

INSTRUMENTOS COM EVIDÊNCIAS PSICOMÉTRICAS PARA AVALIAR O DESEMPENHO DE PROFESSORES UNIVERSITÁRIOS: REVISÃO DE LITERATURA

Joene Vieira-Santos

Centro Universitário Adventista de São Paulo, UNASP, Hortolândia, SP

Marcelo Henrique Oliveira Henklain

Universidade Federal de Roraima, UFRR, Boa Vista, RR

Resumo

O objetivo deste estudo foi identificar instrumentos brasileiros com evidências psicométricas para avaliação do desempenho de professores universitários. Foi realizada revisão de literatura a partir do Scielo, Pepsic e Portal de Periódicos da CAPES. Dois grupos de descritores foram utilizados: (a) escala, inventário, questionário, instrumento, checklist e avaliação; (b) professor e docente. Nas bases foi utilizado um par de descritores por vez, associados pelo operador lógico AND. Foram encontrados 24 instrumentos que, de modo geral, apresentaram o Alfa de Cronbach como evidência de fidedignidade e evidências de validade relacionadas ao conteúdo e à estrutura interna. Estudos futuros sobre construção e adaptação de medidas psicométricas para avaliar o desempenho de professores devem definir com precisão os construtos e ampliar as evidências psicométricas investigadas, envolvendo estudos de validade de critério. Este estudo contribuiu com a identificação dos instrumentos que cientistas e educadores têm a disposição no Brasil, bem como com a proposição de investigações que são necessárias para o aperfeiçoamento da avaliação de professores universitários.

Palavras-chave: Professores; Instrumentos; Avaliação do desempenho docente; Ensino Superior; Evidências psicométricas.

Instruments with psychometric evidence to evaluate college teachers' performance: literature review

Abstract

The aim of this study was to identify Brazilian instruments to assess the performance of college teachers. A literature review was performed from the Scielo, Pepsic and Portal of Periodicals of CAPES. Two groups of descriptors were used: (a) scale, inventory, questionnaire, instrument, checklist, and assessment; (b) professor and teacher. In the databases a pair of descriptors was used at a time, associated by the logical operator AND. We found 24 instruments that, in general, presented the Cronbach's Alpha as evidence of reliability and evidence of validity related to content and internal structure. Future studies on the construction and adaptation of psychometric measures to assess teacher performance must define precisely the constructs and expand the investigated psychometric evidence, involving studies of criterion validity. This study contributed to the identification of the instruments that scientists and educators have at their disposal in Brazil, as well as to the proposition of investigations that are necessary to improve the evaluation of college teachers.

Keywords: Faculty; Instruments; Teacher performance assessment; Higher Education; Psychometric evidence.

Instrumentos con evidencias psicométricas para evaluar el desempeño de profesores universitarios: revisión de literatura

Resumen

El objetivo de este estudio fue identificar instrumentos brasileños para la evaluación del desempeño de docentes universitarios. Se realizó una revisión de literatura de Scielo, Pepsic y Periódicos CAPES. Se utilizaron dos grupos de descriptores: (a) escala, inventario, cuestionario, instrumento, lista de verificación y evaluación; (b) profesor y docente. En las bases, se utilizó un par de descriptores a la vez, asociados por el operador lógico AND. Fueron encontrados 24 instrumentos que generalmente presentaban el alfa de Cronbach como evidencia de confiabilidad y evidencia de validez relacionada con el contenido y la estructura interna. Los estudios futuros sobre la construcción y adaptación de Medidas psicométricas para evaluar el desempeño docente deben definir con precisión los constructos y expandir la evidencia psicométrica investigada, involucrando estudios de validez de criterio. Este estudio contribuyó a la identificación de los instrumentos que los científicos y educadores tienen a su disposición en Brasil, así como a la propuesta de investigaciones que son necesarias para mejorar la evaluación de los profesores universitarios.

Palabras clave: Maestros; Instrumentos; Evaluación del desempeño docente; Enseñanza superior; Evidencia psicométrica.

O professor universitário pode e, nas universidades públicas brasileiras, deve desempenhar atividades que extrapolam aquelas específicas de sala de aula (Portaria nº 554, 2013). Exemplos dessas atividades incluem a condução de pesquisas científicas, a proposição e gestão de projetos de extensão, a participação na administração da universidade, a elaboração de pareceres para aprovação de projetos, artigos ou avaliação de cursos, dentre outros. Apesar dessa multiplicidade de tarefas, ensinar continua a ser uma das principais atividades do professor e um dos maiores desafios da carreira (Komarraju, 2013). E o que significa ensinar no nível superior?

Essa é uma questão complexa e cujo exame foge ao escopo deste estudo. Genericamente, é possível propor, com base em Gusso (2013), que se trata de um tipo de ensino que ocorre em contextos específicos (que variam em função das instituições de ensino), diferentes daqueles característicos, por exemplo, da educação básica, e que deve gerar resultados igualmente específicos. O contexto, geralmente, inclui alunos de diferentes idades (geralmente jovens adultos) e com repertórios diversos, grande volume e complexidade dos objetivos de ensino que devem ser alcançados em um curto espaço de tempo e muita concorrência entre atividades diversas que precisam ser realizadas por professores (exemplificadas no parágrafo anterior) e alunos (p. ex., iniciação científica, participação em congressos e projetos de extensão, envolvimento no movimento estudantil, etc.). Os resultados esperados incluem capacitar os alunos a dominar processos de conhecer e os conhecimentos gerais da profissão, transformar conhecimentos em ações, cooperar e compartilhar conhecimentos, objetivos e ações, desenvolver habilidades relacionadas a raciocínio lógico, sentido estético, capacidades físicas, comunicação, memória e demais habilidades favorecedoras da autonomia.

Em resumo, o que parece crucial no ensino de nível superior é criar condições para que os alunos aprendam não apenas conteúdos, mas como agir na sociedade em função do conhecimento científico disponível, de modo a lidar eficazmente com certa classe de fenômenos característicos de uma profissão (Gusso, 2013). Mas o que o professor deve fazer para arranjar condições facilitadoras desse tipo de aprendizado?

Pesquisas em áreas como a Psicologia e a Educação têm tentado responder a essas questões (p. ex., Alves, 2007; Cândido, Assis, Ferreira, & Souza, 2014; Komarraju, 2013). Cândido et al. Souza (2014), por exemplo, solicitaram que 294 universitários de diversos cursos completassem a seguinte frase “Para mim, um bom professor é aquele que...”. As respostas foram classificadas em duas categorias: técnica do professor (envolvendo didática, conhecimento, busca de aperfeiçoamento e motivação) e relação professor-aluno (envolvendo boa interação com os discentes, preocupação com os alunos e atributos pessoais positivos). Um dos aspectos mais interessantes desses estudos é que seus resultados podem ser utilizados como respostas para uma questão atual e urgente: que critérios eleger para avaliar professores de nível superior?

Por um lado, o estabelecimento de critérios é necessário para viabilizar a avaliação que, por sua vez, é necessária para que o professor tenha informações sobre como aperfeiçoar o seu desempenho (Cassettari, 2014). Por outro, surgem questões como: quais critérios adotar para avaliar um profissional com múltiplas funções e, por vezes, com poucas condições adequadas de trabalho? Que interesses fundamentam as avaliações e que objetivos se espera alcançar por meio dessa prática? (Ribeiro, 2012). Não há como se esquivar do enfrentamento dessas questões porque já existe previsão legal para que os professores de nível superior sejam avaliados (ver Lei n. 10.861/2004 que institui o Sistema Nacional de Avaliação da Educação Superior, SINAES).

O que parece viável é identificar, examinar e criticar (de modo propositivo) as formas de avaliação existentes, bem como propor novos objetivos e modelos de avaliação e testar os resultados que produzem. Ao examinar as avaliações em vigor, parece existir certa ênfase sobre aspectos do desempenho do professor baseados em indicadores de produção de artigos, orientações concluídas, dentre outros critérios dessa natureza (Vosgerau, Orlando, & Meyer, 2017). Embora sejam critérios relevantes, evidências sugerem que eles não estão associados com o desempenho do professor para ensinar (Palazzo & Gomes, 2009). Ou seja, publicar muito, por exemplo, não parece estar associado a quanto o professor é capaz de ensinar, motivo pelo qual indicadores e medidas específicas são necessários para

avaliar os resultados que professores e alunos conseguem alcançar em termos de aprendizado.

A partir desse contexto duas perguntas são relevantes: (a) Quais instrumentos estão disponíveis no Brasil para auxiliar nesse tipo de avaliação sobre o desempenho do professor universitário para ensinar? (b) O quanto eles medem os construtos a que se propõem com a menor margem de erro possível? Responder a essas perguntas envolve avaliar a qualidade do instrumento. Nas áreas de Educação e Psicologia, o *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014) é uma das principais referências que apresenta diretrizes para construir medidas e investigar evidências psicométricas.

Os *Standards* definem validade como o “grau em que evidência e teoria suportam as interpretações de pontuações do teste para usos propostos de testes” (American Educational Research Association et al., 2014, p. 11, tradução livre). O processo de validação envolve cinco fontes: (a) *Evidências baseadas no conteúdo do teste*, que se referem a relação entre o conteúdo do instrumento e o construto que pretende medir; (b) *Evidências baseadas na estrutura interna*, que consiste no grau em que as relações entre os itens e as dimensões do instrumento estão de acordo com o construto em que as interpretações do score do instrumento estão baseadas; (c) *Evidências baseadas nas relações com outras variáveis*, que corresponde ao grau em que as relações com outras variáveis relevantes em relação ao construto que o instrumento avalia são consistentes com as interpretações do score da medida; (d) *Evidências baseadas no processo de resposta*, que são relacionadas aos processos cognitivos envolvidos na tarefa de responder ao instrumento; e (e) *Evidências baseadas nas consequências da testagem*, que referem-se ao quão consistentes são as interpretações propostas para os usos pretendidos do instrumento.

Também é importante verificar o quanto as pontuações obtidas na aplicação do teste são confiáveis, ou seja, ao grau em que a medida é realizada sem erro (Zanon & Huck, 2015). Os *Standards* (American Educational Research Association et al., 2014) sugerem três amplas categorias de avaliação da confiabilidade (ou precisão): (a) coeficientes resultantes da aplicação de formas alternativas e equivalentes em sessões de testagem

independentes (denominado de formas paralelas); (b) coeficientes obtidos pela aplicação da mesma forma do teste, mas em ocasiões diferentes (denominado de processo de teste-reteste ou estabilidade temporal); e (c) coeficientes baseados na correlações entre as pontuações em itens individuais ou subconjuntos de itens dentro de um teste (conhecido como coeficientes de consistência interna). Os *Standards* também indicam que é possível estimar a precisão do instrumento através de altos índices de concordância entre avaliadores expertos sobre a adequação dos itens.

Considerando o exposto, este estudo consiste em uma revisão da literatura nacional que busca identificar quantos são e quais as características dos instrumentos que avaliam o desempenho de professores universitários, de instituições públicas e privadas, em relação ao ensino e que apresentam evidências de validade e confiabilidade para a população brasileira. A expressão *desempenho do professor universitário para ensinar* diz respeito a comportamentos apresentados pelo professor universitário no contexto de sala de aula e/ou na relação com os alunos que contribuam para a aprendizagem dos estudantes, bem como para o seu engajamento no estudo. Esses comportamentos do professor estão relacionados, por exemplo, à manutenção de disciplina, exposição e discussão de conteúdo, mediação das relações entre aluno-aluno e aluno-conhecimento e organização do ambiente físico, social e afetivo.

Método

Este estudo é uma revisão de literatura que, a partir de uma pergunta formulada de forma clara, utiliza procedimentos sistemáticos para selecionar e avaliar criticamente pesquisas científicas (Galvão, Pansani, & Harrad, 2015). Os procedimentos realizados foram agrupados em três fases.

Fase I – Busca de artigos sobre medidas para avaliar o desempenho do professor. As buscas por artigos foram realizadas durante os meses de novembro e dezembro de 2017 e uma atualização do banco foi realizada em fevereiro de 2020. As seguintes bases de dados foram consultadas: Scientific Electronic Library Online - Scielo (www.scielo.org), Portal de Periódicos

Eletrônicos de Psicologia - Pepsic (<http://pepsic.bvsalud.org/>) e Portal de Periódicos da Capes (<http://www.periodicos.capes.gov.br>). Dois grupos de descritores foram utilizados. Relacionados a medidas: escala, inventário, questionário, instrumento, checklist e avaliação. O segundo grupo referia-se ao público-alvo: professor e docente.

Em cada base de dados foi inserido um par de descritores por vez, associados pelo operador lógico AND (e.g., escala AND professor). Nas duas primeiras bases, as buscas foram restritas aos campos “resumo” e “título”. Já no Portal da Capes, a restrição foi apenas ao campo “títulos”, pois nessa base não constava o campo “resumo” para a busca avançada.

Foi utilizado o *software* Microsoft Excel® 2010 para a construção de um banco de dados. Os resultados obtidos foram armazenados em tabelas que relacionavam a base de dados consultada, o descritor utilizado e os resultados encontrados.

Fase II – Seleção de artigos. Após remover todos os estudos duplicados, o banco de dados foi analisado, independentemente, por dois pesquisadores com base em critérios de inclusão e exclusão. Os critérios de inclusão foram: (a) estudo que empregasse na coleta de dados instrumento(s) que avalie(m) a frequência ou grau de concordância em relação a ocorrência do, ou relevância do ou satisfação com o desempenho do professor universitário em sala e/ou na relação com seus alunos; (b) estudo desenvolvido com amostras de professores e/ou estudantes brasileiros de nível superior; e (c) estudo que indicasse, pelo menos, uma evidência de validade ou de fidedignidade. Os critérios de exclusão foram: (a) artigo publicado em outros veículos que não periódicos científicos; (b) instrumentos que avaliavam, além do desempenho do professor, outros aspectos, tais como disciplina, estrutura organizacional, etc.; (c) estudos que examinavam características e/ou efeitos de políticas de avaliação de professores ou que propunham métodos ou técnicas de avaliação do desempenho do professor, mas que não tenham empregado um instrumento psicométrico cuja finalidade fosse a de avaliar o desempenho do professor tal como definido na seção de introdução; (d) estudos que descreviam métodos de coleta de dados diferente do uso de instrumentos psicométricos (e.g., entrevistas, portfólios); (e) estudos que

analisavam instrumentos produzidos para outros contextos culturais que não o brasileiro.

Cada pesquisador leu todos os títulos dos artigos e, com base nos critérios de inclusão e exclusão, aqueles que faziam referência explícita a estudos sobre desempenho do professor de nível superior foram selecionados. Para os casos de títulos que não evidenciavam o tipo de estudo realizado, foram lidos os resumos para tomar a decisão de mantê-los ou não na amostra de artigos. Quando mesmo o resumo era insuficiente, foram consultados objetivo e método do estudo para avaliar se o artigo deveria ou não ser selecionado.

Um segundo levantamento de artigos foi realizado nas referências dos artigos selecionados. O mesmo procedimento descrito acima foi aplicado (título, resumo e artigo) para verificar se o artigo atendia aos critérios de inclusão. Aqueles que atendiam tais critérios, que ainda não estavam no banco e que cujo texto completo estava disponível, passaram a compor a amostra final desta pesquisa.

Fase III – Extração de dados. Os procedimentos de análise foram realizados em função das seguintes perguntas: (a) Quantos e quais instrumentos existem sobre o desempenho do professor universitário com evidências psicométricas para a população brasileira? (b) Quais são os aspectos do desempenho do professor que essas medidas avaliam e que tipos de respostas adotam (e.g., escala Likert)? (c) Quem responde a esses instrumentos e de que modo?; (d) Quais as evidências de validade e de fidedignidade que apresentam? Para responder a essas questões, as seguintes informações foram extraídas: construto avaliado (segundo os autores do artigo), forma de resposta solicitada, respondente, forma de aplicação (individual ou coletiva), meio de aplicação (impresso ou *online*) e tipos de evidência de validade e de fidedignidade reportadas.

Resultados e discussão

Ao final das buscas foram encontrados 1931 artigos. Após a remoção de duplicações, restaram no banco de dados 1283. A aplicação dos critérios de inclusão e exclusão resultaram em uma amostra de 21 artigos. As buscas nas referências desses artigos

resultaram na seleção de mais seis artigos. Ao final, o corpus desta revisão foi composto por 27 artigos, cujas informações extraídas serão descritas a seguir. Três artigos tratavam de medidas já contempladas em outros estudos do banco de dados. Portanto, foram identificados 24 instrumentos. Serão descritas características da amostra de artigos e, em seguida, informações sobre os instrumentos.

Características da amostra de artigos analisadas

Em relação às características dos artigos analisados, três aspectos foram examinados: autoria, objetivos e referencial teórico. Quanto a autoria, 74.07% dos artigos foram escritos por 2 ou 3 autores, o que parece sinalizar que o trabalho de construir (ou adaptar) e examinar as evidências psicométricas de um instrumento é complexo e requer a colaboração de alguns pesquisadores. No entanto, dos 84 autores identificados nos trabalhos selecionados, 91.67% eram responsáveis por um único artigo, o que parece indicar que muitos têm um envolvimento pontual com a investigação sobre a avaliação docente a partir de medidas específicas.

Também foi possível observar que 51.85% dos artigos possuíam como principal objetivo “apresentar evidências psicométricas”. Esse era um resultado esperado uma vez que esta revisão buscou identificar instrumentos com evidências de validade e fidedignidade voltadas para a avaliação do desempenho do professor na Educação Superior. Porém também foram localizados artigos que tinham como objetivo correlacionar o desempenho do professor com outras variáveis (18.52%) e identificar variáveis importantes para o desempenho do professor (29.63%). Por fim, em termos da área de conhecimento sobre a qual se fundamentava o artigo, verificou-se que 55.56% estavam relacionados à Educação, 40.74% à Psicologia e um único artigo adotou como referencial a Fonoaudiologia.

Instrumentos identificados

Cada artigo apresentou um único instrumento que avaliava o desempenho do professor. Contudo, três artigos (Alencar & Fleith, 2004a, b; Souza, Witter, & Araújo, 2014) se referiram ao Inventário de Práticas Docentes de Criatividade na Educação Superior

(Alencar & Fleith, 2004b) e dois artigos (Pereira, Loiola, & Gondim, 2016; Dourado, Gondim, Loiola, Ferreira, & Alberton, 2018) se referiam ao Inventário Aprendizagem/Competências, Suportes Organizacionais e Desempenho Docente (ASOD – Pereira et al., 2016). Assim, a partir dos 27 artigos analisados, foram identificados 24 instrumentos distintos. A Tabela 1 apresenta os instrumentos identificados, as evidências de validade e/ou fidedignidade mencionadas e os diferentes construtos relacionados ao desempenho do professor que foram avaliados.

É possível observar na Tabela 1 que os construtos mais avaliados foram desempenho do professor (33.33% dos instrumentos) e competências do professor (12.50%). No entanto, o construto avaliado pelo instrumento era indicado pelo(s) autor(es) do artigo e, algumas vezes, sem apresentar uma definição clara e precisa ao que estava se referindo. Assim, o fato de dois ou mais instrumentos avaliarem um mesmo construto não significa que eles estão medindo exatamente os mesmos aspectos. Por exemplo, o construto trabalho do professor foi avaliado nos instrumentos Avaliação do trabalho do professor pelo discente e Instrumento de Avaliação Docente, ambos elaborados por Comissões Próprias de Avaliação, respectivamente, da Faculdade Metodista de Santa Maria e de uma IES de Belo Horizonte. No primeiro, o construto é apresentado como sinônimo de desempenho do professor, mas sem maiores informações (Marchesan et al., 2011), e no segundo relacionado a expressão “gostar ou não gostar da aula do docente” (Gomes & Borges, 2008).

Segundo os *Standards* (American Educational Research Association et al., 2014), definir a estrutura conceitual que subsidiará a elaboração do teste é uma parte importante do processo de elaboração e validação do instrumento. É a estrutura conceitual que irá (a) delimitar os atributos do construto que devem ser representados e que serão avaliados, (b) indicar como o construto representado deve se distinguir de outros e como deve se relacionar com outras variáveis, (c) orientar os possíveis usos do teste e a maneira de interpretar suas pontuações, e (d) indicar os tipos de evidências adequadas para examinar se a interpretação sugerida para a pontuação do teste é coerente com a finalidade para a qual foi construído. Nessa direção, alguns dos artigos seriam mais claros se definissem de maneira precisa o construto que

Tabela 1.

Instrumentos e suas evidências de fidedignidade e validade

Instrumento (Autor)	Construto indicado pelo(s) autor(es)	Evidências de Fidedignidade			Evidências de Validade baseadas		
		Consistência interna	Precisão de avaliadores	Formas paralelas	Conteúdo do teste	Estrutura interna	Outras variáveis
Avaliação do trabalho docente pelo discente (CPA da FAMES, Marchesan et al., 2011)	Trabalho docente	Alpha	Não	Não	Não	Clusters e ACP	Não
Avaliação Docente (CPA de IES do Distrito Federal, Apio & Silvino, 2013)	Desempenho docente	Não	Não	Não	Não	AF	Sim
Avaliação Docente (CPA da UNIVALI, Kraemer & Verdinelli, 2017)	Desempenho docente	Não	Não	Não	Não	Não	Sim
Avaliação Docente da IES (CPA de IES não identificada, Primi & Moraes, 2005)	Desempenho docente	Alpha	Não	Não	Não	AFE e Clusters	Não
Escala de Avaliação do Professor Entusiasmado - EAPE (Rodrigues et al., 2009)	Entusiasmo	Alfa e Lambda	Não	Não	Não	ACP	Não
Escala de Avaliação Docente - EADoc (Silva et al., 2017)	Bom Professor	Alfa e Ômega hierárquico	Sim	Não	Juízes e Pop-alvo	AFE e TRI	Não
Escala de Esforço Instrucional (Rowe et al., 2013)	Investimento e empenho pessoal na atividade de ensino	Não	Não	Não	Não	ACP	Não
Escala de Práticas Docentes para a Criatividade na Educação Superior (Alencar & Fleith, 2010)	Criatividade	Alpha	Sim	Não	Pop-alvo	AFE	Não
Instrumento de Avaliação Docente (CPA de IES de Belo Horizonte, Gomes & Borges, 2008)	Trabalho docente	Alpha	Não	Não	Não	AFE	Não

(Continua)

Tabela 1.
Instrumentos e suas evidências de fidedignidade e validade (Continuação)

Instrumento (Autor)	Construto indicado pelo(s) autor(es)	Evidências de Fidedignidade			Evidências de Validade baseadas		
		Consistência interna	Precisão de avaliadores	Formas paralelas	Conteúdo do teste	Estrutura interna	Outras variáveis
Instrumento de Avaliação Docente da UNISUL (De Bem et al., 2011)	Desempenho docente	Sim*	Não	Não	Não	AFC e AFE	Não
Inventário Aprendizagem / Competências, Suportes Organizacionais e Desempenho Docente - ASOD (Pereira et al., 2016)	Aprendizagem de competências, suportes à transferência de aprendizagem e desempenho no cargo	Alpha	Sim	Não	Juízes e Pop-alvo	AFE	Não
Inventário de Habilidades Sociais Educativas do Professor Universitário – versão aluno (IHSE-PU-Aluno – Vieira-Santos et al., 2018)	Habilidades sociais educativas	Não	Sim	Não	Juízes e Pop-alvo	Não	Não
Inventário de Práticas Docentes de Criatividade na Educação Superior (Souza et al., 2014)	Criatividade	Alpha	Sim	Não	Pop-alvo	AFE	Não
Nome não informado (Avrichir & Dewes, 2006)	Desempenho docente	Alpha	Sim	Sim	Juízes e Pop-alvo	AFE	Não
Nome não informado (Borges et al., 2014)	Competências docentes	Não	Não	Não	Juízes e Pop-alvo	Não	Não
Nome não informado (CPA de IES do Distrito Federal, Palazzo & Gomes, 2009)	Desempenho docente	Não	Não	Não	Pop-alvo	AF	Sim
Nome não informado (Grohmann & Ramos, 2012)	Competências docentes	Alpha	Não	Não	Não	AFE	Não
Nome não informado (Kühl et al., 2013)	Competências docentes	Alpha	Sim	Não	Juízes e Pop-alvo	AFE	Não
Nome não informado (Ordóñez & Osorio, 2019)	Estilo de pensamento	Alpha	Não	Não	Não	Não	Não

(Continua)

Tabela 1.

Instrumentos e suas evidências de fidedignidade e validade (Continuação)

Instrumento (Autor)	Construto indicado pelo(s) autor(es)	Evidências de Fidedignidade			Evidências de Validade baseadas		
		Consistência interna	Precisão de avaliadores	Formas paralelas	Conteúdo do teste	Estrutura interna	Outras variáveis
Nome não informado (Peres-Dos-Santos & Laros, 2007)	Prática de ensino e relacionamento professor-aluno	Alpha	Não	Não	Juízes e Pop-alvo	AFE	Não
Questionário de avaliação de professor e da disciplina (Pinent et al., 1993)	Desempenho docente	Alpha	Não	Não	Não	Não	Sim
Questionário de Avaliação do Aluno - QAA2 (Romano et al., 2011)	Expressividade	Não	Sim	Não	Juízes	Não	Não
Questionário de avaliação do desempenho do Professor - Modelo A (Moreira, 1981)	Desempenho docente	Não	Não	Não	Pop-alvo	Não	Não
Questionário de Avaliação do Desempenho do Professor - Modelo C - Disciplinas de Física Geral (CPA do Instituto de Física da UFRGS, Silveira & Moreira, 1999)	Desempenho docente	Alpha	Não	Não	Não	AF	Sim

Nota: IES – Instituição de Ensino Superior; CPA – Comissão Própria de Avaliação; AF – Análise Fatorial; AFE – Análise Fatorial Exploratória; AFC – Análise Fatorial Confirmatória; ACP – Análise de Componentes Principais; TRI – Teoria de Resposta ao Item; Clusters – Análise de Clusters; Pop-alvo – População-alvo; Alpha – Alpha de Cronbach; * Não descreve qual coeficiente foi utilizado.

o instrumento se propõe a avaliar. Essa falta de definição pode ter ocorrido, em parte, porque o desempenho do professor é um fenômeno multideterminado e que pode assumir nuances distintas a partir do enfoque teórico sobre o qual é abordado. Portanto, cabe a sugestão para que novas pesquisas voltadas para a elaboração ou aplicação de instrumentos que avaliem o desempenho do professor apresentem de maneira clara e precisa o construto que está sendo examinado.

Para avaliar as evidências de fidedignidade e validade dos instrumentos, tomou-se como base as normas dos *Standards* (American Educational Research Association et al., 2014). Como é possível observar na Tabela 1, a evidência de fidedignidade mais utilizada foi a consistência interna, a qual foi relatada em 66.70% dos instrumentos, e o coeficiente mais empregado foi o Alpha de Cronbach. Apesar do alpha ser o coeficiente mais empregado nas medidas da área da Psicologia e

Educação, seu uso é problemático, pois, na maioria das vezes, os pressupostos para a sua aplicação (por exemplo, correlação linear entre os itens e covariância entre os itens igual para todos os pares de variáveis) são violadas (Damásio, 2012; Dunn, Baguley, & Brunsten, 2014; Zumbo, Gadermann, & Zeisser, 2007). Isso faz com que o α seja inflado e, conseqüentemente, diminui a estimativa de consistência interna do instrumento. Uma alternativa é usar coeficientes de consistência interna que não estejam baseados em tais pressupostos, tais como o ω (Dunn et al., 2014), o α ordinal e o θ ordinal (Zumbo et al., 2007), bem como métodos de análise fatorial confirmatória que permitam verificar a adequação de ajuste do modelo fatorial (Damásio, 2012).

Quanto aos demais procedimentos para analisar evidências de fidedignidade, verificou-se que 33.33% dos instrumentos foram examinados por avaliadores e um instrumento passou por procedimentos para estimar sua estabilidade temporal. No entanto, nenhum dos 24 instrumentos apresentou evidências de fidedignidade baseadas na equivalência de formas paralelas. Por isso, esse procedimento não foi especificado na Tabela 1.

Já em relação às evidências de validade, 70.83% dos instrumentos apresentavam evidências baseadas na estrutura interna, sendo que a análise fatorial exploratória (AFE) foi o principal procedimento utilizado. A AFE busca verificar a quantidade e composição dos fatores necessários para explicar a variância comum, ou seja, a parcela de variância que os itens compartilham entre si, permitindo, assim, identificar os traços ou construtos latentes mensurados pelo instrumento (Damásio, 2012). Também foram observados instrumentos cuja análise da estrutura interna foi realizada por análise de componentes principais, TRI, análise fatorial confirmatória ou análise de clusters.

Observou-se que 54.17% dos instrumentos apresentavam evidências de validade baseadas em análise de conteúdo, sendo possível verificar que em alguns instrumentos tal análise foi realizada apenas por juízes, em outros apenas pela população-alvo (análise semântica) e em outros por ambos. Para Primi, Muniz e Nunes (2009), tais evidências de validade consistem em investigações sobre o grau em que os itens do instrumento representam adequadamente características

do processo psicológico sobre o qual se pretende obter algum tipo de informação. Esse tipo de evidência de validade tende a ser privilegiado em estudos psicométricos, pelo menos, por dois motivos: (a) é fundamental para que se possa afirmar que o instrumento mensura o construto que se propõe avaliar, com base no julgamento de especialistas e dos usuários da medida; e (b) trata-se de uma evidência mais simples de ser investigada (a quantidade de participantes, por exemplo, é menor do que nos estudos cujos dados são submetidos a análises fatoriais).

Verificou-se, ainda, que 33.33% dos instrumentos apresentavam evidências de validade baseadas em outras variáveis, sendo que em quatro deles tais evidências referiam-se à validade de critério. Nessa direção, foram tomados como indicadores relevantes para avaliar o desempenho do professor: a nota do aluno/respondente, características dos professores valorizadas pelo plano de carreira da IES (tais como, titulação, tempo de experiência do professor e publicações), grau de importância que o aluno atribui a disciplina, autoavaliação que o aluno faz do próprio desempenho, percepção do aluno sobre o quanto aprende na disciplina, a atuação do professor em diferentes turmas e semestre que o aluno cursava. Também foi identificado um instrumento que apresentava evidência de validade convergente: o estudo de Apio e Silvino (2013) investigou se a aula pública realizada no processo de seleção docente é um bom preditor de desempenho do professor no ensino superior. Para tanto, comparou os resultados do instrumento Avaliação Docente da IES (elaborado pela CPA de IES do Distrito Federal) com o resultado do desempenho na aula pública (instrumento cujas evidências de validade não foram descritas no texto).

Nenhum instrumento apresentou evidências de validade baseadas no processo de resposta e/ou em estudos experimentais ou quase experimentais e, por isso, essas informações não foram acrescentadas na Tabela 1. Isso ocorreu, provavelmente, porque estudos que avaliam tais evidências são mais complexos e requerem um tempo maior de dedicação por parte do investigador. Conforme foi apresentado, a maior parte dos autores publicaram apenas um estudo relativo ao instrumento que adaptaram, construíram ou aplicaram. Pesquisas de natureza experimental ou quase experimental requerem mais recursos, por exemplo, em

termos de tempo e um instrumento em estágio avançado de investigação de suas evidências psicométricas. Contudo, parece que os estudos sobre instrumentos brasileiros para avaliação do desempenho do professor têm sido descontinuados. Os instrumentos são desenvolvidos ou adaptados e poucas pesquisas ou nenhuma se seguem após a primeira divulgação da medida.

Foi possível verificar que todos os instrumentos apresentaram alguma evidência de validade nos artigos analisados, mas nem sempre se referiam às evidências de fidedignidade – como, por exemplo, a Escala de Esforço Instrucional (Lopes et al., 2001 citado por Rowe, Bastos, & Pinho, 2013). Isso pode ter ocorrido porque nem todos os autores tenham conhecimento ou considerem importante seguir todo o protocolo de análises estatísticas esperado pela comunidade de psicólogos que estuda e trabalha com Psicometria (Damásio, 2012). Isso sugere existir uma necessidade de divulgar mais os conceitos e a importância das evidências psicométricas para toda a comunidade científica, englobando não só psicólogos como também educadores e profissionais de outras áreas interessados na construção de instrumentos válidos e precisos.

Além das evidências de fidedignidade e validade, também foram examinadas as seguintes características dos instrumentos: (a) estrutura fatorial, (b) se os itens estavam disponíveis no artigo, (c) quantidade de itens no instrumento, (d) tipo de escala de resposta, (e) aspecto avaliado, (f) forma, (g) meio de aplicação e (h) quem responde ao instrumento. Esses dados são apresentados na Tabela 2.

Quanto a estrutura fatorial dos instrumentos analisados, foram identificadas quatro: (a) simples, ou seja, estrutura fatorial unidimensional ou multidimensional que avaliava uma ou mais dimensões distintas do mesmo construto; (b) modelo de ordem superior (ou de segunda ordem), no qual uma solução fatorial hierárquica indicava que os fatores de primeira ordem (que eram mais específicos) podiam ser agrupados em fatores mais gerais (Laros, 2005); (c) conjuntos de escalas, situação em que um instrumento avaliava mais de um construto e, cada construto, era avaliado por um escala distinta; (d) não informado, ou seja, o artigo não descrevia a estrutura fatorial da medida. Em 70.83% dos casos, as medidas apresentaram estrutura fatorial simples.

Em relação a quantidade de fatores nos instrumentos com estrutura simples, o mínimo foi de um fator e o máximo foi de seis. Os instrumentos que apresentavam um único fator referiam-se à atuação do professor em sala de aula, a saber: (a) Avaliação Docente da IES, elaborado pela CPA de uma IES não identificada e descrito por Primi e Moraes (2005), (b) Escala de Avaliação Docente - EADoc (Silva et al., 2017), e (c) Instrumento de Avaliação Docente, elaborado pela CPA de uma IES de Belo Horizonte e descrito por Gomes e Borges (2008). Houve um caso de instrumento com um fator que se referiu à expressividade vocal do professor: Questionário de Avaliação do Aluno - QAA2 (Romano, Alves, Secco, Ricz, & Robazzi, 2011). Já os instrumentos que apresentavam dois fatores, avaliavam o relacionamento professor-aluno e a prática de ensino. Os instrumentos dessa categoria foram a Avaliação Docente da IES, elaborado pela CPA de uma IES do Distrito Federal e descrito por Apio e Silvino (2009), e os instrumentos descritos por Avrichir e Dewes (2006), Peres-dos-Santos e Laros (2007), Palazzo e Gomes (2009). A exceção foi o Questionário de Avaliação do Desempenho do Professor, elaborado pela CPA do Instituto de Física da Universidade Federal do Rio Grande do Sul e descrito por Silveira e Moreira (1999), cujos fatores foram teoria e laboratório. Os demais instrumentos avaliavam aspectos diversos do desempenho do professor. Tal diversidade pode ser atribuída, entre outros fatores, ao referencial teórico sobre o qual se fundamentou a elaboração da medida, bem como a forma como os itens foram redigidos.

Dos instrumentos identificados, 79.17% deles disponibilizavam os itens no artigo, permitindo ao leitor ter acesso aos comportamentos observados que refletiam o fator (ou construto latente) que estava sendo avaliado. A quantidade de itens do instrumento variou entre 13 e 82, mas a maioria dos instrumentos apresentavam no máximo 45 itens. Instrumentos muito extensos, além de serem cansativos, implicam na necessidade de um tempo maior de aplicação, o que representa um problema quando a coleta deve ocorrer em sala, durante o período de aula e, portanto, podem inviabilizar o seu uso em larga escala no contexto de avaliação institucional do desempenho do professor.

A escala do tipo *Likert* foi o formato de resposta adotado em 70.84% dos instrumentos, com predomínio

Tabela 2.

Características dos instrumentos

Características	N	%
<i>Estrutura fatorial</i>		
Simple	17	70.83
Modelo de ordem superior	1	4.17
Conjunto de escalas	2	8.33
Não informado	4	16.67
<i>Disponibiliza os itens</i>		
Sim	19	79.17
Não	5	20.83
<i>Quantidade de itens</i>		
Até 15 itens	4	16.67
Entre 16 e 30 itens	8	33.33
Entre 31 e 45 itens	6	25.00
Mais que 45 itens	4	16.67
Não informado	2	8.33
<i>Tipo de escala de resposta</i>		
Escala Likert de 5 pontos	15	62.50
Escala tipo Likert de 7 pontos	1	4.17
Escala tipo Likert de 10 pontos	1	4.17
Múltipla escolha	2	8.33
Não informado	5	20.83
<i>Aspecto avaliado*</i>		
Aquisição	1	4.17
As respostas possíveis mudam em função do item avaliado	2	8.33
Concordância	9	37.50
Frequência de ocorrência do comportamento	2	8.33
Importância	2	8.33
Domínio	2	8.33
Qualidade	1	4.17
Não informado	7	29.17
<i>Forma de aplicação</i>		
Coletiva	6	25.00
Individual	11	45.83
Não informado	7	29.17
<i>Meio de aplicação</i>		
Impresso	8	33,33
Online	8	33,33
Não informado	8	33,34
<i>Respondente</i>		
Aluno	17	70.83
Aluno e pesquisador	1	4.17
Professor / Tutor	6	25.00

Nota: * A soma das porcentagens desta categoria é maior do que 100% porque alguns instrumentos solicitavam mais de um tipo de resposta.

da escala com cinco pontos. Em 37.50% dos instrumentos essa escala avaliava o nível de concordância do respondente com as afirmações contidas nos itens.

A aplicação foi realizada de forma individual em 45.83% dos estudos. Já em relação ao meio de aplicação, um terço dos instrumentos foram aplicados de maneira impressa, um terço através da internet (online) e um terço não informou o meio de aplicação.

Os instrumentos foram elaborados para serem respondidos, em sua maioria (70.83%), por alunos, o que parece indicar ser essa uma fonte recorrente de informação sobre o desempenho do professor. O uso da avaliação do professor pelo discente é polêmica. Costin, Greenough e Menges (1971), por exemplo, fizeram uma revisão de estudos sobre avaliação do professor pelo discente, considerando a produção entre as décadas de 1920 e 1970, e reportaram diversas evidências de confiabilidade e validade de diferentes instrumentos. Destacam-se os seguintes achados: (a) avaliações de estudantes são estáveis ao longo do tempo (mais estáveis que a avaliação do professor por pares, por exemplo); (b) existe correlação entre a avaliação do professor e o quanto os alunos consideram ter alcançado (ou efetivamente alcançaram) os objetivos de ensino; (c) estudantes parecem ser capazes de distinguir certas habilidades do professor que favorecem a aprendizagem e motivação; (d) correlação inexistente ou fraca entre notas e a avaliação do professor pelo discente; (e) o tamanho da turma parece afetar apenas certos tipos de avaliação (p. ex., a correlação é positiva em relação à preparação do professor e negativa em relação à promoção de pensamento original/criativo); etc. Revisões mais recentes, como a de Wilson e Ryan (2012), confirmam os achados de Costin et al. (1971).

Considerações finais

O objetivo deste estudo foi identificar quantos e quais as características de instrumentos com evidências psicométricas e que avaliam o desempenho de professores universitários brasileiros no ensino. Foram encontrados 24 instrumentos, construídos por pesquisadores ou por CPA's de universidades brasileiras. Sobre cada instrumento selecionado, geralmente, existe

apenas um estudo no Brasil, o que sugere a necessidade de que mais pesquisas sejam conduzidas e/ou divulgadas em periódicos científicos sobre esse tema e que, preferencialmente, utilizem ou dialoguem com as medidas de desempenho já desenvolvidas.

Com relação às evidências de fidedignidade, observou-se a predominância do Alfa, apesar das suas limitações, e pouca referência de outros tipos de evidências de fidedignidade como o teste-reteste. Sobre o uso do alfa, segundo Dunn et al. (2014), a sua prevalência e o fato de que as alternativas a ele são pouco empregadas, estão relacionadas a formação dos pesquisadores e a literatura disponível sobre os métodos alternativos. Para esses autores, em sua maioria, os cursos de pós-graduação não oferecem formação consistente sobre métodos de análise de confiabilidade. Aliado a isso, "(...) os artigos que condenam o alfa tendem a ser muito técnicos e, portanto, não apenas inacessíveis a muitos psicólogos, mas também não oferecem alternativas 'realistas'" (Dunn et al., 2014, p. 400, tradução livre). No caso do teste-reteste, uma variável importante pode ser o custo envolvido de uma nova coleta com a mesma população.

Em termos de evidência de validade, predominaram as de conteúdo (via consulta a especialistas ou a população-alvo) e relativas à estrutura interna (a AFE é a técnica de análise mais comum). Segundo a literatura apresentada, essa ênfase encontrada faz sentido, pois são estudos e análises iniciais e fundamentais para a construção ou adaptação de medidas. Contudo, importa salientar que também é preciso contemplar estudos que avaliem evidências de validade de critério, convergente, discriminante, etc. Além disso, notou-se a importância de uma descrição mais clara e precisa dos construtos investigados por cada instrumento, pois, muitas vezes, ela não foi apresentada ou estava descrita de forma vaga. Essa dúvida impõe dificuldades para o planejamento de estudos de validade convergente com outros instrumentos brasileiros.

Estudos como este de revisão contribuem com o avanço da ciência porque organizam e divulgam o conhecimento disponível sobre determinado assunto em um momento específico. Permitem identificar lacunas nesse conhecimento e, assim, podem impulsionar pesquisas empíricas (Luna, 1996). Nesse sentido, parece fundamental ampliar os estudos de investigação

de evidências psicométricas dos instrumentos existentes e avaliar as relações entre eles, de modo a descrever de forma mais precisa os aspectos do desempenho do professor que têm sido estudados a partir desses instrumentos e que têm sido considerados essenciais para o aperfeiçoamento dos professores e a promoção de um ensino superior de maior qualidade.

Ressalta-se que os critérios utilizados nesta revisão, como o uso de descritores em português, não permitiram a identificação de todos os tipos de instrumento que contribuem ou dizem respeito, direta ou indiretamente, ao trabalho do professor. Por exemplo, o Questionário de Avaliação da Docência, construído por Pasquali (1984), não apareceu em nenhum resultado das buscas. Outros instrumentos, por sua vez, que continham itens para avaliação do desempenho do professor, mas possuíam também itens para tratar de aspectos como a estrutura física da universidade, não foram selecionados para a amostra, embora tenham relação com o que foi discutido neste trabalho. Além disso, foram encontrados artigos que reportavam o uso de instrumentos de avaliação do desempenho do professor em universidades, mas sem evidências psicométricas. Devem existir também muitos casos de universidades que possuem seus instrumentos de avaliação do desempenho do professor, mas nunca publicaram nada a esse respeito (e pode ser que essas medidas disponham ou não de evidências psicométricas). Finalmente, existem instrumentos que avaliam a satisfação dos alunos com o ensino, com a instituição de ensino, com a experiência acadêmica, entre outros construtos, e que contém itens relativos ao desempenho do professor. Todo esse arsenal de medidas é importante e poderia compor o objetivo de estudos de revisão futuros.

Com efeito, trata-se de um campo vasto de investigação, que requer organização e investimento em termos de pesquisa, uma vez que tem repercussões para a formação inicial e continuada de professores, bem como para a progressão dos mesmos na carreira. Em última análise, esses instrumentos podem contribuir com a criação de oportunidades para o aperfeiçoamento dos professores e do ensino de nível superior no país. Essa é uma necessidade premente dos brasileiros que esperam de seus profissionais de nível superior a capacidade de lidar com um ambiente complexo nas mais

diversas dimensões, tais como política, meio ambiente, consumo, tecnologias e questões sociais diversas.

Referências

- *Alencar, E. M. L. S., & Fleith, D. D. S. (2004a). Creativity in University Courses: Perceptions of Professors and Students. *Gifted and Talented International*, 19(1), 24–28. doi: 10.1080/15332276.2004.11673029
- *Alencar, E. M. L. S., & Fleith, D. S. (2004b). Inventário de práticas docentes que favorecem a criatividade no ensino superior. *Psicologia: Reflexão e Crítica*, 17(1), 105–110. doi: 10.1590/S0102-79722004000100013
- *Alencar, E. M. L. S., & Fleith, D. S. (2010). Escala de práticas docentes para a criatividade na Educação Superior. *Avaliação Psicológica*, 9(1), 13–24. Recuperado de <https://bit.ly/2IqAjmY>
- Alves, W. F. (2007). A formação de professores e as teorias do saber docente: contextos, dúvidas e desafios. *Educação e Pesquisa*, 33(2), 263–280. doi: 10.1590/S1517-97022007000200006
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.
- *Apio, A. L., & Silvino, A. M. D. (2013). A aula pública como instrumento preditor de desempenho docente no ensino superior. *Avaliação: Revista da Avaliação da Educação Superior (Campinas)*, 18(2), 277–298. doi: 10.1590/S1414-40772013000200003
- *Avruchir, I., & Dewes, F. (2006). Construção e validação de um instrumento do desempenho docente. *Revista Eletrônica de Ciência Administrativa*, 5(2), 1–16. doi: 10.5329/RECADM.20060502005
- *Borges, J. P. F., Coelho Junior, F. A., Faiad, C., & Rocha, N. F. (2014). Diagnóstico de competências individuais de tutores que atuam na modalidade a distância. *Educação e Pesquisa*, 40(4), 935–951. doi: 10.1590/s1517-97022014121642
- Cândido, C. M., Assis, M. R., Ferreira, N. T., & Souza, M. A. (2014). A representação social do

- “bom professor” no ensino superior. *Psicologia & Sociedade*, 26(2), 356-365. doi: 10.1590/S0102-71822014000200012
- Cassettari, N. (2014). Avaliação de professores: Uma questão de escolhas. *Estudos em Avaliação Educacional*, 25(57), 166-197, 2014. doi: 10.18222/eaecae255720142829
- Costin, F., Greenough, W. T., & Menges, R. J. (1971). Student ratings of college teaching: Reliability, validity, and usefulness. *Review of Educational Research*, 41(5), 511-535. doi: 10.3102/00346543041005511
- Damásio, B. F. (2012). Uso da análise fatorial exploratória em psicologia. *Avaliação Psicológica*, 11(2), 213–228. Recuperado de <https://bit.ly/2EFTFc3>
- *De Bem, A. B., Lanzer, E. A., Tambosi Filho, E., Sanchez, O. P., & Bernardi Junior, P. (2011). Validade e confiabilidade de instrumento de avaliação da docência sob a ótica dos modelos de equação estrutural. *Avaliação: Revista da Avaliação da Educação Superior (Campinas)*, 16(2), 375–401. doi: 10.1590/S1414-40772011000200008
- *Dourado, P. C., Gondim, S. M. G., Loiola, E., Ferreira, A. S. M., & Alberton, G. D. (2018). Aprendizagem individual, suporte organizacional e desempenho percebido: um estudo com docentes universitários. *Educação em Revista*, 34, e178191. doi: 10.1590/0102-4698178191
- Dunn, T. J., Baguley, T., & Brunnsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, 105(3), 399–412. doi: 10.1111/bjop.12046
- Galvão, T. F., Pansani, T. S. A., & Harrad, D. (2015). Principais itens para relatar Revisões sistemáticas e Meta-análises: A recomendação PRISMA. *Epidemiologia e Serviços de Saúde*, 24(2), 335-342. doi: 10.5123/S1679-49742015000200017
- *Gomes, C. M. A., & Borges, O. (2008). Limite da validade de um instrumento de avaliação docente. *Avaliação Psicológica*, 7(3), 391–401. Recuperado de <https://bit.ly/2Jern5r>
- *Grohmann, M. Z., & Ramos, M. S. (2012). Competências docentes como antecedentes da avaliação de desempenho do Professor: percepção de mestrandos de administração. *Avaliação: Revista da Avaliação da Educação Superior (Campinas)*, 17(1), 65–86. doi: 10.1590/S1414-40772012000100004
- Gusso, H. L. (2013). *Avaliação da eficiência de um procedimento de apresentação semanal de consequências informativas ao desempenho de alunos em nível superior* (Tese de doutorado). Centro de Filosofia e Ciências Humanas, Universidade Federal de Santa Catarina, Florianópolis, SC. Recuperado de <https://goo.gl/AAqvQM>
- Komaraju, M. (2013). Ideal teacher behaviors: Student motivation and self-efficacy predict preferences. *Teaching of Psychology*, 40(2), 104-110. doi: 10.1177/0098628312475029
- *Kraemer, M. E. P., & Verdinelli, M. A. (2017). Desempenho no Enade e sua relação com a avaliação docente pelo discente. *Revista Gestão Universitária Na América Latina - GUAL*, 10(3), 204–227. doi: 10.5007/1983-4535.2017v10n3p204
- *Kühl, M. R., Maçaneiro, M. B., Cunha, J. C., & Cunha, S. K. (2013). O valor das competências docentes no ensino da Administração. *Revista de Administração*, 48(4), 783–799. doi: 10.5700/rausp1121
- Laros, J. A. (2005). O uso da análise fatorial: algumas diretrizes para pesquisadores (pp. 163–184). In: Pasquali, L. (Org.), *Análise fatorial para pesquisadores*. Brasília, DF: LabPAM.
- Lei nº 10.861 de 14 de abril de 2004 (2004, 14 abril). Institui o Sistema Nacional de Avaliação da Educação Superior – SINAES e dá outras providências. Diário Oficial da União, Brasília.
- Luna, S.V. (1996). *Planejamento de pesquisa: uma introdução*. São Paulo: EDUC.
- *Marchesan, T. M., Souza, A. M., & Menezes, R. (2011). Avaliação do processo de ensino: uma abordagem multivariada. *Production*, 21(2), 271–283. doi: 10.1590/S0103-65132011005000018
- *Moreira, M. A. (1981). Avaliação do professor pelo aluno como instrumento de melhoria do ensino universitário. *Educação e Seleção*, 4, 109–124. Recuperado de <https://bit.ly/2uIwAz3>
- Nunes, C. H. S. S., & Primi, R. (2010). Aspectos técnicos e conceituais da ficha de avaliação dos testes psicológicos (pp. 101-127). In: Conselho Federal

de Psicologia (Org.), *Avaliação psicológica: diretrizes na regulamentação da profissão*. Brasília: CFP.

- *Ordóñez, L. J. L., & Osorio, A. C. N. (2019). The thinking styles of the university teacher. *Psicologia da Educação*, (49), 13-30. doi: 10.5935/2175-3520.20190015
- *Palazzo, J., & Gomes, C. A. (2009). Plano de carreira e avaliação dos professores: encontros e desencontros. *Ensaio: Avaliação e Políticas Públicas em Educação*, 17(63), 205–234. doi: 10.1590/S0104-40362009000200003
- Pasquali, L. (1984). Questionário de Avaliação da Docência. *Educação e Seleção*, 9, 71-98. Recuperado de <https://goo.gl/kk3rjf>
- *Pereira, L. M. R., Loiola, E., & Gondim, S. M. G. (2016). Aprendizagem de competências, suporte à transferência de aprendizagem e desempenho docente: evidências de validação de escala e teste de relações. *Organizações & Sociedade*, 23(78), 438–459. doi: 10.1590/1984-92307856
- *Peres-Dos-Santos, L. F., & Laros, J. A. (2007). Avaliação da prática pedagógica do professor de ensino superior. *Estudos em Avaliação Educacional*, 18(36), 75–96. doi: 10.18222/ae183620072101
- *Pinent, C. E. C., Silveira, F. L., & Moraes, R. (1993). Avaliação do professor pelo aluno: questionário avaliativo e testes de validação. *Estudos em Avaliação Educacional*, 8, 65–79. doi: 10.18222/ae00819932342
- Portaria nº 554 (2013, 20 junho). Estabelece as diretrizes gerais para o processo de avaliação de desempenho para fins de progressão e de promoção dos servidores pertencentes ao Plano de Carreiras e Cargos de Magistério Federal das Instituições Federais de Ensino vinculadas ao Ministério da Educação. Diário Oficial da União, nº 118, Brasília, DF.
- *Primi, R., & Moraes, I. F. (2005). Validade de um instrumento de avaliação de desempenho de docentes do ensino superior. *Avaliação: Revista da Rede de Avaliação Institucional da Educação Superior*, 10(3), 41-59. Recuperado de <https://bit.ly/2IpG1FE>
- Primi, R., Muniz, M., & Nunes, C. H. S. S. (2009). Definições contemporâneas de validade de testes psicológicos (pp. 243-265). In: Hutz, C. S. (Org.), *Avanços e polêmicas em avaliação psicológica*. São Paulo: Casa do Psicólogo.
- Ribeiro, E. A. (2012). As atuais políticas públicas de avaliação para a educação superior e os impactos na configuração do trabalho docente. *Avaliação: Revista da Avaliação da Educação Superior*, 17(2), 299-316. doi: 10.1590/S1414-40772012000200002
- *Rodrigues Jr, J. F., Pasquali, L., & Moura, C. F. (2009). Dimensões do construto entusiasmo como percebido em professores universitários. *Avaliação Psicológica*, 8(3), 391–403. Recuperado de <https://bit.ly/2GThjXk>
- *Romano, C. C., Alves, L. A., Secco, I. A. O., Ricz, L. N. A., & Robazzi, M. L. C. C. (2011). A expressividade do docente universitário durante sua atuação na sala de aula: análise dos recursos verbais utilizados e suas implicações para a enfermagem. *Revista Latino-Americana de Enfermagem*, 19(5), 1188–1196. Recuperado de <https://bit.ly/2E8VYxa>
- *Rowe, D. E. O., Bastos, A. V., & Pinho, A. P. M. (2013). Múltiplos comprometimentos com o trabalho e suas influências no desempenho: um estudo entre professores do Ensino Superior no Brasil. *Organizações & Sociedade*, 20(66), 501–522. Recuperado de www.revistaoes.ufba.br
- *Silva, M. A., Machado, W. L., Pilotto, L. M., Backes, B., Zanon, R. B., Machado, P. V., ... Bandeira, D. R. (2017). Construção e estudo de evidências de validade da Escala de Avaliação Docente. *Revista Brasileira de Educação*, 22(70), 690–707, 2017. doi: 10.1590/s1413-24782017227035
- *Silveira, F. L., & Moreira, M. A. (1999). Estudo da validade de um questionário de avaliação do desempenho do professor de física geral pelo aluno. *Ensaio Pesquisa Em Educação Em Ciências (Belo Horizonte)*, 1(1), 54–65, 1999. doi: 10.1590/1983-21171999010106
- *Souza, J. R. S., Witter, G. P., & Araújo, R. S. (2014). Criatividade na prática de docentes na área da psicologia: perspectivas de alunos e de professores. *Boletim de Psicologia*, 64(141), 143–157. Recuperado de <https://bit.ly/2H7VktR>
- *Vieira-Santos, J., Del Prette, A., & Del Prette, Z. A. P. (2018). Inventário de habilidades sociais

- educativas do professor universitário - versão aluno (IHSE-PU-Aluno): dados preliminares. *Avaliação Psicológica*, 17(2), 260-270. doi: 10.15689/ap.2018.1702.14611.12
- Vosgerau, D., Orlando, E., & Meyer, P. (2017). Produtivismo acadêmico e suas repercussões no desenvolvimento profissional de professores universitários. *Educação & Sociedade*, 38(138), 231-247. doi: 10.1590/ES0101-73302016163514
- Wilson, J. H., & Ryan, R. G. (2012). Formative teaching evaluations: Is student input useful? (pp. 22-29). In: Kite, M. E. (Org.), *Effective evaluation of teaching: A guide for faculty and administrators*. Recuperado de <https://goo.gl/pSx91e>
- Zanon, C., & Hauck Filho, N. (2015). Fidedignidade (pp. 85-95). In: Hutz, C. S., Bandeira, D. R., & Trentini, C. M. (Orgs.). *Psicometria*. Porto Alegre: Artmed.
- Zumbo, B. D., Gadermann, A. M., & Zeisser, C. (2007). Ordinal versions of coefficients Alpha and Theta for likert rating scales. *Journal of Modern Applied Statistical Methods*, 6(1), 21-29. doi: 10.22237/jmasm/1177992180
- Nota: * = Artigos que compuseram o corpus de manuscritos analisados nesta pesquisa.
- Recebido em: 27/10/2019
Revisado em: 18/02/2020
Aprovado em: 06/04/2020

Sobre os autores:

Joene Vieira-Santos

Centro Universitário Adventista de São Paulo – UNASP, Hortolândia – SP
ORCID: <https://orcid.org/0000-0001-9529-8417>
E-mail: joenesantos@yahoo.com.br

Marcelo Henrique Oliveira Henklain

Universidade Federal de Roraima – UFRR, Boa Vista – RR
ORCID: <https://orcid.org/0000-0001-9884-8592>
E-mail: marcelo.henklain@ufr.br

Contato com os autores:

Rua Pastor Hugo Gegembauer, 265, Portaria, Parque Hortolândia
Hortolândia-SP, Brasil
CEP: 13184-010
Telefone: (19) 98172-8594.