período de 10 anos (1976-1986), cerca de 500.000 participantes, matriculados em mais de 20.000 disciplinas, representando mais de 50 departamentos acadêmicos da UCLA responderam ao SEEQ. O conjunto de itens que compuseram o questionário foi baseado em pesquisas sobre práticas correntes entre os professores, entrevistas com estudantes e professores e revisão da literatura sobre avaliação. Diversos tipos de questionários (*surveys*) piloto foram aplicados em aulas de diferentes departamentos acadêmicos. Os estudantes eram convidados a avaliar, com escala, e a indicar os itens que eles sentiam que eram mais importantes para descrever a qualidade do ensino. Os professores também eram questionados sobre quais itens que eles acreditavam que poderiam dar um melhor *feedback* sobre o seu ensino. A análise fatorial resultou na identificação de nove dimensões: aprendizado/valor, entusiasmo, organização, rapport/empatia, interação com o grupo, amplitude na abordagem, exames/avaliações, tarefas/leituras e sobrecarga/dificuldade (Marsh, 1984, 1990). O questionário demonstrou ser confiável quando aplicado em grupos com mais de 15 alunos e ter estabilidade temporal. Posteriormente, novos estudos foram conduzidos com o mesmo instrumento e forneceram um maior suporte para o uso do SEEQ (Marsh e Roche, 1992). No período de 1978-1982, foram aplicados mais de 250.000 questionários em mais de 24.000 cursos

na Universidade Ocidental da Califórnia (Marsh e Hocevar, 1991) e sua validade e reprodutibilidade têm sido confirmadas internacionalmente (Coffey e Gibbs, 2000; Marsh e Roche, 1992; Mijic, 2010; Nash, 2012). O uso desses instrumentos já validados internacionalmente pode auxiliar no processo de avaliação do ensino ao serem utilizados parcialmente ou na sua totalidade pelas instituições.

Um processo de tradução e validação cultural busca adaptar um instrumento utilizado internacionalmente para a realidade do país em questão. Porém, antes de iniciar o processo, é recomendado que sejam avaliadas duas possibilidades: a de utilização do instrumento na língua original e o desenvolvimento de um novo instrumento (Vallerand apud Cassep-Borges, Balbinotti e Teodoro, 2010, p. 508). No Brasil, a primeira opção é uma alternativa inviável, pois a maioria dos estudantes é monolíngue. A segunda, por sua vez, exige muito tempo de pesquisa e limita a comparação com outras pesquisas no exterior. Logo, a indicação, nesses casos, é utilizar o instrumento internacional validado para o país ou língua em questão.

3. Metodologia

A tradução e validação cultural do questionário seguiu os critérios descritos por Cassep-Borges, Balbinotti e Teodoro (2010), na qual foram cumpridas as 4 (quatro) etapas descritas abaixo:

1. Tradução reversa (*back translation*)
2. Consolidação da revisão preliminar
3. Validação de conteúdo da versão preliminar
4. Teste piloto – Versão final

Estas etapas foram cumpridas para todos os itens da parte objetiva do SEEQ. As questões subjetivas e as informações pessoais não passaram pelo processo de validação. O projeto foi aprovado pelo Comitê de Ética em Pesquisa da Universidade Federal de Santa Maria (CAAE: 51967515.2.0000.5346).

3.1. Tradução reversa (*back translation*)

Para a tradução do questionário, participaram quatro (4) tradutores. Inicialmente foram realizadas duas traduções do documento original, da língua inglesa para a língua portuguesa. Os dois tradutores eram profissionais bilíngues da área do ensino e com independência entre eles, de forma que os possíveis erros e divergências de opiniões entre os itens do questionário pudessem ser detectados.

Posteriormente, foram realizadas duas retraduições, onde as versões traduzidas para o português foram retraduzidas para o inglês (*back translation*) por outros dois tradutores, que também eram bilíngues e envolvidos com a educação. Entretanto, estes não conheciam o instrumento original –tradução às cegas– e não tinham participado da primeira etapa de tradução. O contato com cada um deles deu-se via correio eletrônico e posteriormente houve um encontro pessoal, onde os tradutores foram informados sobre algumas características do instrumento, a metodologia utilizada e os possíveis vieses da adaptação cultural (Cassep-Borges, Balbinotti e Teodoro, 2010).

3.2. Consolidação da versão preliminar

Depois da tradução reversa, foi realizado o procedimento de consolidação das traduções, para unificar uma versão preliminar do instrumento. As 4 versões do questionário (duas em português e duas em inglês) foram analisadas por um comitê composto por quatro pessoas: um pesquisador mestre, um pesquisador doutor (que participou através de correspondência eletrônica), um dos tradutores do questionário original do inglês para o português e um membro externo à pesquisa, da área de Educação. As análises do comitê foram feitas de maneira sistemática. Cada item retraduzido era comparado com a versão original do mesmo item. Quando a tradução era praticamente ou exatamente igual, o item traduzido era aprovado. Quando as traduções divergiam, havia discussão até a chegada de um acordo sobre o item. Dessa forma, ao finalizar todos os itens, foi obtida a versão preliminar do questionário.

3.3. Validação do conteúdo da versão preliminar

Para esta etapa, juízes-avaliadores verificaram se o teste em questão media o que se propunha a medir, pelo viés de conteúdo. Em um período de 4 meses, foram convidados a participar da pesquisa 14 juízes-avaliadores, todos eles doutores e pesquisadores na área de avaliação do ensino superior ou de avaliação de disciplinas. Nenhum deles teve participação na pesquisa em momento anterior. Como a metodologia prevê que sejam incluídos de 3-5 juízes, definiu-se que os 5 primeiros a devolver a planilha avaliada seriam incluídos. Para cada um dos 12 juízes que aceitaram participar da pesquisa foi enviada uma planilha de avaliação com todas as questões objetivas, na qual eles deveriam avaliar os critérios de clareza de linguagem (CL), pertinência prática (PP), relevância teórica (RT) e dimensão, conforme orientações descritas a seguir:

- ✓ *Clareza de linguagem:* Considera a linguagem utilizada nos itens, tendo em vista as características da população respondente. Exemplo: “O senhor acredita que a linguagem de cada item é suficientemente clara, compreensível e adequada para esta população (estudantes de graduação)? Em que nível?”.
- ✓ *Pertinência prática:* Considera se cada item foi elaborado de forma a avaliar o conceito de interesse em uma determinada população. Analisa se de fato cada item possui importância para o instrumento. Exemplo: “O senhor acredita que os itens propostos são pertinentes para esta população? Em que nível?”.
- ✓ *Relevância teórica:* Considera o grau de associação entre o item e a teoria. Visa analisar se o item está relacionado com o construto. Exemplo: “O senhor acredita que o conteúdo deste item é representativo de uma das dimensões dele, considerando a teoria em questão? Em que nível?”.
- ✓ *Dimensão teórica:* Investiga a adequação de cada item à teoria estudada. Exemplo: “O senhor acredita que este item pertence a que dimensão? Assinale apenas aquela que melhor representa o item avaliado”.

Os juízes deveriam marcar em uma escala do tipo Likert, com pontuação de 1 a 5, onde 1 indica pouquíssima concordância (com o critério considerado), 2 pouca concordância, 3 média concordância, 4 muita concordância e 5 muitíssima concordância. Na mesma planilha os juízes ainda deveriam avaliar a qual dimensão a questão pertencia (figura 1).

Questão	Critérios															Observações									
	Clareza de linguagem					Pertinência					Relevância teórica						Dimensão avaliada								
O professor era adequadamente acessível aos estudantes durante os horários de atendimento extra-classe. Escala de respostas: Discordo fortemente - Discordo - Não concordo nem discordo - Concordo - Concordo fortemente	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	A	B	C	D	E	F	G	H	I	

Figura 1. Modelo de planilha enviada aos juízes-avaliadores
 Fonte: Elaboração própria.

Na apresentação original do instrumento, os itens estão dispostos dentro de suas dimensões. Porém, para a validação, as 34 questões do AQEE foram randomizadas por computador, de forma que fossem apresentadas aos juízes-avaliadores em uma ordem distinta da do questionário original. Essa modificação procurou evitar que os itens pertencentes a uma mesma categoria estivessem muito próximos uns dos outros e que a identificação da dimensão fosse facilitada.

A partir dos resultados dos juízes-avaliadores foi calculado o coeficiente de validade de conteúdo (CVC) para os três primeiros critérios (CL, PP e RT) e considerados válidos os itens que obtiveram $CVC > 0,8$. Já para a análise da dimensão teórica foi utilizado o coeficiente Kappa. No entanto, como o Kappa avalia a concordância entre dois avaliadores, e neste trabalho participaram 3, foi utilizado o Kappa médio (Hernandez-Nieto apud Cassep-Borges, Balbinotti e Teodoro, p. 511, 2010), calculado através da média dos coeficientes Kappa entre a avaliação feita por cada juiz e a dimensão original proposta por Marsh (1984). A interpretação foi feita conforme Landis e Kock (1977) (tabela 1).

Tabela 1. Critérios para a interpretação do coeficiente Kappa

COEFICIENTE KAPPA	CLASSIFICAÇÃO
< 0	Discordancia
0-0,20	Quase nenhum
0,20-0,40	Pequeno
0,40-0,60	Moderado
0,60-0,80	Substancial
0,60-0,80	Quase perfeito

Fonte: Elaboração própria.

3.4. Teste piloto

De posse da versão experimental, foi realizado um estudo piloto em uma pequena amostra da população-alvo com a finalidade de verificar se os procedimentos estavam adequados e se algum item permanecia incompreensível àquele grupo. O estudo foi realizado em uma turma de graduação do curso de ciência e tecnologia de alimentos da Universidade Federal do Pampa, na cidade de Itaqui, no estado do Rio grande do Sul. No início da aula, em espaço cedido por um professor, foi realizada uma breve apresentação do questionário e explicado o objetivo do projeto piloto. Os 22 alunos presentes concordaram em participar da pesquisa mediante a assinatura do TCLE. Juntamente com o instrumento, os estudantes receberam uma folha de avaliação onde eles puderam

identificar cada questão e avaliá-la quanto à compreensão, clareza, vocabulário, dúvidas e outras considerações, se assim julgassem necessário. Somente esta folha de avaliação é que foi considerada na pesquisa. As respostas dos questionários fornecidas pelos estudantes neste projeto piloto foram excluídas logo após a aplicação.

Cabe salientar que a apresentação do questionário sofreu duas modificações em relação à estrutura original, sem alterações de conteúdo, conforme orientações de Mattar (1994) (Quadro 1). A primeira foi a transferência das perguntas abertas para o início do questionário, visando deixar o respondente mais à vontade para responder as perguntas restantes, já que segundo Chagas (2000), o primeiro contato do respondente com o questionário é que define a sua vontade de respondê-lo ou até mesmo a decisão de não respondê-lo. A outra alteração foi à transferência das questões relativas aos dados pessoais do respondente para o final do questionário, com o objetivo de evitar vieses. Embora o questionário seja anônimo, os dados pessoais apresentados logo no início do questionário podem levar o respondente a distorcer as respostas com vistas a proteger-se de alguma possível identificação (Chagas, 2000). As modificações na estrutura da apresentação do questionário estão demonstradas no tabela 2.

Tabela 2. Modificações na estrutura de apresentação da avaliação

ANTES	DEPOIS
1) Dados pessoais	1) Questões abertas
2) Questões fechadas	2) Questões fechadas
3) Questões abertas	3) Questões pessoais

Fonte: Elaboração própria.

O teste piloto obedeceu às recomendações indicadas ao instrumento original, sendo aplicado o mais próximo possível do final do semestre, preferencialmente por outra pessoa que não o professor da disciplina –no caso–, o pesquisador –e em grupos maiores do que 15 alunos–.

4. Resultados

4.1. Tradução reversa (*back translation*)

Na primeira etapa, de tradução do inglês para o português, ambas as traduções foram bastante semelhantes. As divergências foram, principalmente, de vocabulário específico da área. Por exemplo, a palavra *course* foi traduzida tanto como curso quanto como disciplina e a palavra *lectures* como aulas e palestras. Também foram observadas diferenças relacionadas aos tempos verbais, como *have learned* traduzida como aprendeu e tem aprendido. Como esperado, as diferenças na tradução levaram a diferenças maiores na *back translation*. Por exemplo, a expressão *course materials*, traduzida inicialmente como materiais da disciplina e materiais do curso, e na *back translation* passou para *matters of the curricular subject* ou *materials of the course*. Diferenças como estas foram encontradas em aproximadamente 20 itens do questionário.

4.2. Consolidação da revisão preliminar

Na reunião do comitê os itens foram avaliados um a um, para que houvesse um consenso sobre qual a tradução mais adequada para aquele item. Em relação ao vocabulário, optou-se por utilizar a palavra mais adequada à realidade brasileira. No exemplo citado

anteriormente, a palavra *course* foi traduzida como disciplina, pois, embora o questionário possa ser utilizado também para cursos individuais, disciplina é, ainda, o termo mais comum para referir-se a um componente curricular. Nas questões de conjugações verbais, percebeu-se que as traduções estavam adequadas mediante as condições em que foram realizadas. Coube ao grupo analisar cada item divergente até que a versão preliminar estivesse pronta para ser submetida ao processo de validação de conteúdo. O instrumento teve o seu nome traduzido para avaliação da qualidade do ensino pelos estudantes (AQEE).

4.3. Validação de conteúdo da versão preliminar

Dos 12 juízes que concordaram em participar, somente três avaliaram o instrumento. Nesta etapa, como o objetivo era verificar a adequação desta versão preliminar, foram feitas as análises dos escores obtidos na avaliação destes três juízes. A tabela 3 apresenta a média do CVC de cada item nos 4 critérios avaliados –clareza de linguagem (CL), pertinência prática (PP) e relevância teórica (RT), bem como o CVC do instrumento total. A tabela 4 demonstra em qual dimensão o item foi classificado pelos juízes e pelo autor da escala (Marsh, 1984, 1990).

Tabela 3. Valores de CVC

Nº	ITEM	CRITÉRIO		
		CL	PP	RP
1	Você considerou a disciplina intelectualmente desafiadora e estimulante.	0,9	0,9	0,93
2	Você aprendeu algo que considera importante.	0,83	0,83	0,76
3	Seu interesse no conteúdo aumentou como consequência dessa disciplina.	0,9	0,83	0,83
4	Você aprendeu e entendeu os materiais de apoio (livros, slides, etc.) desta disciplina.	0,7	0,83	0,83
5	O professor era entusiasmado em relação a ministrar essa disciplina.	0,83	0,76	0,83
6	O professor era dinâmico e enérgico na condução da disciplina.	0,63	0,9	0,63
7	O professor aprimorava as aulas com uso de humor.	0,76	0,63	0,7
8	O modo de o professor dar aulas mantinha seu interesse durante todo o período.	0,9	0,96	0,76
9	As explicações do professor eram claras.	0,96	0,96	0,76
10	Os materiais didáticos da disciplina estavam bem preparados e foram cuidadosamente explicados.	0,83	0,96	0,76
11	Os objetivos apresentados coincidiam com aqueles realmente ensinados, de forma que eu sabia onde a disciplina estava indo.	0,7	0,83	0,63
12	O professor dava aulas que facilitavam a tomada de notas/anotações.	0,83	0,7	0,7
13	Os estudantes eram encorajados a participar das discussões nas aulas.	0,9	0,76	0,83
14	Os estudantes eram convidados a compartilhar suas ideias e conhecimentos.	0,9	0,9	0,83
15	Os estudantes eram encorajados a perguntar durante as aulas, e as respostas dadas eram relevantes.	0,76	0,9	0,83
16	Os estudantes eram encorajados a expressar suas próprias ideias e/ou a questionar as do professor.	0,83	0,9	0,83
17	O professor era amigoso/cordial com os estudantes, individualmente.	0,9	0,83	0,83

Fonte: Elaboração própria.

Tabela 3. Valores de CVC (continuação)

Nº	ITEM	CRITÉRIO		
		CL	PP	RP
18	O professor fazia os alunos sentirem-se bem-vindos ao pedir ajuda/conselhos, dentro ou fora da aula.	0,63	0,7	0,7
19	O professor tinha um interesse sincero nos estudantes, individualmente.	0,63	0,7	0,56
20	O professor era adequadamente acessível aos estudantes durante os horários de atendimento extra-classe.	0,63	0,76	0,76
21	Em relação ao conteúdo, o professor comparava as implicações de várias teorias.	0,83	0,83	0,83
22	O professor apresentava o contexto ou a origem das ideias/conceitos desenvolvidos nas aulas.	0,9	0,9	0,76
23	O professor apresentava outros pontos de vista além do seu, quando apropriado.	0,76	0,83	0,63
24	O professor discutia de forma adequada os avanços na área.	0,83	0,7	0,76
25	Os comentários do professor sobre as provas e trabalhos corrigidos foram importantes.	0,9	0,76	0,83
26	Os métodos de avaliação eram justos e apropriados.	0,9	0,96	0,83
27	Os conteúdos das provas e trabalhos estavam de acordo com os enfatizados pelo professor.	0,76	0,9	0,76
28	O material de estudo e/ou bibliografia recomendados foram importantes.	0,76	0,83	0,76
29	As leituras, tarefas para casa, etc. contribuíram para a compreensão e gosto pelo conteúdo.	0,83	0,83	0,83
30	A dificuldade desta disciplina, em relação a outras, foi: Muito fácil - fácil - Média - Difícil - Muito difícil	0,83	0,83	0,76
31	A exigência/carga de trabalho desta disciplina, em relação às outras disciplinas foi: Muito leve - Leve - Média - Pesada - Muito pesada	0,76	0,63	0,7
32	O ritmo de andamento da disciplina foi: Muito lento - Lento - Médio - Rápido - Muito rápido	0,9	0,83	0,83
33	O número de horas necessárias para as atividades extraclasses foi de: 0 a 2 - 2 a 5 - 5 a 7 - 7 a 12 - Acima de 12	0,83	0,7	0,76
34	O seu nível de interesse na disciplina antes dela iniciar era de: Muito baixo - Baixo - Médio - Alto - Muito alto	0,76	0,7	0,7
<i>CVCt</i>		<i>0,81</i>	<i>0,82</i>	<i>0,76</i>

Fonte: Elaboração própria.

Tabela 4. Avaliação da dimensão teórica

Nº	ITEM	DIMENSÃO*									
		A	B	C	D	E	F	G	H	I	MARSH
1	Você considerou a disciplina intelectualmente desafiadora e estimulante.	1	2								A
2	Você aprendeu algo que considera importante.	2	1								A
3	Seu interesse no conteúdo aumentou como consequência dessa disciplina.	1	1			1					A
4	Você aprendeu e entendeu os materiais de apoio (livros, slides, etc.) desta disciplina.	3									A

*A- Aprendizagem; B- Entusiasmo; C- Organização; D- Interação com o grupo; E- Empatia; F- Amplitude da abordagem; G- Avaliação; H- Atividades extra-classe; I- Sobrecarga/Dificuldade.

Fonte: Elaboração própria.

Tabela 4. Avaliação da dimensão teórica (continuação)

Nº	ITEM	DIMENSÃO*										MARSH
		A	B	C	D	E	F	G	H	I		
5	O professor era entusiasmado em relação a ministrar essa disciplina.		3									B
6	O professor era dinâmico e enérgico na condução da disciplina.	1	1			1						B
7	O professor aprimorava as aulas com uso de humor.					2						B
8	O modo de o professor dar aulas mantinha seu interesse durante todo o período.	1										B
9	As explicações do professor eram claras.	2										C
10	Os materiais didáticos da disciplina estavam bem preparados e foram cuidadosamente explicados.	1		2								C
11	Os objetivos apresentados coincidiam com aqueles realmente ensinados, de forma que eu sabia onde a disciplina estava indo.	1		2								C
12	O professor dava aulas que facilitavam a tomada de notas/anotações.			3								C
13	Os estudantes eram encorajados a participar das discussões nas aulas.				3							D
14	Os estudantes eram convidados a compartilhar suas ideias e conhecimentos.				3							D
15	Os estudantes eram encorajados a perguntar durante as aulas, e as respostas dadas eram relevantes.	1			2							D
16	Os estudantes eram encorajados a expressar suas próprias ideias e/ou a questionar as do professor.			1	2							D
17	O professor era amistoso/cordial com os estudantes, individualmente.					3						E
18	O professor fazia os alunos sentirem-se bem-vindos ao pedir ajuda/conselhos, dentro ou fora da aula.					3						E
19	O professor tinha um interesse sincero nos estudantes, individualmente.		1			1						E
20	O professor era adequadamente acessível aos estudantes durante os horários de atendimento extra-classe.					1			2			E
21	Em relação ao conteúdo, o professor comparava as implicações de várias teorias.	1				1	1					F
22	O professor apresentava o contexto ou a origem das ideias/conceitos desenvolvidos nas aulas.	1				1	1					F
23	O professor apresentava outros pontos de vista além do seu, quando apropriado.					2	1					F
24	O professor discutia de forma adequada os avanços na área.					2	1					F

*A- Aprendizagem; B- Entusiasmo; C- Organização; D- Interação com o grupo; E- Empatia; F- Amplitude da abordagem; G- Avaliação; H- Atividades extra-classe; I- Sobrecarga/Dificuldade.

Fonte: Elaboração própria.

Tabela 4. Avaliação da dimensão teórica (continuação)

Nº	ITEM	DIMENSÃO*									
		A	B	C	D	E	F	G	H	I	MARSH
25	Os comentários do professor sobre as provas e trabalhos corrigidos foram importantes.	1						2			G
26	Os métodos de avaliação eram justos e apropriados.							3			G
27	Os conteúdos das provas e trabalhos estavam de acordo com os enfatizados pelo professor.							3			G
28	O material de estudo e/ou bibliografia recomendados foram importantes.	1		1						1	H
29	As leituras, tarefas para casa, etc. contribuíram para a compreensão e gosto pelo conteúdo.								2		H
30	A dificuldade desta disciplina, em relação a outras, foi: Muito fácil - fácil - Média - Difícil - Muito difícil									3	I
31	A exigência/carga de trabalho desta disciplina, em relação às outras disciplinas foi: Muito leve - Leve - Média - Pesada - Muito pesada									3	I
32	O ritmo de andamento da disciplina foi: Muito lento - Lento - Médio - Rápido - Muito rápido				3						I
33	O número de horas necessárias para as atividades extraclasse foi de: 0 a 2 - 2 a 5 - 5 a 7 - 7 a 12 - Acima de 12									3	I
34	O seu nível de interesse na disciplina antes dela iniciar era de: Muito baixo - Baixo - Médio - Alto - Muito alto			2							I

*A- Aprendizagem; B- Entusiasmo; C- Organização; D- Interação com o grupo; E- Empatia; F- Amplitude da abordagem; G- Avaliação; H- Atividades extra-classe; I- Sobrecarga/Dificuldade.
Fonte: Elaboração própria.

Observou-se que grande parte dos itens apresentou coeficiente de validade de conteúdo (CVC) insatisfatório em pelo menos um critério (CL, PP, RT). O critério de relevância teórica foi o que apresentou mais itens (n=21) com CVC abaixo de 0,80, o que deixou sua média também abaixo do ponto de corte (CVC=0,76). Os critérios de clareza de linguagem e pertinência prática tiveram CVC de 0,81 e 0,82, respectivamente. O CVC do instrumento como um todo foi de 0,76.

A partir destes resultados, o instrumento passou por algumas modificações. Foram excluídos os 14 itens que tiveram dois ou três critérios com classificação abaixo de 0,80 (6, 7, 11, 12, 18, 19, 20, 23, 24, 27, 28, 31, 33, 34). Após essa exclusão, os itens que continham CVC abaixo de 0,80 ou que haviam recebido sugestões de alterações (4, 10, 15, 26 e 29) no critério de clareza de linguagem foram reformulados. A principal observação dos juízes referia-se à necessidade de que cada questão avaliasse somente um atributo, pois, havendo duas situações descritas no mesmo item, uma situação pode ter ocorrido e a outra não, dificultando a resposta do estudante. Para atender as sugestões dos juízes-avaliadores e facilitar a compreensão, três questões passaram por modificações no texto e duas foram divididas em outras duas (tabela 5).

Os itens reformulados foram novamente enviados aos juízes para outra avaliação. No entanto, o cálculo do CVC não foi possível nesta etapa, pois somente dois juízes realizaram essa nova avaliação - o método exige, no mínimo, três. Na Tabela 5 abaixo, entretanto, são apresentadas as avaliações desses dois juízes no critério de clareza de linguagem. Observou-se que somente a questão 10 manteve a mesma avaliação, todas as demais tiveram aumento na escala concordância. Com isso, entendeu-se que os itens melhoraram quanto à clareza de linguagem após as modificações e optou-se por utilizar no instrumento estas versões reformuladas. Os demais critérios de pertinência prática, relevância teórica e dimensão mantiveram-se com a mesma avaliação.

Tabela 5. Avaliação dos itens pelos juízes avaliadores (A1 e A2) antes e depois da reformulação, no critério de clareza de linguagem

ITEM	ITEM	A1	A2	ITEM REFORMULADO	A1	A2
4	Você aprendeu e entendeu os materiais de apoio (livros, slides, etc.) desta disciplina.	3	2	Você entendeu os materiais de apoio (livros, slides, etc.) desta disciplina.	5	3
10	As tarefas para casa contribuíram para a compreensão e gosto pelo conteúdo.	3	3	As tarefas para casa contribuíram para a compreensão pelo conteúdo.	3	3
15	Os métodos de avaliação eram justos e apropriados	3	4	Os métodos de avaliação eram apropriados.	5	5
26	Os estudantes eram encorajados a perguntar durante as aulas, e as respostas dadas eram relevantes.	3	3	Os estudantes eram encorajados a perguntar durante as aulas	5	4
				As respostas do professor, em relação às perguntas feitas pelos estudantes, eram relevantes.	5	3
29	Os materiais didáticos da disciplina estavam bem preparados e foram disciplina foram cuidadosamente explicados.	3	3	Os materiais didáticos da disciplina estavam bem preparados.	5	4
				Os materiais didáticos da disciplina foram cuidadosamente explicados.	5	4

1- pouquíssima concordância; 2- pouca concordância, 3- média concordância, 4- muita concordância e 5- muitíssima concordância (com o critério considerado).

Fonte: Elaboração própria.

Com relação à dimensão teórica, 18 questões tiveram ao menos dois juízes concordando com Marsh (1982). O índice Kappa do questionário foi de 0,544, considerado moderado (Landis e Kock, 1977). Nem todos os juízes avaliaram todas as questões. As de números 7, 9, 19, 29 e 34 foram avaliadas por somente dois juízes e a questão 8 por somente 1. No espaço destinado a observações sobre cada questão, um juiz avaliador relatou dificuldades em optar por uma dimensão nas questões 5, 8, 9, 19 e 34.

Após as adaptações, o questionário passou de 34 para 22 questões, pois foram excluídas 14 devido ao baixo CVC e inseridas 2 por sugestão dos juízes avaliadores. Ao proceder a nova avaliação, o CVC total do questionário foi de 0,80, o CVC do critério clareza de linguagem foi de 0,88, de pertinência prática 0,85 e de relevância teórica 0,81. A escala, enfim, estava apta à aplicação piloto.

4.4. Teste piloto

O tempo de aplicação do AQEE foi de cerca de 20 minutos. Na planilha de avaliação, os estudantes fizeram algumas sugestões de alterações de vocabulário, como substituição da palavra cordial (questão 15); as palavras amistoso/cordial foram substituídas por respeitoso/gentil. As questões 10, 11 e 12 foram relatadas como semelhantes, e a questão 11 acabou sendo excluída. Após as alterações, o questionário foi considerado finalizado com 21 questões objetivas (Anexo 1).

5. Discussão

O instrumento validado AQEE apresentou concordância na tradução e obteve índices de validade de conteúdo satisfatórios para a aplicação no Brasil. Nas etapas da tradução e consolidação da versão preliminar, as alterações de vocabulário eram esperadas devido às diferenças entre os sistemas educativos dos dois países. No entanto, em relação às conjugações verbais, percebeu-se que os tradutores deveriam ter sido comunicados sobre em qual etapa do semestre o instrumento seria aplicado. Esse procedimento poderia ter reduzido as diferenças de conjugações verbais durante a tradução.

A exclusão de 14 itens do instrumento melhorou o CVC do instrumento, que passou de 0,76 para 0,80, e também dos seus critérios de CL, PP e RT. Essa exclusão, todavia, não deixa de ser um fato preocupante, pois determinadas dimensões perderam boa parte dos seus itens. A dimensão empatia, por exemplo, passou de quatro para somente dois itens. A falta de itens para mensurar cada dimensão é um problema considerado crucial na validação de instrumentos (Haynes, Richard e Kubany, 1995) e que deverá ser melhor investigado quando forem realizadas as avaliações psicométricas da escala. Para evitar um maior número de exclusões, foi definido que seriam mantidos os itens que continham somente um critério CVC abaixo de 0,80. Ressalta-se que essa alteração no método foi definida durante o processo e teve por objetivo não alterar demais o instrumento original.

No intuito de reduzir ainda mais o número de questões a serem excluídas, chegou-se a discutir duas possibilidades, que foram rejeitadas posteriormente. A primeira seria relativizar o ponto de corte, uma alternativa sugerida quando os avaliadores possuem formações muito distintas (Cassepp, Borges e Teodoro, 2010). No entanto, preferiu-se manter no trabalho o valor de 0,80 tendo em vista que já havia sido feita a alteração descrita anteriormente, de manter itens que tinham somente um critério com CVC abaixo do ponto de corte. Outra alternativa seria reformular as questões com baixo CVC no critério clareza de linguagem e enviar para os juízes avaliadores antes de serem excluídas. Dessa forma, algumas questões (6, 11, 23, 27 e 28), caso obtivessem uma avaliação mais elevada, poderiam ter sido mantidas no instrumento. A justificativa para não proceder desta forma deveu-se ao fato de que as questões excluídas já tinham outro critério com CVC abaixo de 0,80, e, na metodologia, são enviadas para uma nova avaliação somente os itens com CVC baixo em um único critério, o de clareza de linguagem. As observações e considerações dos juízes foram essenciais para refinar o instrumento, pois, mesmo questões que obtiveram CVC adequado, obtiveram sugestões de melhorias em relação à compreensão e clareza.

Na avaliação das dimensões, a análise do Kappa demonstrou que houve concordância moderada entre os juízes avaliadores e o autor da escala (Marsh, 1984, 1991). Acredita-se que esse resultado deva-se a um número de dimensões elevadas (9), que dificulta a classificação dos itens. É necessário considerar, também que, em se tratando da área do ensino, alguns itens podem ser interpretados como pertencendo a mais de uma dimensão, pois muitos temas estão interligados. Para citar apenas um exemplo, no instrumento original a questão 9 (“As explicações do professor eram claras”) encontra-se dentro da dimensão organização (Marsh, 1984), mas neste trabalho foi classificado como pertencendo à dimensão aprendizagem pelos três juízes avaliadores. Ou seja, os juízes concordaram entre eles, mas não concordaram com o autor da escala. Em casos como este, quando há constructos bastante relacionados, pode-se oferecer aos juízes uma tabela nas quais ele possa marcar mais de uma dimensão. Porém, conforme relatado por Pasquali (1999), essa ação, apesar de poder tornar a avaliação mais adequada, pode abrir espaço para que sejam atribuídas muitas dimensões para itens relacionados a apenas um fator, o que dificulta a avaliação. Dessa forma, foi optado pela planilha com somente uma opção de resposta e era previsto que ocorressem situações como esta. Os itens, no entanto, não foram modificados por tais diferenças de classificação.

A discussão sobre avaliação do ensino superior é polêmica e atual. Nos Estados Unidos da América (EUA), o tema é discutido desde a década de 20 e hoje é tratado de forma bem objetiva e analítica, com indicadores de qualidade sendo constantemente revisados (Lucas, Gibbs, Hughes, Jones e Wisker, 1997; Ramsden, 1991). Na Europa, a avaliação passou a ter um papel de destaque desde a Declaração de Bolonha, pois ela trouxe a necessidade de definição de critérios e metodologias que pudessem ser comparados entre os países constituintes da União Europeia (Declaração de Bolonha, 1999). No Brasil, o SINAES vem passando por estudos e avaliações desde a sua implementação (Sobrinho, 2010).

Embora seja impossível separar a dupla professor-ensino, necessário lembrar que a atividade do professor é influenciada pelas suas condições de ensino. Darling-Hammond (2007) define que a qualidade do ensino é resultado da qualidade do professor somada às condições de ensino. Ou seja, se um professor de alta qualidade tiver muita carga horária, ou ministrar aulas em áreas distintas da sua especialidade, não haverá condições para que ele exerça a sua alta qualidade. As más condições de ensino podem gerar resultados insatisfatórios numa avaliação, penalizando o professor, por isso, quando realizada, a avaliação feita pelos alunos sobre o trabalho realizado dentro da sala necessita ser ponderada pelas reais condições a que o professor trabalha.

Qualquer tipo de avaliação - seja ela feita pelos alunos ou pelos professores - não deve ser única no processo. Para que possuam credibilidade, elas devem incluir tanto características somativas quanto formativas, pois ambas possuem seus limites de atuação (Depresbiteris, 2008; Fernandes, 2008; Sobrinho, 2004). É sabido que avaliações do tipo quantitativa são alvo constante de críticas devido à dificuldade de definição de critérios numa área tão ampla quanto a educação. Porém, avaliações do tipo qualitativa também possuem suas limitações, visto que seu objetivo, de permitir a livre expressão de práticas e ideias que envolvem o ensino, pode levar a uma ausência de critérios mensuráveis e inviabilizar, assim, uma avaliação comparativa (Herrera et al., 2012). A difícil definição de quais critérios serão incluídos não pode ser um fator impeditivo para a realização das

avaliações, afinal, os alunos devem ser ouvidos de forma oficial pelos cursos de graduação (Bobbio, 2012).

Além do SEEQ, há outros SETs (*Student Evaluation of Teaching*) utilizados internacionalmente que podem auxiliar o trabalho de avaliação, como o *Course Experience Questionnaire* (CEQ) (Ramsden, 1991), o *Module Experience Questionnaire* (MEQ) (Lucas et al., 1997) e o *Experiences of Teaching and Learning Questionnaire* (ETLQ) (Entwistle, McCune e Hounsell, 2002). Todos são multidimensionais e resultaram de pesquisas tanto na literatura quanto com estudantes sobre as características desejáveis para um bom ensino ou um bom professor. O SEEQ, entretanto, é o instrumento mais utilizado nos EUA (Coffey and Gibs, 2001) e mais amplamente pesquisado no mundo (Richardson, 2005), tendo sido validado em diversos países (Mijic, 2010; Morais, Almeida e Montenegro, 2006; Nash, 2012).

O SEEQ, assim como qualquer instrumento, possui suas limitações. Há pesquisadores que questionam a capacidade dos SETs para propósitos somativos e formativos dentro da avaliação (Spooren, Brockx e Mortelmans, 2013). Também, Marsh (2007) demonstrou que os SETs, sozinhos, não são insuficientes para melhorar a efetividade do ensino ao longo do tempo. As inconsistências observadas nesse tipo de avaliação têm sido estudadas para que sejam buscadas novas abordagens na avaliação (Hammond, 2012). No entanto, a dificuldade de definição de critérios ou os limites de abrangência dos SETs não devem ser utilizadas como justificativa para a ausência de avaliação. Novamente, é necessário lembrar que os estudantes necessitam ter voz dentro dos seus cursos.

Conforme afirma o pesquisador brasileiro Sobrinho (2010), “sempre haverá contradições e imperfeições num sistema de avaliação, pois se trata de um fenômeno social e de uma construção histórica” (p. 223). Ele diz, ainda, que uma avaliação não é neutra nem, tampouco, detentora da verdade absoluta, mas que precisa se esforçar para ser justa e socialmente eficaz. Entende-se que uma avaliação do ensino não deve ser constituída por um único instrumento. Cada um que fizer parte desse amplo processo deve ser cuidadosamente estudado e cumprir, assim, com a sua parcela na avaliação do ensino.

6. Considerações finais

A avaliação do ensino é um tema com muitas afinidades e contradições. Assim como a maioria concorda com a importância da sua realização, não há consenso sobre critérios e métodos a serem utilizados. No caso da avaliação feita pelos alunos, há outros instrumentos além do SEEQ que podem ser utilizados parcial ou totalmente dentro dos cursos e/ou instituições. O SEEQ, entretanto, é um dos mais extensamente estudados internacionalmente, e pode auxiliar os cursos de graduação na avaliação da qualidade do ensino.

Neste trabalho, o método utilizado mostrou-se adequado para adaptar o instrumento entre as culturas. Após a tradução e validação de conteúdo, tanto o CVC total do instrumento quanto o dos seus critérios -clareza de linguagem, pertinência prática, relevância teórica- ficaram acima de 0,80, demonstrando que o instrumento está apto à aplicação em pesquisas dentro do Brasil. Futuramente, é importante que sejam

investigadas as características psicométricas da escala, a fim de que ela possa ser utilizada pelas universidades brasileiras para avaliar a qualidade do ensino ministrado.

Referências bibliográficas

- Bobbio, N. (1992). *A era dos direitos 11*. São Paulo: Campus.
- Cassep-Borges, V., Balbinotti, M. A. A. e Teodoro, M. L. M. (2010). Tradução e validação de conteúdo: Uma proposta para a adaptação de instrumentos. Em L. Pasquali (Ed.), *Instrumentação psicológica: Fundamentos e prática* (pp. 506-520). Porto Alegre: Artmed.
- Chagas, A. T. R. (2000). O questionário na pesquisa científica. *Administração online*, 1(1). Disponível em http://www.fecap.br/adm_online/art11/anival.htm
- Chonko, L. B., Tanner, J. F. e Davis, R. (2002). What are they thinking? Students' expectations and self-assessments. *Journal of Education for Business*, 77(5), 271-281. <https://doi.org/10.1080/08832320209599676>
- Cohen, P. A. (1980). Effectiveness of student rating feedback for improving college instruction. A metaanalysis of multisection validity studies. *Research in Higher Education*, 13(4), 321-341. <https://doi.org/10.1007/bf00976252>
- Darling-Hammond, L. (2007). Recognizing and enhancing teacher effectiveness: A policy maker's guide. Em L. Darling-Hammond e C. D. Prince (Eds.), *Strengthening teacher quality in high-need schools-policy and practice* (pp. 1-26). Nova Iorque, NY: Council of Chief State School Officers.
- Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E. e Rothstein, J. (2012). Evaluating teacher evaluation. *The Phi Delta Kappa*, 93(6), 8-15. <https://doi.org/10.1177/003172171209300603>
- Declaração de Bolonha. Declaração conjunta dos ministros da educação europeus, assinada em Bolonha, de 19 de Junho de 1999. Disponível em http://media.ehea.info/file/Ministerial_conferences/05/3/1999_Bologna_Declaration_Portuguese_553053.pdf
- Depresbiteris, L. (2005). Auto-avaliação das instituições de ensino superior. *Estudos em Avaliação Educacional*, 16(32), 7-24. <https://doi.org/10.18222/ea163220052133>
- Entwistle, N., McCune, V. e Hounsell, J. (2002). Approaches to studying and perceptions of university teaching-learning environments: Concepts, measures and preliminary findings. *Enhancing Teaching and Learning Environments in Undergraduate Courses Occasional Report*, 1.
- Feldman, K. A. (2007). Identifying exemplary teachers and teaching: Evidence from student ratings. Em R. P. Perry e J. C. Smart (Eds.), *The scholarship of teaching and learning in higher education: An evidence-based perspective* (pp. 93-143). Dordrecht: Springer.
- Fernandes, D. (2008). *Avaliação do desempenho docente: Desafios, problemas e oportunidades*. Lisboa: Texto Editores.
- Fernández, J. M., Sánchez, J. F. e García, M. A. M. (1996). Valoración por parte del profesorado de la evaluación docente realizada por los alumnos. *Psicothema*, 8(1), 167-172.
- Haynes, S. N., Richard, D. e Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological assessment*, 7(3), 238-247. <https://doi.org/10.1037//1040-3590.7.3.238>

- Hernández-Nieto, R. A. (2002). *Contribuciones al análisis estadístico*. Mérida: Universidad de Los Andes/IESINFO.
- Herrera, F., Cano, M. A, Hernáñez, L. e Morales, M. A. (2012). Sistema de evaluación del desempeño profesional docente para las licenciaturas del área de las ciencias biológicas y de la salud. *Revista Iberoamericana de Evaluación Educativa*, 5(2), 124-146.
- Keane, E. e Labhrainn, I. M. (2005). Obtaining student feedback on teaching & course quality. *Briefing paper*, 2, 1-9.
- Landis, J. R. e Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159-174. <https://doi.org/10.2307/2529310>
- Lei nº 10.861, de 14 de Abril de 2004. Institui o Sistema Nacional de Avaliação da Educação Superior – SINAES. Diário Oficial da União, Seção 1, de 15 de abril de 2004, p. 3.
- Lucas, L., Gibbs, G., Hughes, S., Jones, O. e Wisker, G. (1997). A study of the effects of course design features on student learning in large classes at three institutions: A comparative study. Em C. Rust e G. Gibbs (Eds.), *Improving student learning: Improving student learning through course design* (pp. 10-24). Oxford: Oxford Centre for Staff and Learning Development.
- Marsh, H. W. e Hocevar, D. (1991). The multidimensionality of students' evaluations of teaching effectiveness: The generality of factor structures across academic discipline, instructor level, and course level. *Teaching and Teacher Education*, 7(1), 9-18. [https://doi.org/10.1016/0742-051x\(91\)90054-s](https://doi.org/10.1016/0742-051x(91)90054-s)
- Marsh, H. e Roche, L. (1992). The use of student evaluations of university teaching in different settings: The applicability paradigm. *Australian Journal of Education*, 36(3), 278-300. <https://doi.org/10.1177/000494419203600305>
- Marsh, H. W. (1984). Student's evaluations of university teaching: Dimensionality, reliability, validity, potential biases and utility. *Journal of Educational Psychology*, 76(5), 707-754. <https://doi.org/10.1037//0022-0663.76.5.707>
- Marsh, H. W. (2007). Do university teachers become more effective with experience? A multilevel growth model of students' evaluations of teaching over 13 years. *Journal of Educational Psychology*, 99(4), 775-790. <https://doi.org/10.1037/0022-0663.99.4.775>
- Mattar, F. e Motta, S. (1999). *Pesquisa de marketing*. São Paulo: Elsevier.
- Mijic, D. (2010). Measuring teaching quality in higher education: Instrument for collecting student feedback. En M. Gusev (Ed.), *ICT Innovations Web Proceedings* (pp. 117-128). Disponível em <http://ictinnovation.org>
- Morais, N., Almeida, L. e Montenegro, I. (2006). Percepções do ensino pelos alunos: Uma proposta de instrumento para o ensino Superior. *Análise Psicológica*, 1(24), 73-86.
- Nash, J. L. (2012). *Using student evaluations at a Cambodian University to improve teaching effectiveness* (Dissertação). Lehigh University, Bethlehem.
- Ramsden, P. (1991). A performance indicator of teaching quality in higher education: The course experience questionnaire. *Studies in Higher Education*, 16(2), 129-150. <https://doi.org/10.1080/03075079112331382944>
- Richardson, J. T. (2005). Instruments for obtaining student feedback: A review of the literature. *Assessment & Evaluation in Higher Education*, 30(4), 387-415. <https://doi.org/10.1080/02602930500099193>

Sistema Nacional de Avaliação do Ensino Superior (SINAES). (2003). *Bases para uma nova proposta de avaliação do ensino superior*. São Paulo: Ministério da Educação.

Spooren, P., Brockx, B. e Mortelmans, D. (2013). On the validity of student evaluation of teaching the state of the art. *Review of Educational Research*, 83(4), 598-642. <https://doi.org/10.3102/0034654313496870>

Sobrinho, J. D. (2004). Avaliação ética e política em função da educação como direito público ou como mercadoria. *Educação e Sociedade*, 25(88), 703-725. <https://doi.org/10.1590/s0101-73302004000300004>

Sobrinho, J. D. (2010). Avaliação e transformações da educação superior brasileira (1995-2009): Do provão ao SINAES. *Avaliação: Revista da Avaliação da Educação Superior*, 15(1), 195-224. <https://doi.org/10.1590/s1414-40772010000100011>

Vallerand, R. J. (1989). Vers une méthodologie de validation transculturelle de questionnaires psychologiques: Implications pour la recherche en langue française. *Psychologie Canadienne*, 30(4), 662-680. <https://doi.org/10.1037/h0079856>

Anexo I

Instruções

As três questões abaixo referem-se à disciplina em curso. Deixe em branco apenas se ela não for relevante.

Por favor, indique características importantes da disciplina e/ou do professor que você acha que foram relevantes para o seu aprendizado.

Por favor, indique aspectos desta disciplina e/ou do professor que você considera que poderiam ser melhorados.

Por favor, use este espaço adicional para explicar melhor alguma de suas respostas ou para fazer outro comentário.

Instruções

O quadro abaixo apresenta, na primeira coluna, diversas afirmações sobre a disciplina.

Responda, numa escala de 1 (discordo totalmente) a 5 (concordo totalmente), qual o seu grau de concordância com cada uma das afirmações seguintes. Deixe em branco apenas se ela não for relevante.

	Discordo totalmente	Discordo	Não concordo nem discordo	Concordo	Concordo totalmente
Aprendizagem					
Você achou a disciplina desafiadora e estimulante intelectualmente.	1	2	3	4	5
Você aprendeu algo que considera importante.	1	2	3	4	5
Seu interesse no conteúdo aumentou como consequência desta disciplina.	1	2	3	4	5
Você entendeu os materiais de apoio (slides, livros, etc.) desta disciplina	1	2	3	4	5
Entusiasmo					
O professor era entusiasmado em relação a ministrar esta disciplina.	1	2	3	4	5
O modo de o professor dar aulas mantinha seu interesse durante todo o período.	1	2	3	4	5
Organização					
As explicações do professor eram claras.	1	2	3	4	5
Os materiais didáticos da disciplina estavam bem preparados.	1	2	3	4	5
Os materiais didáticos da disciplina foram cuidadosamente explicados.	1	2	3	4	5
Interação do grupo					
Os estudantes eram encorajados a participar das discussões nas aulas.	1	2	3	4	5
Os estudantes eram encorajados a perguntar durante as aulas.	1	2	3	4	5
As respostas do professor, em relação às perguntas feitas pelos estudantes, eram relevantes.	1	2	3	4	5
Os estudantes eram encorajados a expressar suas próprias ideias e/ou a questionar as do professor.	1	2	3	4	5
Empatia					
O professor era amigável/cordial com os estudantes, individualmente.	1	2	3	4	5
Amplitude na abordagem dos assuntos					
Em relação ao conteúdo, o professor comparava as implicações de várias teorias.	1	2	3	4	5
O professor apresentava o contexto ou a origem das ideias/conceitos desenvolvidos nas aulas.	1	2	3	4	5
	Discordo totalmente	Discordo	Não concordo nem discordo	Concordo	Concordo totalmente
Avaliação da aprendizagem					
Os comentários do professor sobre as provas e trabalhos corrigidos foram importantes.	1	2	3	4	5
Os métodos de avaliação da aprendizagem eram apropriados.	1	2	3	4	5
Atividades/Atribuições					
As tarefas para casa contribuíram para a compreensão do conteúdo.	1	2	3	4	5
Sobrecarga/Dificuldade					
A dificuldade desta disciplina, em relação a outras, foi: 1) Muito fácil 2) Fácil 3) Média 4) Difícil 5) Muito difícil	1	2	3	4	5
O ritmo de andamento da disciplina foi: 1) Muito lento 2) Lento 3) Médio 4) Rápido 5) Muito rápido	1	2	3	4	5

Dados do respondente

Curso: _____

Instituição: _____

Disciplina: _____ Data: _____

Idade: _____ anos Sexo: () Feminino () Masculino

Formação no ensino médio: () Pública () Particular

Ano de finalização do ensino médio (ou equivalente): _____

Exerce atividade remunerada? () Sim () Não

Breve CV dos autores

Joice Trindade Silveira

Docente do curso de Nutrição na Universidade Federal do Pampa (UNIPAMPA), na área de Gestão de Unidades de Alimentação e Nutrição e Deontologia. Possui graduação em Nutrição e Mestrado em Microbiologia Agrícola e do Ambiente pela Universidade Federal do Rio Grande do Sul (UFRGS). É estudante de doutorado no PPG Educação em Ciências - Química da vida e saúde na UFRGS, desenvolvendo trabalhos na área de avaliação de disciplinas e professores. Email: joicetsilveira@gmail.com

João Batista Teixeira da Rocha

Possui Graduação em Ciências Biológicas pela Universidade Federal do Rio Grande do Sul - UFRGS (1986), e Doutorado em Ciências Biológicas (Bioquímica) pela Universidade Federal do Rio Grande do Sul - UFRGS (1996). Realizou seu Pós-Doutorado na UFRJ, Departamento de Bioquímica Médica no Laboratório de Bioenergética coordenado pelo Prof. Leopoldo de Meis (1997-1998). Atualmente trabalha na área de bioquímica, toxicologia e farmacologia de organocalcogênio, papel do estresse oxidativo em patologias humanas e experimentais e educação em ciências. Têm experiências na área de toxicologia, sistema nervoso central (incluindo psicofarmacologia e desenvolvimento), farmacologia de organocalcogênios e interação entre calcogênios e tióis. Participa de atividade relacionados ao ensino de ciências e como melhorar o ensino de ciências por meio da interação entre cientistas e estudantes e professores do ensino médio e fundamental. ORCID ID: 0000-0003-3829-0595. Email: jbrocha@yahoo.com.br