

A Teoria de Resposta ao Item Aplicada em Avaliações da Educação Profissional e Tecnológica: uma revisão sistemática da literatura

THIAGO KORB^I

DALTON FRANCISCO DE ANDRADE^{II}

PATRÍCIA DE SÁ FREIRE^{III}

<http://dx.doi.org/10.22347/2175-2753v11i33.2095>

Resumo

A avaliação da educação profissional e tecnológica é uma lacuna, não existindo sistema de avaliação nacional e nem metodologia consolidada que atendam suas particularidades. Este artigo tem como objetivo identificar quais experiências de aplicação da Teoria de Resposta ao Item (TRI) na avaliação da educação profissional e tecnológica já foram pesquisadas e registradas cientificamente. A metodologia utilizada foi a Revisão Sistemática da Literatura (RSL). Após análise na base de dados Scopus, foram escolhidos cinco artigos que configuram a amostra final deste estudo. Os artigos selecionados são relativamente recentes (todos a partir do ano 2000) e relatam experiências internacionais sobre o problema de pesquisa. Constatou-se que algumas iniciativas foram implementadas nesse intuito, contudo ainda são pouco exploradas, demonstrando a necessidade de pesquisas científicas para aprofundamento e instrumentos de avaliação com viés prático para aferição das competências profissionais.

Palavras-chave: Teoria de Resposta ao Item. Educação Profissional e Tecnológica. Avaliação.

Submetido em: 14/12/2018

Aprovado em: 21/10/2019

^I Universidade Federal de Santa Catarina (UFSC), Florianópolis (SC), Brasil; <http://orcid.org/0000-0002-4362-4273>; e-mail: thiagokorb@sc.senai.br.

^{II} Universidade Federal de Santa Catarina (UFSC), Florianópolis (SC), Brasil; <http://orcid.org/0000-0002-4403-980X>; e-mail: dalton.andrade@ufsc.br.

^{III} Universidade Federal de Santa Catarina (UFSC), Florianópolis (SC), <http://orcid.org/0000-0002-9259-682X>; e-mail: patriciadesafreire@gmail.com.

Item Response Theory Applied in Vocational and Technological Education Assessments: A Systematic Review of the Literature

Abstract

The evaluation of professional and technological education is a gap, and there is no national evaluation system and no consolidated methodology that meets its particularities. This article aims to identify which experiences of application of the Item Response Theory (IRT) in the evaluation of professional and technological education have been researched and registered scientifically. The methodology used was the Systematic Literature Review (SLR). After analysis in the Scopus database, five articles were chosen and configure the final sample of this study. Selected articles are relatively recent (all from the year 2000) and report international experiences on the research problem. It was verified that some initiatives were implemented in this way, but they are still little explored, demonstrating the need for scientific research that deepens and for evaluation tools with a practical bias for measuring of professional skills.

Keywords: Item Response Theory. Professional and Technological Education. Evaluation.

Teoría de Respuesta al Ítem Aplicada en Evaluaciones de la Educación Profesional y Tecnológica: una revisión sistemática de la literatura

Resumen

La evaluación de la educación profesional y tecnológica es un vacío, y no existe un sistema nacional de evaluación ni metodología consolidada que cumpla con sus particularidades. Este artículo tiene como objetivo identificar qué experiencias de aplicación de la Teoría de Respuesta al Ítem (TRI) en la evaluación de la educación profesional y tecnológica han sido investigadas y registradas científicamente. La metodología utilizada fue la Revisión Sistemática de la Literatura (RSL). Después del análisis en la base de datos Scopus, se eligieron cinco artículos que constituyen la muestra final de este estudio. Los artículos seleccionados son relativamente recientes (todos a partir del año 2000) e informan experiencias internacionales sobre el problema de investigación. Se descubrió que algunas iniciativas se han implementado con este fin, pero aún se exploran poco, lo que demuestra la necesidad de investigación científica para profundización, e instrumentos de evaluación con sesgo práctico para medir las habilidades profesionales.

Palabras clave: Teoría de Respuesta al Ítem. Educación Profesional y Tecnológica. Evaluación.

Introdução

O Decreto nº 7.690, de 2 de março de 2012 (BRASIL, 2012), que estabelece as competências da Secretaria de Educação Profissional e Tecnológica (Setec), do Ministério da Educação (MEC), corrobora com a perspectiva de que é necessário “estabelecer diretrizes para as ações de expansão e avaliação da educação profissional e tecnológica em consonância com o Plano Nacional de Educação (PNE)”. Entretanto, não há uma política pública de avaliação nessa vertente no país, o que demanda reflexão e análise urgente frente à ampliação da oferta de vagas nessa modalidade, catalisada em especial pelo Programa Nacional de Acesso ao Ensino Técnico e Emprego (Pronatec), instituído pela Lei nº 12.513, de 26 de outubro de 2011 (BRASIL, 2011). Em 2015, o MEC estruturou um grupo de trabalho para discutir uma política nessa vertente, porém o processo caminha lentamente e sem previsão de implantação.

Atualmente, no Brasil, não há um sistema de avaliação de larga escala para essa frente, a exemplo do Sistema Nacional de Avaliação da Educação Superior (Sinaes) e o Sistema de Avaliação da Educação Básica (Saeb), o que aponta uma lacuna e o não atendimento dos artigos 206, inciso VII, que estabelece como princípio a “garantia do padrão de qualidade” e 209, inciso II, “autorização e avaliação de qualidade pelo Poder Público”, respaldados pela Lei de Diretrizes e Bases da Educação Nacional (LDB), Lei nº 9.394, de 20 de dezembro de 1996 (BRASIL, 1996). Nos sistemas elencados, a análise do desempenho de proficiência dos estudantes é um dos principais componentes, configurando-se como um importante resultado e entendimento de que essa modalidade de ensino poderia seguir o mesmo caminho.

Diante dessa necessidade, surge a demanda de criar uma metodologia que seja capaz de fazer esse tipo de avaliação. Entretanto, o modelo convencional, por meio de testes cognitivos, não supre todas as necessidades. A sua complexidade é maior, como destaca Chico Soares, ex-presidente do Instituto Nacional de Estudos e Pesquisas Educacionais (INEP): “as questões de um exame para estudantes, por exemplo, não podem ser lápis e papel. Uma das saídas é utilizar simuladores nas situações problema. O importante na educação profissional (e tecnológica) é o saber fazer” (AFONSO, 2015). Esta declaração indica a necessidade de aplicar testes práticos para avaliação das competências técnicas, principalmente no que tange ao domínio psicomotor da aprendizagem.

Para a realização de exames dessa magnitude, é necessário fundamentar-se em alguma teoria com respaldo na comunidade científica e que apresente resultados possíveis de comparabilidade entre os sujeitos. A Teoria de Resposta ao Item (TRI) apresenta fatores importantes que colaboram para sua elegibilidade, como a experiência brasileira em sistemas públicos de avaliação e seus pressupostos epistemológico de análise longitudinal dos resultados. Andrade, Tavares e Valle (2000) elucidam que a TRI é um conjunto de modelos matemáticos que visa representar a probabilidade que um indivíduo possui em acertar um determinado item (questão) em função da sua habilidade (traço latente) do respondente, sendo que quanto maior a habilidade, maior a probabilidade de acerto.

Ante ao exposto, este artigo apresenta como problema de pesquisa o seguinte questionamento: quais experiências de aplicação da Teoria de Resposta ao Item, na avaliação da educação profissional e tecnológica, já foram pesquisadas e registradas cientificamente?

Desse modo, para responder esta questão, foi realizada uma pesquisa exploratória-descritiva, por meio de Revisão Sistemática da Literatura (RSL), para analisar os casos já relatados anteriormente em publicações científicas.

Pressupostos da Teoria de Resposta ao Item

Os estudos e conceitos básicos da TRI tiveram início com o precursor Lawley (1943), fortalecidos posteriormente com as pesquisas desenvolvidas por Lord (1952). Nas décadas de 1950 e 1960, sua aplicação estava voltada para os testes de inteligência, contudo os modelos e algoritmos aplicados eram complexos de ser operacionalizado para a época. No Brasil, as primeiras aplicações começaram em 1995 por meio da pesquisa AVEJU, da Secretaria de Estado da Educação de São Paulo, também no Sistema de Avaliação do Rendimento Escolar do Estado de São Paulo (SARESP) e no Sistema de Avaliação da Educação Básica (ANDRADE; TAVARES; VALLE, 2000).

Pasquali (1997) afirma que recentemente a utilização da TRI ampliou-se para várias áreas do conhecimento, sobretudo a da avaliação educacional, que emprega modelos para aferição de traços latentes, entendidos como características que não podem ser observadas diretamente. Desse modo, entende-se por TRI um conjunto de modelos matemáticos e estatísticos que expressam a relação entre a

probabilidade de responder de modo correto ou positivo um grupo de itens e o traço latente necessário para isso, de modo que quanto maior o nível do traço latente, maior a probabilidade de acerto ao item.

Assim, a TRI caracteriza-se por ser uma poderosa ferramenta que supre necessidades da limitada Teoria Clássica da Medida (TCM) ou Teoria Clássica do Teste (TCT). Isso porque se tem como essência os itens e não a prova como um todo, o que permite a comparabilidade entre populações distintas submetidas, inclusive, a provas diferentes (ANDRADE; VALLE, 1998). Essa teoria também amplia a possibilidade de avaliar parâmetros que vão além da dificuldade do item (como usado pela TCT), podendo ser empregados outros como a discriminação e a probabilidade de resposta correta dada por indivíduo de baixa habilidade (ANDRADE; TAVARES; VALLE, 2000).

Demars (2010, p. 3) aponta que os itens avaliativos podem ser dicotômicos, quando envolvem apenas duas categorias, tais como certo ou errado, sim ou não, concordo ou discordo. Ou politômicos (mais de duas categorias), como nos casos utilizados para pesquisas de qualidade de vida com escala *Likert*¹, como exemplo. Independentemente do número de categorias, o construto medido poderá ser aplicado para aferição de proficiência acadêmica ou aptidão, ou até mesmo uma atitude ou crença.

Como sua abrangência é ampla e possibilita a construção de diferentes tipos de escala, a TRI começou a ser empregada em diferentes contextos, não somente no campo educacional, mas no social e econômico também. Especificamente sobre o campo da educação, a TRI vem sendo amplamente utilizada na educação básica, citando como exemplos os casos do SAEB, do Exame Nacional do Ensino Médio (ENEM) e do *Programme for International Student Assessment (PISA)*. Já para as outras dimensões, como a educação profissional e tecnológica, pouco se discute sobre avaliação de larga escala e principalmente aplicação de um ou outro tipo de teoria avaliativa.

¹ Criada por Rensis Likert, trata-se de um tipo de escala de respostas psicométricas em que os participantes especificam o nível de concordância/discordância com uma determinada afirmação, sendo utilizadas, usualmente, cinco categorias do tipo "discordo totalmente" – "discordo parcialmente" – "indiferente" – "concordo parcialmente" – "concordo totalmente", entretanto encontram-se variações com sete ou nove níveis.

Contextualização da educação profissional e tecnológica no Brasil

As primeiras iniciativas em estabelecer uma relação entre educação e trabalho foram advindas da economia de subsistência praticada pelas colônias. A Companhia de Jesus migrou padres jesuítas para praticarem no Brasil suas especialidades profissionais e, concomitantemente, ensinar os escravos para desenvolver atividades manuais e suprir a carência de mão de obra da época. Daí advém o preconceito contra os trabalhos que exigiam esforço físico ou utilização das mãos, os quais eram realizados por índios e afrodescendentes (MANFREDI, 2002, p. 68-69).

No período da Primeira República (1889–1929), o ensino profissional e tecnológico passou por uma metamorfose. O presidente Nilo Peçanha instituiu a criação das escolas de aprendizes e artífices para o ensino profissional gratuito, facilitando acesso às classes proletarizadas para vencer as “dificuldades sempre crescentes na luta pela existência” (BRASIL, 1909). Após, com o lançamento de um projeto industrial no país, a Constituição Federal de 1937 (BRASIL, 1937) determinou a obrigatoriedade da organização de escolas de aprendizes pelas empresas e sindicatos, nascendo então os Serviços Nacionais de Aprendizagem para atender essa demanda. Esse foi um marco, pois esse tipo de educação sempre esteve associado às pessoas de baixa renda, sem identidade, desafortunados e de desvalidos de sorte (KIRSCHNER, 1993, p. 14).

Ao longo dos anos, ocorreram poucos avanços em termos de regulamentação. Entretanto, a partir do Decreto nº 5.154, de 23 de julho de 2004 (BRASIL, 2004), e, posteriormente, pela Lei nº 11.700, de 13 de junho de 2008 (BRASIL, 2008), ocorreram mudanças significativas nas políticas dessa modalidade de educação no País. Isso porque, anteriormente, grande parte da oferta dessa modalidade era realizada pela sociedade civil e como mostra o censo de 2003, 55% pela iniciativa privada. Mais recentemente, entre 2011 e 2013, essa relação se inverteu chegando a 53% da oferta pela rede pública, conforme dados do INEP de 2014, aumento esse atribuído ao Pronatec (CGEE, 2015, p. 104-108).

A educação profissional e tecnológica está regulamentada pela LDB que “no cumprimento dos objetivos da educação nacional, integra-se aos diferentes níveis e modalidades de educação e às dimensões do trabalho, da ciência e da tecnologia” (BRASIL, 1996). Desse modo, vincula-se a três níveis: básico, técnico e tecnológico. Os

cursos básicos são destinados a qualquer cidadão, independentemente da escolaridade prévia; os técnicos ofertados concomitantemente ao ensino médio ou para os egressos, possuindo organização curricular própria; e os tecnológicos são cursos de nível superior (BRASIL, 1996).

Avaliar, portanto, esse objeto de estudo, não é somente um requisito legal em que obriga o estado a “garantia de padrão de qualidade” (BRASIL, 1988), mas, sobretudo, um assunto emergente, visto a lacuna de um sistema de monitoramento e controle quanto à oferta dos cursos e acompanhamento do desempenho dos estudantes nessa dimensão.

Procedimentos Metodológicos

A pesquisa realizada pode ser classificada como uma pesquisa de abordagem qualitativa, com fins exploratório-descritivos. A pesquisa exploratória está fundamentada no estabelecimento de critérios, métodos e técnicas para oferecer informações sobre o objeto de pesquisa (CERVO; BERVIAN; SILVA, 2007). Já a pesquisa descritiva, conforme apresentado por Perovano (2014), objetiva à identificação, registro e análise das características, fatores ou variáveis que se relacionam com o objeto de estudo.

Como procedimento exploratório, foi utilizada a RSL, que se caracteriza por ser um método que se propõe em realizar uma revisão planejada, para responder a uma questão específica de pesquisa, utilizando técnicas e procedimentos explícitos e organizados, possíveis de ser reproduzidos (DICKSON, 1999). Nesse sentido, foram identificados, selecionados e avaliados criticamente os estudos já elaborados e correlacionados com a questão de pesquisa, visando eliminar tendenciosidades na escolha das fontes analisadas.

A RSL realizada para este artigo segue as orientações da publicação *Cochrane Handbook* (HIGGINS et al., 2019), que determina uma sequência de sete passos: formulação da pergunta; localização e seleção das bases de dados para pesquisa; avaliação crítica dos estudos; coleta de dados; análise e apresentação dos dados; interpretação dos dados e aprimoramento e atualização da revisão.

Após a formulação da pergunta – “quais experiências de aplicação da TRI na avaliação da educação profissional e tecnológica já foram pesquisadas e registradas cientificamente?” – foi escolhida como base de dados a plataforma Scopus (<https://www.scopus.com/>), por ser a maior fonte referencial de literatura técnica e científica, possuir mais de 19.500 títulos de mais de 5.000 editoras. É uma

base internacional e que abrange publicações em várias áreas do conhecimento (multidisciplinar), além de possuir inteligência na indexação de artigos.

Para avaliação crítica e coleta de dados, primeiramente foi inserido o termo "Item Response Theory" na base de dados, pesquisado apenas em títulos, abstracts e keywords, resultando em 4.994 publicações. Além disso, limitou-se a pesquisa para que as publicações possuíssem essa terminologia como palavra-chave, reduzindo para 2.266 documentos encontrados, com o objetivo de que os estudos localizados tivessem esse tema central. Na sequência foram inclusos descritores que remetessem a essa modalidade de educação como: "professional education", "professional training", "professional development", "vocational education" ou "work education", resultando 18 publicações. Das 18 publicações selecionadas, apenas nove estavam disponíveis gratuitamente. Dessas, quatro foram descartadas por focar exclusivamente na educação básica (ensino de matemática e ciência), sendo que não atendem a pergunta de pesquisa, restando apenas cinco artigos. O Quadro 1 resume os passos utilizados na RSL.

Quadro 1 - Filtros da Revisão Sistemática da Literatura

Filtro	Resultados
TITLE-ABS-KEY ("Item Response Theory").	4.994
AND (LIMIT-TO (EXACTKEYWORD, "Item Response Theory").	2.266
AND ("professional education" OR "professional training" OR "professional development" OR "vocational education" OR "work education").	18
Artigos disponíveis gratuitamente.	09
Excluídos artigos cuja temática central era da educação básica.	05

Fonte: Os autores (2018).

Os cinco artigos restantes foram traduzidos e avaliados em profundidade, permitindo a análise descritiva e interpretação dos dados. Nota-se que as experiências encontradas se referem a cursos de nível superior, que por definição da legislação, não se configuram como educação profissional e tecnológica, entretanto margeiam desafios similares e permitem estabelecer paralelismos. Por fim, são apontadas as possibilidades de aprimoramento e atualização da RSL.

Análise dos Resultados

A TRI é um tema recente e amplamente pesquisado por diferentes áreas do conhecimento. Em uma primeira análise, foram encontradas 4.994 publicações na base de dados Scopus. A primeira publicação encontrada foi por Wolfe (1981) e a culminância no ano de 2015, com 532 documentos encontrados (10,7% do total). Na

linha do tempo, verifica-se uma tendência de crescimento no número de estudos sobre o tema, em especial a partir de 2004, sendo que mais de 81,2% das publicações são registradas a partir deste ano. As publicações se dividem em 28 áreas do conhecimento, com destaque para a área de psicologia (47,6% do total). Há também uma pluralidade quanto à origem das publicações, com 86 países diferentes, sendo os Estados Unidos o mais representativo (53,6% do total). Os estudos se classificam em 12 tipos diferentes, sendo que 83,1% são artigos. Além do termo "Teoria de Resposta ao Item, não existe outro descritor predominante, sendo encontrado 154 palavras-chave diferentes.

Para a análise qualitativa, dos cinco artigos resultantes, dois estão entre os mais citados, o que demonstra grande influência dos autores e relevância do estudo (terceiro e quinto artigo apresentado no Quadro 2). A seguir, elenca-se a amostra final dos artigos desta análise, informando-se o título, nome do(s) autor(es) e ano de publicação, listados por ordem cronológica decrescente.

Quadro 2 – Artigos selecionados para análise qualitativa

Nº	Título	Autor(es)	Ano de publicação
01	Modeling the Development of Vocational Competence: a Psychometric Model for Economic Domains (Modelando o desenvolvimento da competência profissional: um modelo psicométrico para domínios econômicos)	Klotz, V. K.; Winther, E.; Festner, D.	2015
02	Modeling the Competencies of Prospective Business and Economics Teachers: Professional Knowledge in Accounting (Modelagem das competências na perspectiva dos professores de administração e economia: conhecimento profissional em contabilidade)	Schnick-Vollmer, K. Berger, S.; Bouley, F.; Fritsch, S.; Schmitz, B.; Seifried, J.; Wuttke, E.	2015
03	Modeling and Measuring the Structure of Professional Vision in Preservice Teachers (Modelagem e medição da estrutura da visão profissional em professores do pré-atendimento)	Seidel, T. Stürmer, K.	2014
04	Measuring nursing competencies in the operating theatre: Instrument development and psychometric analysis using Item Response Theory (Medindo competências de enfermagem na sala de operação: desenvolvimento de instrumento e análises psicométricas utilizando a Teoria de Resposta ao Item)	Nicholson, P. Griffin, P. Gillis, S. Wu, M. Dunning, T.	2013
05	Progress Testing for Postgraduate Medical Education: a Four-Year Experiment of American College of Osteopathic Surgeons Resident Examinations (Teste de progresso em pós-graduação em medicina: a experiência de quatro anos do Colégio Americano em Residência Cirúrgica de Medicina Osteopática)	Shen, L.	2000

Fonte: Os autores (2016).

O primeiro artigo e mais recente (KLOTZ; WINTHER; FESTNER, 2015) relata a experiência no contexto da educação profissional e tecnológica da Alemanha. Embora o artigo tenha sido publicado em 2015, os dados foram coletados em 2013 em quatro escolas de formação profissional, selecionadas de acordo com o maior volume de matrículas no país. A experiência se baseou em criar uma escala que conseguisse avaliar o nível de competência dos estudantes de economia em dois distintos momentos: iniciantes – com até 0,5 anos de formação inicial profissional e avançados – de 1,3 a 2,8 anos de formação inicial profissional.

Como base conceitual, apropriou-se dos axiomas de Dreyfus e Dreyfus (1980) que estabelece dois grupos de competência: “*domain-linked competence*” (domínio ligado a competência) não requer conhecimentos específicos para realizar a tarefa e “*domain-specific competence*” (domínio específico de competência) que requer conhecimentos específicos para realizar a tarefa. Os autores elucidam que, no primeiro grupo, por exemplo, estaria a competência em calcular uma taxa de câmbio, que, embora seja uma atividade prevista no currículo do curso, pode ser concluída com habilidade numérica geral e não específica. Já no segundo grupo estaria a correta aplicação da teoria do contrato econômico em uma determinada situação, que requer conhecimentos específicos em economia para a resolução.

A cada um dos grupos de estudantes (iniciantes e avançados) foram aplicadas um conjunto de tarefas (itens), sendo elaborados 46 itens diferentes, no qual 16 eram em comum aos dois grupos. Do total, 28 itens objetivavam avaliar domínios ligados à competência e 18 domínios específicos de competência. Noutra classificação: 15 itens tinham como base avaliação conceitual, a qual busca averiguar se o estudante possui conhecimento em um determinado assunto; 21 eram processuais, visando identificar a habilidade dos estudantes em aplicar conhecimentos em uma determinada situação; e 10 eram interpretativos, sendo o nível mais elevado e que exige a capacidade do estudante em tomar decisões mais estratégicas a partir da interpretação.

Os itens foram elaborados em formato de situação-problema e estudo de caso, onde o respondente precisava ler e interpretar. A avaliação, portanto, era teórica e não prática. As respostas eram abertas e avaliadas por “peritos” (avaliadores) que classificavam a resposta na escala de três categorias: incorreta, parcialmente correta e correta.

Quanto aos resultados, os autores observaram que o instrumento teve boa aderência dos estudantes, pois foi utilizado como preparação para o exame final. Por

meio da TRI, os itens do instrumento foram classificados em quatro níveis de progressão de aprendizagem: nível 1 – numérico e literária, nível 2 – domínio específico de conhecimento, nível 3 – domínio específico de procedimentos e nível 4 – domínio específico de interpretação. Nos três primeiros níveis ficou evidente a progressão de aprendizagem entre os grupos dos iniciantes e avançados, demonstrando que o tempo de formação profissional contribui diretamente para a aquisição dessas competências.

Os pesquisadores apontaram ainda algumas limitações da pesquisa como a sua extensão, podendo ser reproduzida e aplicada apenas naquele momento e naquele local, não podendo ser replicada para outros grupos e países. Outra questão apontada também foi a seleção da amostra, sendo que não foi probabilística (não aleatória), sem permitir a estimação de parâmetros de grau de confiança. Todavia, ressaltam que este é um “novo campo sistemático de pesquisa” que carece de avanço na compreensão da aprendizagem profissional, alertando quanto a necessidades de investigações mais profundas.

No segundo artigo (SCHNICK-VOLLMER et al., 2015), embora o foco não seja proficiência de estudantes, a experiência é válida, pois se busca avaliar a competência dos professores. Os autores sustentam o conhecimento profissional do docente como fator crucial para o ensino de sucesso, uma vez que impacta diretamente no desempenho dos estudantes.

Amparados teoricamente por Shulman (1986), dividem o conhecimento profissional em dois grupos: conhecimento de conteúdo (CK) e conhecimento pedagógico de conteúdo (PCK). No âmbito da pesquisa, cujo foco é o estudo de contabilidade como ferramenta para as áreas de administração e economia, o CK é entendido como os conhecimentos técnicos e necessários para estruturação de instruções em sala de aula. Dentro da contabilidade teríamos três principais áreas: (1) o propósito, a relevância e a base jurídica em contabilidade; (2) dupla entrada da contabilidade e (3) compra e venda. Já o PCK é percebido como tornar o conteúdo acessível aos alunos e conhecimento do pensamento do aluno (cognição). Aplicando na contabilidade temos: (1) cognição do estudante e erros típicos, (2) conhecimento das tarefas como ferramenta de ensino; e (3) conhecimento das múltiplas representações e explicações.

O instrumento foi construindo com 49 itens no total, sendo que cada participante respondia apenas 28. O teste foi realizado no formato *paper-pencil-test* (papel e caneta). Os itens podiam ser de três modos: abertos, múltipla escolha ou múltipla

escolha complexa. A escala de resposta foi dicotômica: certo ou errado. A amostra contemplou 1.158 futuros professores de 24 universidades da Alemanha com programas de formação de docentes em economia e administração. Os participantes receberam €\$20,00 para responderem a prova.

Os principais resultados apontam que há uma alta correlação entre os dois tipos de conhecimentos CK e PCK, sendo $R = 0,92$ (muito alta). A confiabilidade apresentada pelo modelo Rasch² é de 0,64 para ambos, sendo considerado satisfatório pelos autores para validação da utilização do modelo. Também foi constatada divergência entre o CK de homens e mulheres, sendo que os participantes do sexo masculino obtiveram desempenho significativamente melhor do que as mulheres. Além disso, uma limitação importante apresentada foi que os itens de PCK foram construídos com situações hipotéticas e não reais, o que poderia distorcer parcialmente a escala elaborada.

Igualmente, no terceiro artigo (SEIDEL; STÜRMER, 2014), o foco não é a avaliação do desempenho dos estudantes, e sim a forma como os professores criam oportunidades de aprendizagem, uma vez que esta influência fortemente o aprendizado do aluno. A pesquisa na formação de professores, conforme relatado no artigo, é um campo novo, pois, até então, as medidas eram puramente subjetivas e baseadas na titulação e em cursos realizados pelos docentes. Desse modo, surge a necessidade por um modelo de teste empírico, sendo este destacado como objeto de estudo deste artigo analisado.

O foco, portanto, é avaliar a "visão profissional" dos futuros docentes, sendo esta entendida como a capacidade de perceber e interpretar características relevantes de situações ocorridas em sala de aula. Os autores acrescentam que os professores em início de carreira encontram-se, muitas vezes, incapazes de lidar com a complexidade do ambiente escolar e apresentam dificuldade na transposição do que aprenderam na universidade para o contexto real de ensino-aprendizagem.

O constructo avaliado foi dividido em três pilares: descrever, explicar e prever situações em sala de aula. A descrição é a capacidade de diferenciar claramente os aspectos relevantes de um ensinamento notado e componente de aprendizagem (ou seja, o objetivo) sem fazer quaisquer julgamentos. A explicação é capacidade de usar o que se sabe para raciocinar sobre uma situação, esclarecer o ensino e as metas de aprendizagem, ativando o conhecimento prévio dos alunos. Já a previsão é a

² Trata-se de um modelo de TRI, também denominado de modelo logístico de um parâmetro (ML1P), codificado pela primeira vez pelo matemático Georg Rasch. Expressa a probabilidade de se obter uma resposta assertiva em função da diferença entre o sujeito e a dificuldade do item, sendo este o parâmetro exclusivo de análise.

capacidade de prever as consequências de eventos observados em termos de aprendizagem dos alunos, aplicados à prática em sala de aula. Os autores abordam que nenhuma pesquisa foi realizada até o momento para testar empiricamente a inter-relação desses três aspectos.

O instrumento de coletas de dados foi elaborado com cliques de vídeos que representassem situações reais e relevantes de sala de aula, utilizando os critérios de autenticidade e promoção da ativação do conhecimento dos professores como parâmetros de seleção. Doze cliques foram elencados e classificados de acordo com dois ou três dos parâmetros de "visão profissional", validados por 40 peritos (professores) que compuseram a banca de seleção. Para cada vídeo foram desenvolvidos 36 itens com respostas na escala *Likert* de quatro pontos: discordo, discordo parcialmente, concordo parcialmente e concordo. Os itens eram baseados em percepções dos vídeos, sendo que as respostas dos avaliados foram comparadas com a dos peritos, que configurou o referencial de critérios. Cada indivíduo verificou seis vídeos, com 90 minutos de duração total, sendo o formato de aplicação online.

Os resultados obtidos foram coletados em três estudos com três amostras distintas, porém com público-alvo semelhante. Os dois primeiros grupos foram com estudantes no quinto semestre de formação e o último com estudantes em fase de conclusão da graduação. Os cliques foram avaliados como autênticos e relevantes pelos participantes, representando bem as três dimensões. O modelo dicotômico de Rasch mostrou maior confiabilidade do que o modelo parcial de créditos, e a perspectiva tridimensional apresentou resultados mais significativos que a unidimensional e bidimensional. Foram construídas escalas para cada uma das dimensões e uma escala integradora com as três abordagens. Além disso, no terceiro grupo foram aplicadas duas vezes o mesmo instrumento, com intervalos de dias diferentes em pequenos grupos. Não foram constatadas diferenças sistemáticas entre os dois pontos de medição ao longo do tempo e os autores não apontaram limitações para a pesquisa.

O quarto artigo (NICHOLSON et al., 2013) está associado à área de enfermagem. Os autores buscaram elaborar um instrumento para avaliar o desempenho clínico de enfermeiros em salas de operação com base em evidências circundantes, sendo este uma ferramenta para atestar se os enfermeiros estavam aptos em gerir e prestar cuidados de qualidade para os pacientes durante procedimentos cirúrgicos. Segundo eles, o uso de testes de múltipla escolha e testes de habilidades clínicas já foram

aplicadas nas escolas de enfermagem da Austrália, porém há uma necessidade e um movimento em direção a sistemas de avaliação baseado em competências. Isso se justifica porque um instrumento deste tipo consegue levar em consideração diferentes variáveis presentes no ambiente clínico que poderiam influenciar o resultado da avaliação.

Para isso, os autores propuseram um instrumento baseado em dois métodos: analítico e holístico. No primeiro buscava-se examinar aspectos específicos do desempenho em relação a um conjunto de critérios, isto é, julgava-se a performance de modo geral, mas para cada aspecto era atribuído uma pontuação separadamente. Na visão dos autores, esse método permite níveis de informações mais detalhados e possibilita um *feedback* mais pontual ao avaliado. Já o holístico verificava a qualidade global do desempenho, não marcando os componentes separadamente. Apesar de prestar poucas informações, entendeu-se válido verificar se o enfermeiro conseguiria alcançar o resultado global da atividade esperada.

Para ambos os métodos, o critério de avaliação adotado era a rubrica, podendo ser entendida nas palavras dos autores como “a ferramenta de pontuação que estabelece as especificações e expectativas para uma missão”. Isto é, a qualidade do desempenho requerida que será verificada por meio de evidências. Para isso foram elaborados descritores comportamentais para cada item e codificados em 0, 1, 2, 3 ou 4, sendo o 4 o nível mais avançado e 0 atribuído quando o desempenho não pudera ser observado. Também foram aplicados itens dicotômicos, categorizados em “competente” ou “ainda não competente”.

O instrumento foi validado por 32 educadores em enfermagem que aplicaram em três enfermeiros de diferentes níveis de experiência. Foi solicitado *feedback* da ferramenta e preenchimento de questionário fundo para levantamento de dados demográficos. Os resultados evidenciaram que a maioria dos itens apresentou confiabilidade utilizando o modelo de Rasch, com limite fixado entre 0,7 e 1,3. Os autores concluem que a evidência é necessária para reduzir a subjetividade, usando uma abordagem de lista de verificação (padrões esperados de comportamento) e que a análise por meio do modelo de Rasch é um bom caminho para a tomada de decisão em competência.

O quinto e último artigo (SHEN, 2000) traz a experiência do Colégio Americano de Cirurgiões Osteopáticos (ACOS) e do Conselho Nacional de Examinadores Médicos Osteopatas (NBOME), que, desde 1994, aplicam testes de progressos no programa de

pós-graduação em medicina. Estes possuem como objetivo fornecer informações aos coordenadores sobre o diagnóstico de aprendizagem dos residentes, sem possuir caráter de aprovação/reprovação.

O método utilizado é a aplicação de instrumentos por escrito, com periodicidade anual e itens no formato dicotômico (certo ou errado). É conduzido nos quatro anos de formação, somando um total de 300 itens por teste, distribuídos em nove temas clínicos e que contemplem todo o currículo (independentemente do ano que os residentes estejam cursando). Outras características identificadas foram: 50 itens formam intersecção com anos anteriores ou posteriores; pessoas e itens foram posicionados na mesma escala interpretativa; e utilização do modelo Rasch, apresentando confiabilidade razoável segundo o autor.

Quanto aos principais resultados, de modo geral, os exames foram classificados como difíceis, uma vez que cada teste de progresso cobriu todo o currículo e os avaliados em níveis de formação mais baixos estavam claramente em desvantagem. O crescimento entre os anos é contínuo, porém não linear. No nível individual, os testes ajudam os coordenadores da pós-graduação em avaliar se os residentes estão cumprindo as normas e verificar o quão rápido estão se aproximando dos padrões necessários. No nível geral, serve como uma ferramenta para avaliar a eficácia de cada programa. As limitações apresentadas foram: necessidade de mais análises psicométricas junto à amostra, em especial a avaliação da validade preditiva, além da demanda pela construção de um banco de itens para sistematização do processo.

Considerações finais

Visto que o objetivo dessa pesquisa era identificar quais experiências de aplicação da TRI na avaliação da educação profissional e tecnológica já foram pesquisadas e registradas cientificamente, os artigos analisados contribuem para o início de uma discussão que demanda aprofundamento, uma vez que a metodologia utilizada apontou cinco estudos, o que representa apenas 0,1% das publicações na área (5 entre 4.994), ainda que outros artigos estivessem indisponíveis para consulta gratuitamente (quatro). Percebe-se também que todos são estrangeiros, predominantemente de origem alemã, sem destaque para produção científica nacional, o que também indica a iminência do assunto. De modo geral, a

pesquisa buscou um caminho para dar visibilidade a uma área pouco explorada, sem necessariamente pretender a exaustão sobre a temática.

Noutra frente, as aplicações encontradas são predominantemente da avaliação da dimensão cognitiva dos estudantes ou professores da educação profissional e tecnológica e pouco se observa a dimensão psicomotora, que poderia ser observada por testes práticos (apenas citados no quarto artigo), não apresentando detalhes específicos sobre a formulação dos itens nesse sentido. Em alguns artigos até ficou evidente a realização dos testes por meio de situações problemas, embora ainda não se observasse aplicação de avaliação prática, buscavam simular no papel o modo como os estudantes deveriam se comportar em uma determinada circunstância. No segundo artigo, por assim citar, os próprios autores alegam a deficiência de trabalhar com situações hipotéticas e não reais, reforçando a necessidade de avaliar na prática se o estudante consegue desenvolver uma determinada competência ou não.

Outro ponto importante é a necessidade de elaboração de itens específicos por especialistas da área que se pretende avaliar. No exemplo do quinto artigo, os itens foram coletados da literatura e não demonstraram bons resultados. Na perspectiva dessa modalidade de educação, é válido destacar que não há itens prontos e que para uma avaliação dessa magnitude estes deverão ser inéditos e construídos por docentes com conhecimento técnico em cada área tecnológica. Em contrapartida, no quarto artigo temos um direcionador chamado rubrica, demandando a avaliação por meio de evidências observáveis e percebido pelos pesquisadores como uma alternativa de estratégia para esse tipo de avaliação.

Os testes de progressos, apresentado no quinto artigo, também contribui com a ideia de avaliação global do currículo, o que faz sentido nessa vertente, uma vez que o estudante formado poderá ser demandado em qualquer uma das competências relacionadas ao seu perfil profissional.

Por fim, cabe destacar sobre a incipiência do tema e dos muitos desafios que estão empregados nessa temática. Aplicar a TRI na educação profissional e tecnológica é viável e já realizado como demonstrado em quatro dos cinco artigos analisados, contudo os testes utilizados focam no domínio cognitivo da aprendizagem e estavam predominantemente restritos ao modelo papel e caneta. Avaliar estudante utilizando um instrumento prático será uma nova vertente e

contribuição, não só para esse segmento educacional, mas para todo o sistema educacional nacional e internacional.

Esta pesquisa poderá ser continuada e aprimorada utilizando outras bases de dados indexadas e cruzando as informações do detalhamento apresentado nesse artigo. Recomenda-se a utilização de bases internacionais em função da ausência de um sistema de avaliação dessa natureza no Brasil, no qual poderá ter maior chance de sucesso. Periodicamente, podem ser realizadas novas buscas de dados, com os descritores já apresentados, para identificar se novos estudos foram incluídos, principalmente porque os artigos mais relevantes foram importados nos últimos três anos.

Referencias

AFONSO, I. Brasil vai avaliar cursos e alunos de educação profissional de todo o país. *Portal da Indústria*. [s. l.], 2015. Disponível em:

<https://noticias.portaldaindustria.com.br/noticias/educacao/brasil-vai-avaliar-cursos-e-alunos-de-educacao-profissional-de-todo-o-pais/>. Acesso em: 3 maio 2016.

ANDRADE, D. F.; VALLE, R. C. Introdução à teoria da resposta ao item: conceitos e aplicações. *Estudos em Avaliação Educacional*, São Paulo, n. 18, p. 13-32, jul./dez. 1998.

ANDRADE, D. F.; TAVARES, H. R.; VALLE, R. C. *Teoria da resposta ao item: conceitos e aplicações*. São Paulo: ABE - Associação Brasileira de Estatística, 2000.

BRASIL. [Constituição (1937)]. *Constituição da República Federativa do Brasil de 1937*. Brasília, DF: Casa Civil, 1937.

BRASIL. [Constituição (1988)]. *Constituição da República Federativa do Brasil de 1988*. Brasília, DF: Senado Federal: Centro Gráfico, 1988. 292 p.

BRASIL. Ministério da Educação. Lei nº 9.394, de 20 de dezembro de 1996.

Estabelece as diretrizes e bases da educação nacional. *Diário Oficial da União*, Brasília, DF, 23 dez. 1996.

BRASIL. Decreto nº 5.154, de 23 de julho de 2004. Regulamenta o § 2º do art. 36 e os arts. 39 a 41 da Lei nº 9.394, de 20 de dezembro de 1996, que estabelece as diretrizes e bases da educação nacional, e dá outras providências. *Diário Oficial da União*, Brasília, DF, 23 jul. 2004.

BRASIL. Decreto nº 7.566, de 23 de setembro de 1909. Crêa nas capitães dos Estados da República Escolas de Aprendizes Artífices, para o ensino profissional primário e gratuito. *Diário Oficial da União*, Brasília, DF, 26 set. 1909.

BRASIL. Lei nº 11.700, de 13 de junho de 2008. Acrescenta inciso X ao caput do art. 4º da Lei nº 9.394, de 20 de dezembro de 1996, para assegurar vaga na escola pública de educação infantil ou de ensino fundamental mais próxima de sua residência a toda criança a partir dos 4 (quatro) anos de idade. *Diário Oficial da União*, Brasília, DF, 13 jun. 2008.

BRASIL. Lei nº 12.513, de 26 de outubro de 2011. Institui o Programa Nacional de Acesso ao Ensino Técnico e Emprego (Pronatec); altera as Leis nº 7.998, de 11 de janeiro de 1990, que regula o Programa do Seguro-Desemprego, o Abono Salarial e institui o Fundo de Amparo ao Trabalhador (FAT), nº 8.212, de 24 de julho de 1991, que dispõe sobre a organização da Seguridade Social e institui Plano de Custeio, nº 10.260, de 12 de julho de 2001, que dispõe sobre o Fundo de Financiamento ao Estudante do Ensino Superior, e nº 11.129, de 30 de junho de 2005, que institui o Programa Nacional de Inclusão de Jovens (ProJovem); e dá outras providências. *Diário Oficial da União*, Brasília, DF, 26 out. 2011.

- BRASIL. Decreto nº 7.690, de 2 de março de 2012. Aprova a Estrutura Regimental e o Quadro Demonstrativo dos Cargos em Comissão e das Funções Gratificadas do Ministério da Educação. *Diário Oficial da União*, Brasília, DF, 2 mar. 2012.
- CERVO, A. L.; BERVIAN, P. A.; SILVA, R. *Metodologia científica*. 6 ed. São Paulo: Pearson Prentice Hall, 2007.
- CGEE. *Mapa da educação profissional e tecnológica: experiências internacionais e dinâmicas regionais brasileiras*. Brasília, DF: Centro de Gestão e Estudos Estratégicos, 2015.
- DEMARS, C. *Item response theory*. New York: Oxford University, 2010. (Understanding Statistics).
- DICKSON, R. Systematic reviews. In: Hamer S.; Collinson G. *Achieving evidence-based practice: a handbook for practitioners*. London: Baillière Tindall, 1999. p. 41-60.
- DREYFUS, S. E., DREYFUS, H. L. *A five-stage model of the mental activities involved in directed skill acquisition*. Berkley: University of California, Berkeley. 1980.
- HIGGINS, J. P. T. et al. (ed.). *Cochrane handbook for systematic reviews of interventions*. 2. ed. Chichester, UK: John Wiley & Sons, 2019.
- KIRSCHNER, T. C. *Modernização tecnológica e formação técnico-profissional no Brasil: impasses e desafios*. Rio de Janeiro: IPEA, 1993.
- KLOTZ, V. K.; WINTHER, E.; FESTNER, D. Modeling the development of vocational competence: a psychometric model for economic domains. *Vocations and Learning*, Springer Netherlands, v. 8, n. 3, p. 247–268, out. 2015.
- LAWLEY, D.N. *On problems connected with item selection and test construction*. *Proceedings Royal Society Edinburgh, Series A*. 1943; 61 (2):273-87.
- LORD, F. M. *A theory of test scores*. New York: Psychometric Society, 1952. (Psychometric Monograph, n. 7).
- MANFREDI, S. M. *Educação profissional no Brasil*. São Paulo: Cortez, 2002.
- NICHOLSON, P. et al. Measuring nursing competencies in the operating theatre: instrument development and psychometric analysis using item response theory. *Nurse Education Today*, Edinburgh, UK, v. 33, n. 9, p. 1088–1093, set. 2013.
- PASQUALI, L. *Psicometria: teoria e aplicações*. Brasília: Editora Universidade de Brasília, 1997.
- PEROVANO, D. G. *Manual de metodologia científica: para a segurança pública e defesa social*. Curitiba: Juruá Editora, 2014.
- SCHNICK-VOLLMER, K. et al. Modeling the competencies of prospective business and economics teachers: professional knowledge in accounting. *Zeitschrift für Psychologie*, [s. l.], v. 223, n. 1, p. 24-30, mar. 2015.

SEIDEL, T.; STÜRMER, K. Modeling and measuring the structure of professional vision in preservice teachers. *American Educational Research Journal*, Washington, v. 51, n. 4, p. 739–771, ago. 2014.

SHEN, L. Progress testing for postgraduate medical education: a four-year experiment of american college of osteopathic surgeons resident examinations. *Advances in Health Sciences Education*, Holanda, v. 5, n. 2, p. 117–129, maio 2000.

SHULMAN, L. S. Those who understand: knowledge growth in teaching. *Educational Researcher*, Thousand Oaks, Ca, v. 15, n. 2, p. 4-14, fev. 1986.

WOLFE, J. H. Optimal item difficulty for the three-parameter normal ogive response model. *Psychometrika*, New York, US, v. 46, n. 4, p. 461-464, dez. 1981.