



XIX COLÓQUIO INTERNACIONAL DE GESTÃO UNIVERSITÁRIA

Universidade e Desenvolvimento Sustentável: desempenho acadêmico e os desafios da sociedade contemporânea

Florianópolis | Santa Catarina | Brasil
25, 26 e 27 de novembro de 2019
ISBN: 978-85-68618-07-3



ESCALAS DE AUTOAVALIAÇÃO DE COMPETÊNCIAS DOCENTES EM ARTIGOS INTERNACIONAIS

FERNANDO RAMOS LENGLER

Faculdade CESUSC

fernandolengler@gmail.com

NÉRIO AMBONI

Universidade do Estado de Santa Catarina (UDESC)

amboni30@yahoo.com.br

RENAN GRIJÓ BÚRIGO

Universidade do Estado de Santa Catarina (UDESC)

rburigo@gmail.com

RESUMO

A competência docente é um construto que tem e vem despertando interesse por diferentes pesquisadores e estudiosos do campo da educação e da gestão. Assim como outros recursos intangíveis, as competências docentes exercem influência sobre o desempenho individual no exercício profissional do professor e sobre os resultados de uma instituição de ensino. O presente trabalho tem por objetivo analisar uma amostra de artigos que utilizaram escalas de autoavaliação de competências docentes em relação aos passos definidos por DeVellis (2017). Os artigos analisados foram selecionados a partir de buscas em bases livres, utilizando os descritores “*teaching competency*”, “*teaching competencies*”, “*scale development*” e “*identification of competencies*”, com o filtro dos que avaliassem por meio de escalas a percepção do professor acerca de suas próprias competências. Os achados evidenciaram que o uso de escalas de autoavaliação de competências docentes é de natureza plural, reforçada pelas diferentes dimensões utilizadas na geração de itens para compor o instrumento de coleta de dados, demonstrando que não há um modelo de autoavaliação das competências docente amplamente aceito.

Palavras-chave: Autoavaliação docente. Escala de mensuração. Competências docentes.

1. INTRODUÇÃO

As competências necessárias para o exercício da docência é um tema que tem despertado crescente interesse por parte de teóricos e pesquisadores, por exercerem influência sobre o desempenho individual do professor e, por conseguinte, sobre os resultados da instituição de ensino (MADURO, 2013; PERRENOUD, 2015). Diferentes interpretações dos conceitos e abordagens de competências são evidenciadas nos estudos e pesquisas, revelando aspectos comuns, principalmente nas características de iniciativa e responsabilidade frente às situações ocupacionais.

As competências, desta forma, são avaliadas de diferentes maneiras, sendo uma delas, através do uso de escalas. As escalas podem ser formadas para mensurar uma ou mais características de um fenômeno, a partir de um grupo de itens, que se referem a um conjunto de indicadores ou questões (URBINA, 2007). As escalas de medidas são muito utilizadas em pesquisas na área de Ciências Sociais Aplicadas. Escalas podem referir a técnica de escalonamento, como técnicas comparativas ou não comparativas. A escala pode classificar um nível de medida e mensuração de variáveis que a pesquisa possui e busca responder, utilizando parâmetros nominais, ordinais, intervalares e de razão (STEVENS, 1946; HAIR JR et al., 2005).

As escalas são construídas com o intuito de buscar características dos objetos e fenômenos, que podem ser chamados de construtos. Os construtos podem ser observáveis, que conseguem ser mensurados diretamente, ou latentes, que geram manifestações observáveis e podem ser classificados como formativos, onde os itens formam o construto, ou reflexivos, onde o construto reflete os itens. Os construtos também podem ser unidimensionais, que apresenta somente uma dimensão, ou multidimensionais, que apresentam mais de duas dimensões para alcançar as medidas desejadas (HAIR JR et al., 2005; COSTA, 2011; PASQUALLI, 2012; DEVELLIS, 2017).

Neste contexto, o artigo tem por objetivo analisar artigos internacionais que utilizaram escalas de autoavaliação de competências docentes em relação aos passos definidos por DeVellis (2017). A autoeficácia ou autoavaliação, segundo Bandura (1986), baseia-se na percepção de competência, pois está relacionada à ação e é dependente da situação, domínio e tarefas específicas ao contexto em que a atividade está sendo desenvolvida. Ela mostra a percepção que o professor tem a respeito de suas capacidades no exercício da atividade docente.

Para o alcance do objetivo geral foram selecionados 38 artigos internacionais, acerca de “escalas de medida das competências docentes”, a partir de buscas livres na internet com o uso dos descritores “*teaching competency*”, “*teaching competencies*”, “*scale development*” e “*identification of competencies*”, podendo estar no título e na busca como um todo, publicados nas diversas bases de dados no período de 2010 a 2017. Deste total, foram selecionados 09 artigos que avaliaram a percepção do professor acerca de suas próprias competências, ou seja, os estudos de autoavaliação ou autoeficácia de competências docentes. Os artigos que tinham por objetivo verificar as competências docentes na ótica dos alunos ou demais envolvidos não foram considerados no estudo.

O artigo está estruturado com a revisão teórica dos assuntos centrais da pesquisa: desenvolvimento de escalas de medidas e competências docentes, a descrição dos procedimentos realizados nos artigos selecionados, análise sobre as escalas de medidas utilizadas nos artigos e considerações finais.

2. REVISÃO TEÓRICA

Este tópico apresenta os fundamentos, a partir do que é apresentado por DeVellis (2017), para o desenvolvimento de escalas de medidas e competências docentes.

2.1. ESCALAS DE MEDIDAS A PARTIR DE DEVELLIS

Robert F. DeVellis (2017), em sua obra “*Scale development: theory and application*”, além de conceituar o desenvolvimento de escalas, traz de maneira acessível partindo de uma visão geral com perspectivas de medição, entendimento da variável latente, confiabilidade de mensuração, validade de escalas, diretrizes de desenvolvimento de escalas, aplicação e relevância da análise fatorial, visão geral da teoria da resposta ao item e o contexto da medição com medidas. Com a base consolidada de conhecimento, DeVellis (2017) disponibiliza uma série de diretrizes para o desenvolvimento de escalas que serve como orientações e passo-a-passo àqueles que precisam construir ou validar suas investigações por meio de escalas de medidas.

As etapas, neste estudo apresentadas apenas como passos de DeVellis (2017), formam uma metodologia de oito momentos, destacados com detalhamentos que expressam a preocupação do desenvolvimento de uma escala, no que tange o que realmente retrata a necessidade de investigação.

No primeiro (1º) passo, **determinar claramente o que se deseja medir**, começa com um embasamento teórico sobre o assunto e construto que está sendo mensurado, ou seja, é necessário defender e sustentar a importância do desenvolvimento da escala com uma descrição do construto e existência do fenômeno como um todo. Como maneira de determinar a clareza da escala é necessário ressaltar a especificidade do fenômeno e a posição do construto para a construção dos itens alinhados à variável latente. Neste processo a subjetividade e experiência do pesquisador faz parte da construção e definição dos itens para a medição. Por fim, no primeiro passo, para ter clareza sobre o que incluir numa medida, os desenvolvedores de escalas devem questionar se o construto avaliado é distinto de outros construtos e se os itens aparentemente similares podem representar outros construtos, ou mesmo serem sensíveis a outros fenômenos (DEVELLIS, 2017).

No segundo (2º) passo, **gerar um pool de itens**, após identificar o que realmente será medido com os itens criados, o instrumento poderá ser construído. Neste momento é recomendável a geração de vários itens que reflitam o propósito da escala para que estes sejam candidatos a uma eventual inclusão no instrumento de medição. Os itens podem ser submetidos a análises de sua força com a variável latente, é importante entender a redundância de alguns itens, mensurar a quantidade suficiente de itens para a escala, diagnosticar e entender o processo de escrever os itens para as características da escala e seus respondentes, além de definir os itens com palavras negativas ou positivas, tudo isto devendo refletir o construto de interesse. O pesquisador pode eliminar alguns itens baseando-se em critérios a priori, como clareza, relevância questionável ou similaridades indesejáveis entre os itens (DEVELLIS, 2017).

No terceiro (3º) passo, **determinar o formato da medição**, a literatura apresenta diversos formatos possível de questões e caberá o autor, de acordo com a compatibilidade do tema, definir a melhor forma de medição. Algumas tipologias utilizam as características como: escala de Thurstone, que respondem diferentemente a específicos níveis de atributos em uma questão; escala de Guttman, que é uma escala de atitudes e cumulativa tipo hierárquica; escala Likert, que traduz a relação em graus de concordância e discordância com intervalos iguais; escala Diferencial Semântico, que contempla um par de adjetivos opostos que é possível responder entre a linha de opostos (DEVELLIS, 2017).

No quarto (4º) passo, **possuir um pool de itens iniciais revisados por especialistas**, é onde as revisões produzidas por experts maximizam e auxiliam na validade do conteúdo da escala. Por meio da verificação de especialistas no tema do construto a relevância dos itens, a clareza e a consistência dos itens são confirmadas para a composição da escala de medida (DEVELLIS, 2017).

No quinto (5º) passo, **considerar a inclusão de itens de validação**, devem ser consideradas as solicitações e indicações para a inclusão e mudança nos itens primariamente propostos. Este processo pode ocorrer após ser detectado de falhas ou problemas no modelo de medida e entendimento da variável latente a ser medida. Este momento ajudará na validade do construto da escala (DEVELLIS, 2017).

No sexto (6º) passo, **administrar itens em uma amostra de desenvolvimento**, é importante que o instrumento de medição seja aplicado em uma amostra estatisticamente compatível para que os resultados possam ser avaliados no passo seguinte sem maiores problemas. O tamanho da amostra deve ser de no mínimo cinco respondentes para cada item de medição e ou com uma amostra de pelo menos trezentos indivíduos. Estes parâmetros ajudam a evitar problemas relacionados a amostras pequenas, ou seja, o padrão de covariação entre os itens pode não ser estável, assim como a amostra pode não representar a população para a qual a escala é destinada (DEVELLIS, 2017).

No sétimo (7º) passo, **avaliar os itens**, um exame inicial da performance dos itens é necessário. Poderá ser realizado através da obtenção de correlação elevada da variável latente, já que os itens deverão estar correlacionados. O pesquisador deve atentar para a ocorrência de itens com a resposta inversa, reescrevendo o item, e atentar também para a correlação dos itens da escala, com o intuito de identificar a variação dos mesmos em relação à amostra. Após a verificação dos itens, é necessário verificar o significado dos itens, através da análise fatorial, que é considerada a melhor maneira para se determinar um grupo de itens compõe alguma dimensão de análise ou se constitui um conjunto unidimensional. Além disto, é importante que seja calculado o coeficiente Alpha de Cronbach para todas as dimensões encontradas, que avalia as correlações dentro de cada uma das dimensões (DEVELLIS, 2017).

No oitavo (8º) e último passo, **otimizar o tamanho da escala**, a proposta é avaliar após os passos anteriores os efeitos do comprimento da escala na confiabilidade do instrumento. Deve-se considerar a brevidade, tamanho e quantidade de itens em vias de verificar a confiabilidade da escala (DEVELLIS, 2017).

O Quadro 1 sucinta os critérios e questionamentos apontados por DeVellis (2017) para o desenvolvimento de escalas, que servirá de base para a avaliação e análise das escalas selecionadas para este estudo.

Quadro 1 – Detalhamento para o desenvolvimento de escalas de DeVellis

nº	Passos	Detalhamento
1º	Determine claramente o que se deseja medir	Teoria como auxílio à clareza
		Especificidade como um auxílio à clareza
		Ser claro sobre o que incluir em uma medida
2º	Gere um pool de itens	Escolha itens que reflitam o propósito da escala
		Entenda a redundância dos itens
		Número suficiente de itens
		Entenda o processo de escrever itens
		Liste as características de bons e maus itens
	Itens com palavras positivas e negativas	
3º	Determine o formato da medição	Tipos de escala (Thurstone, Guttman, Likert, entre outras)
		Escalas com item igualmente ponderado
		Quantas categorias de resposta
4º	Tenha um <i>pool</i> de itens iniciais revisados por especialistas	Validade, clareza, relevância, consistência, agrupamentos dos itens
		Pré-teste com especialistas
5º	Considere a inclusão de itens de validação	Detectar falhas, problemas e inclusões de itens
		Validade do construto
6º	Administre itens a uma amostra de desenvolvimento	Mínimo de 5 respondentes para cada item
		Caracterização da amostra
		Tamanho da amostra (~300 indivíduos)

		Erro amostral
7°	Avalie os itens	Exame inicial do desempenho dos itens
		Análise fatorial
		Coefficiente Alfa de Cronbach
8°	Otimize o tamanho da escala	Efeito do comprimento da escala na confiabilidade
		Efeito da queda de itens "ruins"
		Tamanho da escala
		Divisão da amostra

Fonte: Adaptado de DeVellis (2017)

As escalas surgiram da necessidade de mecanismos de mensuração que reflitam o construto como um todo, possibilitando a comparação entre estudos e aplicação em diversas realidades. Espera-se no desenvolvimento de uma escala que os pesquisadores possuam os cuidados essenciais da revisão da literatura para a identificação e embasamento dos itens iniciais a compor a escala, consulta com especialistas na temática de estudo, para a verificação de possíveis inclusões, exclusões ou modificações de itens e verificar se os itens da escala se comportam nas dimensões especificadas pelos testes e conceituação teórica (DEVELLIS, 2017).

A seção a seguir, para complementar o embasamento necessário para o desenvolvimento da proposta de estudo, destaca os modelos de competências docentes dominantes na literatura.

2.2. COMPETÊNCIAS DOCENTES

O assunto competências docentes está sendo estudado, a partir de cinco modelos. O primeiro (1º) modelo é de cunho **condutivista ou behaviorista** que analisa a competência por meio das tarefas de cada posto de trabalho. A competência reflete a capacidade da pessoa e descreve o que ela pode fazer e não, necessariamente, o que faz, independentemente da situação ou circunstância. O modelo behaviorista está baseado nas teorias psicológicas (Psicologia de Skinner e Pedagogia dos Objetivos de Bloom) que reconhecem apenas os dados passíveis de tratamento empírico. A identificação de competências considera como ponto de partida o conhecimento das características de um indivíduo que apresenta um desempenho profissional considerado superior, para, em seguida, poder transformá-las em normas de aplicação geral (STEFFEN, 1999; MERTENS, 1996). Este modelo é rotulado de racionalista da competência, porque enfatiza os atributos pessoais do trabalhador frente ao cargo ou vice-versa. (McCLELAND, 1973; BOYATZIZ, 1982; SPENCER; SPENCER, 1993; FLEURY; FLEURY, 2000).

O segundo (2º) modelo trata da **teoria da análise funcional**, baseado na escola do pensamento funcionalista da sociologia, aplicado no sistema de competência trabalhista na Inglaterra. O modelo está associado ao projeto *Management Charter Initiative* (M.C.I.), um empreendimento do *National Council for Vocational Qualifications* (N.V.Q.) e do *Scottish Council for Vocational Qualifications* (S.C.O.T.V.E.C.), entidades que implementaram uma iniciativa governamental com o propósito de estabelecer padrões mínimos de desempenho em nível nacional. Ao contrário da abordagem behaviorista ou condutivista, o modelo está centrado na definição de tarefas e a descritores de resultados, sendo esse último considerado o elemento que materializa, de fato, as competências. Parte dos objetivos da organização ou da área de ocupação, tendo como meta a identificação de um perfil ocupacional desejado para servir de referência para a implementação de programas de capacitação e de certificação de competências (MERTENS, 1996).

O terceiro (3º) modelo de competências é o **construtivista** e tem sua origem na França. A normalização de competências é definida com base nos resultados de aprendizagem e pelo desempenho dos indivíduos diante dos problemas e dificuldades enfrentados. A busca

da solução das disfunções na organização é mediada por estratégias de motivação pessoal. Este modelo toma como padrão não os funcionários com desempenho superior, como no modelo condutivista e, sim, naqueles com desempenho inferior. Na visão de Ramos (2001, p. 94), a metodologia construtivista desenvolvida na França com base, dentre outros, nos estudos de Bertrand Schwartz, “rechaça a defasagem entre construção da competência e a norma por um lado e, por outro, a estratégia de capacitação”, pois se propõe a unificar os processos de investigação das competências requeridas e a capacitação em serviço que leva à sua construção.

O modelo construtivista de competências está fundamentado no modelo interativo desenvolvido por Hodkinson (1992), onde as interações entre pessoas e entre pessoas e seu ambiente são consideradas na explicação da competência em educação, complementado pelo modelo cognitivo em que conhecimento e compreensão subjacentes às ações competentes são importantes (REYNOLDS; SALTERS, 1995). Os autores, também fornecem um lembrete sobre a importância da dimensão pessoal da competência do professor. Short (1985) com a intenção de esclarecer os modelos citados, apresentou um quarto (4º) modelo, **contendo quatro diferentes concepções** de competência docente:

- a) Competência como comportamento ou desempenho, o fazer de determinadas coisas independentemente do propósito ou intenção;
- b) Competência como comando de conhecimento ou habilidades, envolvendo a escolha e saber por que a escolha é apropriada;
- c) Competência como nível de capacidade que tem sido "suficiente" através de algum processo judicioso e público. Este indicador de suficiência pode flutuar desde que envolve um julgamento de valor;
- d) Competência como qualidade de uma pessoa ou estado de ser, incluindo mais do que comportamentos característicos: desempenho, conhecimentos, habilidades, níveis de suficiência e qualquer outra coisa que possa parecer relevante, como intenção, motivos, atitudes ou qualidades particulares.

De acordo com Short (1985), a quarta definição evidencia que muitas teorias sobre a competência do professor podem coexistir. Na mesma linha, Cheetham e Chivers (1996) propõem um **quinto (5º)** modelo, ou seja, um modelo de competência mais holístico, buscando integrar elementos dos diversos modelos existentes, protocolos de descrição de competências, literatura sobre educação e desenvolvimento profissional. O modelo proposto pelos autores abrange quatro competências centrais, inter-relacionadas com as metacompetências, quais sejam:

- a) Competências de conhecimento/cognitiva (conhecimento técnico/teórico especializado, conhecimento tácito/prático, conhecimento procedimental e conhecimento contextual);
- b) Competência funcional (ocupacional específica, organizacional/processo gerenciamento e mental);
- c) Competência pessoal ou comportamental (social/vocacional);
- d) Competência de valores e ética (pessoal e profissional)

As metacompetências expressam as competências que desenvolvem e reforçam outras competências, como por exemplo, a capacidade de refletir e de aprender. O contexto de trabalho em que o indivíduo opera e o ambiente de trabalho, também são considerados. Os autores reconhecem que tanto a personalidade como a motivação podem ter impacto nas competências profissionais e a reflexão como uma metacompetência pode melhorar a competência profissional (CHEETHAM; CHIVERS, 1996). Paiva (2007), a partir de estudos dos autores, acrescentou a componente “política”, desenvolvida em função dos comportamentos pessoal, profissional, organizacional e social, baseados nas relações de poder do exercício da profissão.

A competência docente é entendida, desta forma, como a capacidade de mobilizar elementos relativos aos conhecimentos, habilidades, atitudes e valores para responder às demandas da situação de trabalho. Tão importante quanto o nível de aquisição de competências é o senso que o professor tem sobre o seu domínio e aplicação. Para Batista, Matos e Graça (2011, p. 119) “A percepção de domínio das competências necessárias à atividade profissional surge muitas vezes como fator decisivo da competência profissional”. A autoavaliação ou a autoeficácia é entendida como “um julgamento que o professor faz acerca de suas próprias capacidades para atingir resultados desejados de engajamento e aprendizagem dos alunos, mesmo entre aqueles alunos difíceis ou desmotivados” (TSCHANNEN-MORAN; WOOLFOLK HOY, 2001, p. 783). Os indivíduos com maior percepção de autoeficácia acreditam em suas capacidades para lidar com demandas específicas e, conseqüentemente, agem de acordo com as exigências da situação a ser enfrentada (BANDURA, 1986).

Os trabalhos selecionados para a análise são estudos que avaliaram a autopercepção ou a autoeficácia que os professores têm de suas competências. Os docentes são auto examinadores do seu próprio funcionamento, refletindo a sua eficácia, suas motivações e seus valores.

3. DESCRIÇÃO DE ESTUDOS QUE UTILIZARAM ESCALAS DE MEDIDA DE COMPETÊNCIAS DOCENTES

Com o intuito de explorar os estudos com escalas de medição, esta seção apresenta o resumo, abordagem geral e informações características relativas aos tratamentos estatísticos de dados dos nove artigos selecionados para análise proposta pelo estudo sobre o desenvolvimento e aplicação das escalas em competências docentes.

O primeiro (1º) artigo selecionado, intitulado “*The multicultural teaching competency scale: development and initial validation*”, de **Spanierman et al (2011)**, tem o objetivo de relatar o desenvolvimento e validação inicial de uma escala de medida da competência docente no ensino multicultural (Multicultural Teaching Competency Scale – MTCS). Foram encontrados 548 professores para participação e coletados e utilizados os dados de 506 professores respondentes. Na aplicação da análise fatorial exploratória, foram utilizados os dados de 248 professores e a escala com duas dimensões: (1) habilidade de ensino multicultural; e (2) conhecimento de ensino multicultural, com o total de 16 itens em um instrumento de pesquisa com respostas do tipo Likert de 6 pontos. Após, com os dados dos outros 258 respondentes, foi realizada a análise fatorial confirmatória onde foi constatado que o modelo com as duas dimensões era válido. A escala MTCS demonstrou consistência interna adequada e de maneira significativa relacionada as medidas de conscientização do racismo e atitudes de ensino multiculturais. O artigo não informa a população, o coeficiente Alfa de Conbrach e demais aplicações estatísticas.

O segundo (2º) artigo selecionado, intitulado “*The identification of competencies for online teaching success*”, de **Bigatel et al (2012)**, é um estudo exploratório que examinou os comportamentos, atitudes e crenças dos professores para a atividade no ensino online. Foi utilizado um instrumento de pesquisa com respostas em escala Likert de 7 pontos distribuídos para professores com experiência na modalidade online. Após a revisão do instrumento com professores chegou-se a 64 itens relevantes para a realização da pesquisa. Dos 260 pedidos de participação como respondentes, foram respondidos 197 ao final do período de pesquisa. Uma análise fatorial exploratória constatou sete dimensões confiáveis para a: (1) aprendizagem ativa; (2) administração/liderança; (3) ensino ativo/capacidade de resposta; (4) tecnologia multimídia; (5) postura em sala de aula; (6) competência tecnológica; e (7) aplicação política. Mais da metade dos questionamentos sobre as tarefas de ensino tinham uma classificação de 6,0 ou mais, por isso não foram incluídas em categorias usando a análise fatorial. Um exame

adicional dos resultados se fez necessário para determinar por que estas tarefas de ensino altamente classificadas não foram carregadas em nenhuma dimensão. Os resultados da importância das tarefas formaram a base dos esforços de desenvolvimento dos professores, visando proporcionar ao grupo de professores o desenvolvimento profissional em competências críticas para garantir o sucesso do ensino online. O artigo apresenta o coeficiente Alfa de Cronbach e não informa a realização de validação com especialistas, teste t e demais análises estatísticas.

O terceiro (3º) artigo selecionado, intitulado “*Teaching competencies of physical education teachers in primary education (comparative study between accredited and non-accredited schools)*”, de **Awad e Eid (2013)**, tem o objetivo de comparar o nível de competências dos docentes de educação física em escolas credenciadas e não-credenciadas da educação básica, utilizando o método de pesquisa descritiva. Os docentes foram escolhidos aleatoriamente e divididos em uma amostra básica envolvendo 120 docentes de ambos os sexos. A escala de medida foi dividida em seis dimensões: (1) planejamento; (2) implementação; (3) avaliação; (4) auxílio ao ensino; (5) métodos de ensino; e (6) consciência do ambiente. Com um total de 224 itens o questionário foi validado por 5 especialistas com conhecimento e experiências na área e métodos de ensino. Os resultados mais importantes demonstraram que existem diferenças estatisticamente significantes entre as escolas no nível das competências de ensino dos professores de educação física. O artigo apresenta a aplicação do teste t e demais cálculos estatísticos simples e não informa a população do estudo, a realização de coeficiente Alfa de Cronbach e não aplica de análise fatorial exploratória e confirmatória.

O quarto (4º) artigo selecionado, intitulado “*Competencias docentes en secundaria: análisis de perfiles de profesorado*”, de **Ferrández-Berrueco e Sánchez-Tarazaga (2014)**, tem o objetivo de analisar a importância atribuída aos professores da província de Castellon (Espanha) às competências de ensino e ao estabelecimento de perfis profissionais a partir de suas respostas. A pesquisa foi aplicada com questionário contendo 44 itens, com respostas em escala Likert de 4 pontos e respondido por uma amostra de 136 professores da região, do total de uma população de 2000 professores. A escala de medida foi dividida em quatro dimensões: (1) disciplinares; (2) metodológicas; (3) sociais; e (4) pessoais. Foram calculadas a média, desvio padrão e coeficiente de variação, análise da existência de perfis, análise de cluster (Q-cluster) e realizada a caracterização dos perfis encontrados de variáveis pessoais e contextual utilizando o coeficiente Qui-quadrado. A partir de uma análise das respostas, foram identificados três perfis de profissionais com base na importância que os professores atribuem a uma gama de competências. Embora todas as competências tenham sido altamente valorizadas, os resultados evidenciaram que os professores ainda consideram a entrega tradicional de conhecimento como a competência mais relevante. O artigo não informa a validação de especialistas do instrumento, cálculo de hipótese com teste t e análise fatorial exploratória e confirmatória.

O quinto (5º) artigo selecionado, intitulado “*A suggested model for developing and assessing competence of prospective teachers in faculties of education*”, de **Alqiawi e Ezzeldin (2015)**, tem o objetivo fornecer um modelo para avaliar as competências necessárias para os professores dos cursos superiores em educação. O artigo aborda que a gestão de cursos tem procurado um modelo para a avaliação de competências dos professores. À luz das revisões teóricas e das validações com especialistas, os investigadores concluíram que as competências dos professores se concentram em três dimensões: (1) acadêmica; (2) profissional; e (3) pessoal. Para a pesquisa foi utilizado um instrumento com respostas em escala Likert de 3 pontos com 14 itens no seu modelo final. O trabalho pretende que os itens do modelo sirvam como um guia na formação e avaliação de professores em perspectiva e em serviço. A pesquisa foi aplicada em uma amostra de 80 docentes e utilizou métricas analíticas descritivas com correlação de Pearson. O artigo não informa a população do estudo, a

realização de testes de hipóteses e confiabilidade, como teste t e coeficiente Alfa de Cronbach e não aplica de análise fatorial exploratória ou confirmatória.

O sexto (6º) artigo selecionado, intitulado “*Investigating classroom teaching competencies of pre service elementary mathematics teachers*”, de **Gokalp (2016)**, buscou respostas para duas questões centrais em termos de competências docentes em sala de aula: a situação atual dos programas de Educação em Matemática na Faculdade de Educação e como os professores pré serviço adquirem essas competências. A pesquisa foi conduzida e aplicada a partir de questionário com respostas em escala Likert de 5 pontos, validado por especialistas, com 20 itens com declarações de competências, em uma amostra de 202 docentes que atuam no departamento de Educação em Matemática Elementar da Universidade Estadual na Turquia durante os anos de 2012-2013. Os dados foram analisados por meio de estatística descritiva. Os resultados indicaram uma relação significativa e positiva entre as dimensões de competências em preparação e ensino dos professores. A escala de medida foi dividida em quatro dimensões: (1) sobre objetivos; (2) organização orientadas ao conteúdo; (3) para o uso de métodos de ensino; e (4) utilização de ferramentas e materiais no curso. Além disso, houve uma diferença significativa entre o escore de competência em diferentes níveis de fases dos cursos. De acordo com os resultados e observação, os professores eram significativamente competentes em salas de aula do 6º ano do que nas salas de aula do 8º ano. O artigo apresenta o coeficiente Alfa de Cronbach significativo e aplicação de estatística descritiva, ANOVA e correlação. O artigo não informa se foi realizado o teste t e não apresenta análise fatorial exploratória e confirmatória.

O sétimo (7º) artigo selecionado, intitulado “*Development and Validation of Evaluation Indicators for Teaching Competency in STEAM Education in Korea*”, de **Kim e Kim (2016)**, tem o objetivo desenvolver e validar os indicadores de avaliação da competência docente com o modelo STEAM de educação. As competências de ensino em educação STEAM foram elaboradas utilizando entrevista comportamental (BEI) e uma revisão da literatura. Os itens de avaliação das competências iniciais foram revisados por 15 especialistas e dois testes piloto foram conduzidos. Como indicadores finais de avaliação foram compostos por 35 itens em sete dimensões: (1) compreensão dos sujeitos; (2) métodos de ensino-aprendizagem; (3) induzir os alunos a participar da aprendizagem; (4) compreensão dos alunos; (5) ambientes de aprendizagem e circunstâncias; (6) avaliação de aprendizes; e (7) qualificação individual. A versão final do questionário foi aplicada com 208 professores e os dados da pesquisa foram utilizados para validar o modelo utilizado a partir da análise fatorial exploratória e da análise fatorial confirmatória. Os resultados do estudo mostram que os indicadores de avaliação do instrumento são uma diretriz para entender o que realmente importa na educação com o modelo STEAM de educação e como realizar uma aula de STEAM. O artigo apresenta o coeficiente Alfa de Cronbach e análise fatorial exploratória e confirmatória em amostras diferentes. O artigo não informa se foi realizado o teste t, não apresenta a escala de medição do instrumento e demais métricas estatísticas.

O oitavo (8º) artigo selecionado, intitulado “*Developing a validated instrument to measure preservice teachers’ ICT competencies: Meeting the demands of the 21st century*”, de **Tondeur et al (2017)**, possui como principal objetivo desenvolver um instrumento de autorrelato docente para medir as competências de Tecnologia da Informação e Comunicação (TIC) na educação. O questionário de pesquisa possui 19 itens, com respostas em escala Likert de 5 pontos, baseados em uma estrutura criada com a contribuição a validação de especialistas na área. Os dados foram coletados de uma amostra de 931 professores que atuaram um ano antes da pesquisa em Flandres (Bélgica). A amostra foi dividida para a primeira parte ser utilizada para a análise fatorial exploratória e a segunda parte para a análise fatorial confirmatória. Com isso foi identificada uma estrutura de duas dimensões de competências em TIC: (1) competências para apoiar os alunos na utilização das TIC nas aulas; e (2) competências para utilizar as TIC no design instrucional. A estrutura de duas

3º	O formato de medição foi escolhido com propósito?	SIM	SIM	NÃO	SIM	SIM	SIM	NÃO	SIM	SIM
4º	O instrumento foi validado por especialistas?	SIM	NÃO	SIM	NÃO	SIM	SIM	SIM	SIM	SIM
5º	Foram realizados ajustes, inclusão ou exclusão de itens?	SIM	NÃO	SIM	PAR	SIM	SIM	SIM	SIM	NÃO
6º	A amostra contém pelo menos cinco respondentes por item?	SIM	NÃO	NÃO	NÃO	SIM	SIM	SIM	SIM	SIM
7º	Foram realizados exames de desempenho dos itens?	SIM	SIM	SIM	SIM	SIM	SIM	SIM	SIM	SIM
8º	O tamanho da escala e divisão da amostra é suficiente?	SIM	SIM	SIM	SIM	SIM	SIM	SIM	SIM	SIM

Fonte: Os autores (2018)

O primeiro (1º) questionamento **“O que será medido está definido com teoria e hipóteses?”** foca na construção teórica do construto que o artigo pretende medir e se existe a formulação de hipóteses para o estudo. Neste questionamento, todos os artigos selecionados utilizaram de embasamento teórico para subsidiar a opção das dimensões e construto apresentado no estudo. Apenas dois artigos definem e apresentam as hipóteses a serem confirmadas ou rejeitadas no estudo, que são os artigos de Spanierman et al (2011) e Onyilo e Shamo (2017). Na percepção de DeVellis (2017), é importante que sejam conhecidas as fronteiras do fenômeno pesquisado para que o conteúdo da escala desenvolvida não se desvie dos domínios desejados. Segundo Hair et al (2005), o pesquisador deve conhecer o que está sendo estudado, possuir clara definição dos conceitos a serem medidos e o que é esperado como objetivo e problema da pesquisa, para somente assim conseguir criar um questionário organizado e condizente. Por isso o aprofundamento teórico e conhecimento das hipóteses são relevantes para a consolidação do instrumento.

O segundo (2º) questionamento **“Os itens escolhidos refletem o propósito da escala?”** consolida a geração de itens para a escala utilizada nos artigos, desde a quantidade apresentada até o propósito de cada item fazer parte do instrumento. Diante da análise dos artigos selecionados, todos evidenciam, de alguma maneira, a geração, utilização da teoria para endossar o processo, redesenho de perguntas e ajustes de características dos itens.

Os procedimentos utilizados por Spanierman et al (2011), na geração de itens, exemplificam a questão:

A equipe de pesquisa dividiu-se em três subequipes e gerou itens que refletiu as dimensões mencionadas da competência de ensino multicultural. A lista preliminar de itens gerados por cada subequipe circulou para as outras duas subequipes para revisão. As subequipes incorporaram todos feedbacks, resultando em 57 itens. Para garantir o conteúdo, clareza e parcimônia de cada um dos 57 itens, os itens foram distribuídos para estudantes de pós-graduação que tinham concluído cursos extensivos em relação à competência multicultural. Vários itens foram modificados para refletir o feedback dos alunos. Também foi procurado conselhos adicionais de especialistas em formação de professores (ou seja, três membros do corpo docente de um programa de desenvolvimento de professores com especialização em avaliação educacional e dois professores no campo) para comentar o nível de adequação do conteúdo, bem como a clareza de cada item. Itens que foram acordados pela maioria dos especialistas como congruente com o domínio de conteúdo foram retidos; um item foi descartado devido a ambiguidade, resultando em uma escala preliminar de 56 itens.

DeVellis (2017) reforça a questão quando ressalta que os itens devem ser selecionados e ou criados com o objeto de medida bem definido e que a descrição do item precisa refletir exatamente o construto que a escala pretende medir. Além disto, um bom conjunto de itens é escolhido randomicamente a partir de um universo de itens relacionados ao construto geral, ampliando assim a gama de opções de questionamentos. Segundo Hair et al (2005), os

indicadores de mensuração devem ser determinados desde a construção do questionário considerando sua extensão, estrutura das perguntas, ordem e codificação.

O terceiro (3º) questionamento **“O formato de medição foi escolhido com propósito?”** colabora no entendimento da escolha do método de medição e coleta dos dados para o formato de análise na pesquisa. Na seleção da escala de medição, sete dos nove artigos utilizaram escala Likert como parâmetro de respostas das questões de pesquisa, variando as escalas entre 4 e 7 pontos. Os dois artigos que não apresentam o formato de medição foram os de Awad e Eid (2013) e Kim e Kim (2016). Segundo DeVellis (2017), a definição do formato de medição é uma etapa que deve ocorrer simultaneamente com a geração de itens do instrumento, para assegurar compatibilidade de alcance dos resultados do construto, além de ajudar nas opções de análises que serão aplicadas no estudo. Segundo Hair et al (2005, p. 218), a tipologia de um questionário depende da “natureza do tema, da administração do questionário, da habilidade e vontade da população-alvo de responder, do tipo de análise estatística e de fatores semelhantes”.

O quarto (4º) questionamento **“O instrumento foi validado por especialistas?”** garante a realização de uma das etapas mais relevantes na aplicação e desenvolvimento de uma escala de medição. Dois artigos selecionados não evidenciam a realização desta etapa na aplicação e validação do instrumento de pesquisa: Bigatel et al (2012) e Ferrández-Berruero e Sánchez-Tarazaga (2014). Os outros artigos, sete no total, informam a existência desta etapa e possuem uma variação no perfil dos especialistas, com alunos, professores ou gestores institucionais, com um montante oscilando entre 5 e 15 participantes para avaliação dos itens. O exemplo citado no passo 2 acerca dos procedimentos utilizados por Spanierman et al (2011), reforçam os argumentos de DeVellis (2017) sobre o papel dos especialistas. Para o autor, a validação com especialistas que avança desde a definição das subdimensões dos itens com o intuito de maximizar a validade de conteúdo, verificar a relevância dos itens, avaliar a clareza e consistência dos itens e verificar se há outros itens que não foram trabalhados. Hambleton e Rogers (1991) sugerem que os novos instrumentos de avaliação sejam revisados quanto à qualidade técnica (gramática, redação, randomização de itens e escalonamento) por especialistas da área de avaliação e medição. Haynes et al. (1995) sugerem a avaliação e correção do conteúdo do questionário por parte de especialistas na área de conhecimento como importante etapa de refinamento do instrumento. A confiança e robustez do instrumento aumentará à medida que o número de especialistas aumentar, já que facilitará a identificação e exclusão itens *outliers*.

O quinto (5º) questionamento **“Foram realizados ajustes, inclusão ou exclusão de itens?”** busca entender de que forma a participação dos envolvidos na etapa anterior foram conclusivas e contribuíram para a aplicação da pesquisa. Seis artigos analisados relataram e submeteram os itens prévios para validação de especialistas e conseguiram melhorar, incluindo, excluindo ou ajustando, parte do instrumento antes da aplicação final. Os artigos de Ferrández-Berruero e Sánchez-Tarazaga (2014) relatam ter passado por ajustes nos itens, entretanto não evidencia a realização da etapa de validação com especialistas na área. Os outros dois artigos dos autores Bigatel et al (2012) e Onyilo e Shamo (2017) não informam se ocorreu algum tipo de ajuste, inclusão ou exclusão de itens e afirmam em metodologia que o questionário foi enviado diretamente aos respondentes. Segundo DeVellis (2017), os ajustes são essenciais para detectar de falhas conceituais e garantir a validade do construto da escala. Anastasi (1988) afirma que os itens em um instrumento de avaliação devem ser distribuídos, ou ponderados, de uma maneira que reflita a importância relativa das várias facetas do construto alvo.

O sexto (6º) questionamento **“A amostra contém pelo menos cinco respondentes por item?”**, evidencia uma deficiência das pesquisas que utilizam amostras inferior à necessária para alcançar uma segurança dos resultados. Dos artigos selecionados, três apresentam amostras abaixo do indicado estatisticamente por DeVelis (2012), ou seja, não

apresentam o mínimo de 5 itens por respondentes, são eles: dos autores Bigatel et al (2012), Awad e Eid (2013) e Ferrández-Berrueco e Sánchez-Tarazaga (2014). Os outros seis trabalhos apresentam mais de 5 respondentes por itens do instrumento. Segundo DeVellis (2017), um dos problemas em usar amostra pequena é o impacto no padrão de covariação entre os itens, que pode não ser estável, e a amostra de desenvolvimento pode não representar a população para a qual a escala é destinada. Como alerta para garantia de qualidade da medição, Haynes et al (1995) relata que os procedimentos de validação de conteúdo e índices de validade do instrumento devem ser tratados como categorias importantes no processo de construção e ser reportados sistematicamente no mesmo detalhe que outros componentes de validação.

O sétimo (7º) questionamento **“Foram realizados exames de desempenho dos itens?”** faz menção às aplicações de técnicas estatísticas nas bases de pesquisa, dados e ou resultados da pesquisa realizada pelos estudos. Este questionamento busca evidenciar se a escala passou por exames de ajuste e confiabilidade, além de etapas relevantes para garantia dos melhores resultados finais. Nesta análise, todos artigos selecionados apresentaram evidências e resultados da aplicação de técnicas estatísticas, como: teste t, normalização de dados, confiabilidade de dados, adequação de dados e análise fatorial. Segundo DeVellis (2017) a avaliação do desempenho dos itens é o coração do processo de desenvolvimento da escala e possivelmente a segunda parte mais importante de toda construção da pesquisa e seus resultados

Os fundamentos expostos por Kim e Kim (2016), em seu estudo, reforçam os argumentos relacionados ao exame de desempenho dos itens. Os autores realizaram:

Análises da qualidade, confiabilidade e validade dos itens foram realizadas por meio de dois questionários. Para analisar a qualidade dos itens, foram utilizadas estatísticas descritivas como média (M), desvio padrão (DP), assimetria, curtose e grau de correlação. Para análise de confiabilidade, foi utilizado o coeficiente α de Cronbach para o grau de qualidade dos itens. A confiabilidade foi assegurada por meio de um processo elaborado para excluir ou modificar itens com baixa confiabilidade, confirmando os valores no momento da eliminação do item. Além disso, a validade de construto dos itens do indicador de avaliação foi verificada por meio da análise fatorial exploratória (AFE).

Na sequência, os autores Kim e Kim (2016) realizaram a Validação da Análise Confirmatória com 208 professores que estavam realizando ou realizaram a educação STEAM como sujeitos para a validação dos indicadores de avaliação desenvolvidos. A versão revisada dos indicadores de avaliação da competência docente ficou composta por 36 itens.

O oitavo (8º) e último questionamento **“O tamanho da escala e divisão da amostra é suficiente?”** busca entender após a aplicação do instrumento e análise de resultados a necessidade de ajustes de itens, dimensões e tamanho da escala. No caso, todos artigos conseguiram otimizar e revisar o instrumento a ponto de alcançar resultados conclusivos para os estudos propostos em suas concepções. DeVellis (2017) relata que a otimização da escala é importante para o momento de aplicação da pesquisa, no entanto, se o pesquisador não conseguir atribuir pontuações e significados nas versões mais curtas e otimizadas, de nada valerá o esforço, por isso, a questão da brevidade do instrumento deve ser confiada às situações de aplicabilidade da pesquisa.

A análise individual dos artigos em relação aos critérios propostos na metodologia de DeVellis (2017), revela que o artigo de Spanierman et al (2011), atende 100% dos passos recomendados. Nos artigos de Alqiawi e Ezzeldin (2015), Gokalp (2016) e Tondeur et al (2017) o atendimento é parcial em apenas um critério dos passos de exposição das hipóteses. Os demais artigos possuem pelo menos um dos critérios não atendidos ou não apresentam evidências de realização.

Os achados evidenciaram que o uso de escalas de autoavaliação de competências docentes é de natureza plural, reforçada pelas diferentes dimensões utilizadas na geração de itens para compor o instrumento de coleta de dados, demonstrando que não há um modelo de

autoavaliação das competências docente amplamente aceito. Também, os estudos adotaram uma variedade de procedimentos metodológicos, revelando que nem todos os passos estabelecidos na metodologia de DeVellis (2017) são considerados. Por esta razão, muitos achados de pesquisas acerca do tema têm apresentado fragilidades na estrutura e na confiabilidade dos resultados.

A confiabilidade dos resultados está relacionada com todos os passos descritos pelo autor. Por exemplo, se o pesquisador não definir as fronteiras do fenômeno, a partir de teorias, poderá gerar itens nem sempre relevantes com as dimensões pretendidas pelo estudo. A não definição e entendimento de um dos passos poderá comprometer todos os outros passos por serem interdependentes. Um dos critérios principais de qualidade de um instrumento de coleta de dados é a confiabilidade, que demonstra a capacidade de reprodução dos resultados, independentemente do tempo e de uso por outros pesquisadores, indicando coerência, precisão, estabilidade, equivalência e homogeneidade (TERWEE, 2007).

5. CONSIDERAÇÕES FINAIS

Os artigos selecionados neste estudo enfatizavam a autoavaliação ou autoeficácia de competências docentes. Os artigos foram submetidos a verificação e análise da existência dos oito passos estabelecidos na metodologia de DeVellis (2017), a fim de validar o desenvolvimento e aplicação da escala de medida utilizada.

Os pontos relevantes na construção de escalas destacados por DeVellis (2017) evidenciam a necessidade de ressaltar a especificidade do fenômeno frente ao construto na elaboração dos itens, sempre alinhados à variável latente. Conhecer os limites do fenômeno pesquisado garante maior direcionamento para o alcance dos objetos e domínios desejados na mensuração dos dados e resultados. Para tal, o aprofundamento teórico e o conhecimento claro das hipóteses são condições primordiais para a consolidação do instrumento. Mesmo assim, a subjetividade e experiência do pesquisador também fazem parte da construção e definição dos itens para a escala de medição.

A recomendação na geração de itens é que estes reflitam o propósito da escala e sejam em uma quantidade significativa, para que seja possível eventuais ajustes, agrupamentos e exclusões, com o intuito de consolidar um instrumento de medição válido e eficiente. O processo de escrever itens, entendendo a redundância e mensurando a quantidade suficiente de itens para a escala, devem ser selecionados e ou criados com o objeto de medida bem definido. A descrição do item que comporá o instrumento de pesquisa precisa refletir exatamente o construto que a escala pretende medir, juntamente com uma base teórica e um bom conjunto de opções. Além disto, a literatura apresenta diversos formatos possível de questionamentos para os itens e caberá ao pesquisador, de acordo com a compatibilidade do tema, definir o melhor formato para medição, que deve ocorrer simultaneamente com a geração de itens do instrumento, para assegurar compatibilidade de alcance dos resultados da pesquisa e construto (DEVELLIS, 2017).

Um universo de itens relacionados ao construto geral é necessário para a etapa de seleção e verificação de especialistas, que auxiliarão na maximização do instrumento e validação do conteúdo dos questionamentos. Por meio da verificação de especialistas na área e construto, a relevância dos itens, a clareza e a consistência dos itens são garantidas para a continuidade e confirmação da escala de medida. Neste sentido, o papel dos especialistas é auxiliar no avanço e definição das subdimensões dos itens com o intuito de validar o conteúdo, verificar a relevância dos itens, avaliar a clareza e consistência dos itens e verificar se há outros itens necessários que não foram apresentados. Estes ajustes são a essências para detectar falhas conceituais e garantir a validade do construto da escala (DEVELLIS, 2017).

Para uma boa pesquisa e resultados de relevância científica, o instrumento de medição deve ser aplicado em uma amostra estatisticamente compatível, que deve ser de no mínimo

cinco respondentes para cada item de medição e ou de pelo menos trezentos indivíduos. Se estes parâmetros forem respeitados, os resultados obtidos pelas escalas de medição serão consistentes. Um exame inicial de performance dos itens também é necessário, já que a verificação dos itens e o significado dos itens, através da análise fatorial, é considerada a melhor maneira para se determinar as dimensões de análise. Já o cálculo do coeficiente Alpha de Cronbach avalia as correlações dentro de cada uma das dimensões encontradas e traz segurança no desempenho dos itens e dimensões da escala (DEVELLIS, 2017).

Contudo, a otimização da escala de medida é importante para o momento de aplicação da pesquisa. De nada valerá ter um instrumento relativamente completo, se o pesquisador não conseguir aplicabilidade da pesquisa. Por isso, a questão da brevidade do instrumento deve ser confiada e atribuída nas versões mais curtas e otimizadas, desde que atenda a expectativa de alcance dos resultados do construto pesquisado (DEVELLIS, 2017).

Por fim, o estudo avança na análise e compreensão das escalas de medição das competências docentes como instrumentos embasados teoricamente, confiáveis e válidos. Como contribuição do trabalho considera-se a identificação das pesquisas com o foco na autoavaliação e autopercepção dos docentes, das diferentes dimensões e categorias de análise no tema e subsídios que permitem a replicação em outras realidades, principalmente no cenário brasileiro.

REFERÊNCIAS

ALQIAWI, D. A; EZZELDIN, S. M. A suggested model for developing and assessing competence of prospective teachers in faculties of education. **World Journal of Education**, v. 5, n. 6, p. 65-73, 2015.

ANASTASI, A. **Psychological testing**. New York: Macmillan, 1988.

AWAD, K. T; EID, A. I. Teaching competencies of physical education teachers in primary education (comparative study between accredited and non-accredited schools). **Turkish Journal of Sport and Exercise**, v. 15, n. 2, p. 16-20, 2013.

BATISTA, P.; MATOS, Z.; GRAÇA A. Autopercepção das competências profissionais em profissionais do desporto: efeito da área de intervenção e da experiência profissional. **EBM RECIFE**, v. 7, n. 2, p. 117-131, 2011.

BIGATEL, P. M. et al. The identification of competencies for online teaching success. **Journal of Asynchronous Learning Networks**, v. 16, n. 1, p. 59-78, 2012.

CHEETHAM, G.; CHIVERS, G. Towards a holistic model of professional competence. **Journal of European Industrial Training**, v. 20, n. 5, p. 20-30, 1996.

COSTA, F. J. **Mensuração e desenvolvimento de escalas: aplicações em administração**. Rio de Janeiro: Editora Ciência Moderna Ltda., 2011.

DEVELLIS, R. F. **Scale development: theory and application**. Fourth edition. Los Angeles, CA: Sage Publications, 2017.

FLEURY, A; FLEURY, M. T. L. **Estratégias empresariais e formação de competências**. São Paulo: Atlas, 2000.

FERRÁNDEZ-BERRUECO, R; SÁNCHEZ-TARAZAGA, L. Competencias docentes en secundaria: análisis de perfiles de profesorado. **Revista ELelectrónica de Investigación y Evaluación Educativa**, v. 20, n. 1, p. 01-20, 2014.

GOKALP, M. Investigating classroom teaching competencies of pre service elementary mathematics teachers. **Eurasia Journal of Mathematics, Science & Technology Education**, v. 12, n. 3, p. 503-512, 2016.

HAIR JR., J. F. et al. **Fundamentos de métodos de pesquisa em administração**. Porto Alegre: Bookman, 2005.

HAMBLETON, R. K; ROGERS, H. J. Advances in criterion-references measurement. In: HAMBLETON, R. K; ZAAL, J. N (Eds.), **Advances in educational and psychological testing: Theory and applications**. Boston: Kluwer Academic, 1991, p. 3-43.

HAYNES, S. N; RICHARD, D. C. S; KUBANY, E. S. Content validity in psychological assessment: a functional approach to concepts and methods. **Psychological Assessment**, v. 7, n. 3, p. 238-247. 1995.

HODKINSON, P. Alternative models of competence in vocational education and training. **Journal of Further and Higher Education**, v. 16, n. 2, p. 30-39, 1992.

KIM, B.; KIM, J. Development and validation of evaluation indicators for teaching competency in STEAM education in Korea. **Eurasia Journal of Mathematics, Science & Technology Education**, v. 12, n. 7, p. 1909-1924, 2016.

MADURO, M. R. Identificação do perfil de competências docentes em uma instituição estadual de ensino superior. **TAC**, v. 3, n. 2, p. 79-94, 2013.

MERTENS, L. **Sistemas de competencia laboral: surgimento y modelos**. Resumen Ejecutivo. Guanajuato, México, mayo, 1996.

ONYILO, B. O; SHAMO, I. I. Teacher competence as a tool for effective and sustainable human capital development in the federal capital territory: Implications for career guidance. **International Journal of Psychology and Counselling**, v. 9, n. 3, p. 17-22, 2017.

PAIVA, K. C. M. **Gestão de competências e a profissão docente um estudo de caso em universidades no Estado de Minas Gerais**. 2007. Tese (Doutorado) - Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil.

PASQUALLI, L. **Análise fatorial para pesquisadores**. Laboratório de Pesquisa em Avaliação e Medida (LabPAM) – Instituto de Psicologia. Brasília: Universidade de Brasília, 2012.

PERRENOUD, P. **Dez novas competências para ensinar**. Porto Alegre: Artmed, 2015.

RAMOS, M. N. **Da qualificação à competência: deslocamento conceitual na relação trabalho-educação**. 2001. Tese (Doutorado) - Universidade Federal Fluminense, Niterói, Rio de Janeiro.

REYNOLDS, A. What is competent beginning teaching? A review of the literature. **Review of Educational Research**, v. 62, n. 1, p. 1-35, 1992.

REYNOLDS, M.; SALTERS, M. Models of competence and teacher training. **Cambridge Journal of Education**, v. 25, n. 3, p. 349-359, 1995.

SHORT, E. C. The concept of competence: Its use and misuse in education. **Journal of Teacher Education**, v. 36, n. 2, p. 2-6, 1985.

SPANIERMAN, L. B; OH, E; HEPPNER, P. P; NEVILLE, H. A; MOBLEY, M; WRIGHT, C. V; DILLON, F. R; NAVARRO, R. The multicultural teaching competency scale: development and initial validation. **Urban Education**, v. 46, n. 3, p. 440–464, 2011.

SPENCER, L.; SPENCER, S. **Competence at work: a model for superior performance**. New York: Wiley, 1993.

STEVENS, S.S. On the theory of scales of measurement. **Science**, v. 103, n. 2684, p. 677-680, jun. 1946.

STEFFEN, I. **Modelos de competência profissional**. Turin: OIT, 1999.

TERWEE, C. B. et al. Quality criteria were proposed for measurement properties of health status questionnaires. **Journal of Clinical Epidemiology**, v. 60, n. 1, p. 34-42, 2007.

TSCHANNEN-MORAN, M.; WOOLFOLK HOY, A. Teacher efficacy: capturing an elusive construct. **Teaching and Teacher Education**, v. 17, p. 783-805, 2001.

TONDEUR, J; AESAERT, K; PYNOO, B; BRAAK, J. V; FRAEYMAN, N; ERSTAD, O. Developing a validated instrument to measure preservice teachers' ICT competencies: meeting the demands of the 21st century. **British Journal of Educational Technology**, v. 48, n. 2, p. 462-472, 2017.

URBINA, S. **Fundamentos da testagem psicológica**. Porto Alegre: Artmed, 2007.