

QUESTIONÁRIO CONCEPÇÕES DE AVALIAÇÃO DE PROFESSORES: ADAPTAÇÃO E VALIDAÇÃO PARA A REALIDADE BRASILEIRA *

Karine Votikoske RONCETE[√]
Daniel Abud Seabra MATOS^{√√}

RESUMO

Os objetivos deste trabalho foram: realizar uma revisão da literatura sobre o questionário Teachers' Conceptions of Assessment (TCoA-III) e adaptar e coletar fontes de evidência de validade para a realidade brasileira. Como referencial teórico, a literatura aponta quatro grandes grupos de concepções de avaliação de professores: 1) melhora; 2) responsabilização da escola; 3) responsabilização do aluno e 4) irrelevância. Foram encontrados 25 trabalhos internacionais que utilizaram o TCoA-III. Com relação ao estudo empírico, participaram professores da educação básica. Realizou-se uma Análise Fatorial Exploratória do questionário e os resultados sugerem uma estrutura fatorial diferente para o TCoA brasileiro. Efetuou-se diversas análises excluindo 28 itens que se mostraram problemáticos. O modelo final apresentou um bom ajuste (CFI= .960; RMSEA= .067; SRMR = .045) e quatro fatores: 1) melhora; 2) melhora-ensino; 3) irrelevância-ignorada e 4) responsabilização do aluno. Assim, esse resultado sugere uma estrutura fatorial diferente para o TCoA brasileiro, que não possui a concepção "responsabilização da escola". Acredita-se que a utilização do TCoA em pesquisas futuras é uma alternativa viável para pesquisadores interessados na temática das concepções de avaliação de professores em diversos sentidos.

Palavras-chave: Concepções. Avaliação. Professor. Educação. Análise Fatorial.

* Artigo recebido em 20/10/2023 e aprovado em 25/11/2023.

[√]Doutora em Educação pela Universidade Federal de Ouro Preto (UFOP). Coordenadora de Pesquisa e membro do Comitê Científico do Equidade.info. E-mail: karineroncete@gmail.com.

^{√√}Doutor em Educação pela Universidade Federal de Minas Gerais (UFMG). Professor do Programa de Pós-Graduação em Educação, da Universidade Federal de Ouro Preto (UFOP). E-mail: danielmatos@ufop.edu.br.

1 INTRODUÇÃO

Os estudos sobre avaliação educacional têm obtido um incremento significativo de interesse nos últimos anos, mas os que focam nas concepções de professores ainda são escassos no Brasil, apesar de serem comuns em outros países (Ribeiro *et al.*, 2018). Pesquisas apontam que as concepções dos professores sobre vários aspectos do processo educacional, como ensino, aprendizagem, avaliação e currículo, influenciam fortemente o modo como eles ensinam (SANTOS, 2020). As concepções de avaliação dos professores são moldadas de acordo com a forma que conceituam o ensino e a aprendizagem. As concepções que os docentes têm sobre as avaliações são um produto de suas experiências educacionais como estudantes, sugerindo que concepções semelhantes podem ser encontradas entre os docentes e os discentes. A literatura também indica associações entre as concepções de avaliação dos professores e o desempenho acadêmico dos alunos (Brown, 2008).

Dessa forma, todos os fatores supracitados justificam esta pesquisa, que busca compreender as concepções de avaliação de professores. De maneira mais específica, o objetivo deste trabalho é apresentar a adaptação e a validação do questionário *Teachers' Conceptions of Assessment (TCoA-III)* para o contexto educacional brasileiro.

2 CONCEPÇÕES DE AVALIAÇÃO DE PROFESSORES

O senso comum, por vezes, concebe a avaliação como sinônimo de medida e hierarquização. Porém, dentro de uma perspectiva pedagógica, a dimensão da avaliação torna-se mais ampla. O ato de avaliar não pode resumir-se, somente, a atribuir notas, mas realizar uma série de objetivos que se traduzem em mudanças de comportamento dos alunos (LAMY; GALIETA, 2019). Além disso, ao se pensar a avaliação como um processo que ajuda no ensino e na aprendizagem, as funções mais usuais presentes na literatura são: diagnóstica, formativa e somativa. Miquelante *et al.* (2019) apontam que a avaliação diagnóstica é aquela efetuada antes do início de uma instrução, podendo ser o ponto de partida do ensino (p. 268). Já a formativa, tem uma característica processual, como foco no decorrer do ensino e da aprendizagem (p.269). Por fim, a somativa, ocorre ao final do ensino, ou seja, ela julga

o valor de um programa após ter sido terminado (p.270). Por isso, compreender as funções e os propósitos da avaliação é imprescindível para o estudo das concepções de avaliação dos professores, pois esses conceitos estão relacionados. Lamy e Galieta (2019) enfatizam, ainda, que a formação inicial de professores deve ter como foco reflexões sobre as finalidades da avaliação. Na literatura, identificam-se quatro grandes grupos de concepções de avaliação dos professores (Brown, 2008), que se vinculam com diversas funções e propósitos da avaliação:

A avaliação melhora o ensino e a aprendizagem (**melhora**). A maior premissa dessa concepção é que a avaliação melhora o aprendizado dos alunos e a qualidade do ensino. Às vezes, também é referida como avaliação para o aprendizado ou avaliação formativa. Silva *et al.* (2020) definem que a avaliação formativa (de melhora) deve ajudar o discente a desenvolver suas capacidades cognitivas, por meio de um acompanhamento progressivo. Dessa maneira, essa categoria é baseada na suposição de que o dever dos professores é melhorar a aprendizagem de seus alunos e que a avaliação é um método utilizado para a obtenção de informações que contribuam para esse processo. Logo, os educadores seriam capazes de usar a avaliação formativa para incentivar a aprendizagem (a avaliação é legítima se resultar em ação que consiga mudar o que os alunos sabem ou podem fazer). Na concepção da melhora, são avaliados os conhecimentos com o objetivo de gerar informações para que a melhoria no desempenho dos alunos possa ser alcançada.

A avaliação torna os professores e as escolas responsáveis pela sua eficácia (**responsabilização da escola - accountability**). O termo *accountability* tem sido traduzido como responsabilização, mas ele abrange uma complexidade em sua tradução. Na educação, pode ser definido como uma política de tornar escolas e professores responsáveis pelo progresso acadêmico dos alunos, ligando esse progresso com recursos para salários, manutenção das escolas, entre outros. Desse modo, essa concepção focaliza no uso dos resultados da avaliação para demonstrar publicamente que professores e escolas estão fazendo um bom trabalho, o que impõe consequências caso eles não atinjam os padrões estabelecidos (Araujo; Leite; Passone, 2018). Duas lógicas estão postas aqui: demonstrar, de forma pública, que as escolas e professores promovem um ensino de qualidade e melhorar a qualidade do ensino. As consequências da responsabilização podem ser positivas ou negativas. Além disso, podem tanto possuir um alto impacto quanto um baixo impacto (Araujo *et*

al., 2018). Exemplos: o professor recebe bonificação por uma melhora em resultados de avaliação dos estudantes (consequência positiva e de alto impacto, uma vez que resulta em uma remuneração diferente para o docente); ou uma escola é publicamente exposta por ter resultados de avaliações ruins quando comparadas com outras (consequência negativa e de alto impacto, podendo resultar no recebimento de menos recursos, em críticas dos pais ou na diminuição do número de matrículas).

A avaliação torna os alunos responsáveis pelo seu aprendizado (**responsabilização do aluno**). Essa concepção destaca a noção de que a avaliação torna o aluno individualmente responsável pelo seu aprendizado, mediante notas ou certificados; ou informações de resultados para os pais, futuros empregadores e outros educadores. Essas avaliações podem ter consequências importantes para os estudantes como um diploma, a aprovação em uma disciplina ou um concurso, uma bolsa, entre outras.

A avaliação é irrelevante e os professores respondem negativamente a ela (**irrelevante**). A premissa dessa concepção é que a avaliação não tem um lugar legítimo dentro do ensino e da aprendizagem, baseando-se na visão de que os processos de avaliação são inadequados, imprecisos ou irrelevantes para a habilidade do professor de melhorar o aprendizado do aluno. Assim, não haveria a necessidade de realizar qualquer tipo de avaliação além de processos intuitivos que ocorrem automaticamente enquanto os docentes interagem com os discentes. Isso seria garantido pelo conhecimento dos professores sobre os alunos baseado em um longo relacionamento com eles e em uma profunda compreensão do currículo. Os processos avaliativos podem ser considerados irrelevantes, também, por causa de seus efeitos prejudiciais sobre a autonomia do educador e seu poder de “distração” do verdadeiro propósito de ensinar. Tratar a avaliação como irrelevante, portanto, é comumente ligado às seguintes alegações: a avaliação equivale a testes e há uma pressuposição de que os testes são ruins para a educação e a avaliação responsabiliza os professores, as escolas e os alunos, o que é considerado ruim para a qualidade da educação. Alguns docentes consideram a avaliação irrelevante em função das consequências negativas nos alunos (BROWN, 2008).

Os professores tendem a ter uma orientação direcionada a uma concepção ou uma combinação de várias delas. Dessa maneira, um mesmo docente pode relatar que a avaliação serve para melhorar o aprendizado e que ela é injusta, devendo ser

ignorada (concepções de avaliação que parecem ser opostas e conflitantes). A natureza extremamente complexa do fenômeno educacional experienciado por alunos e professores parece contribuir para essa diversidade de concepções. Outra explicação para essa diversidade de concepções em um mesmo sujeito, reside no fato de existirem múltiplas funções da avaliação (as concepções estão ligadas a essas funções).

2.1 O TEACHERS CONCEPTIONS OF ASSESSMENT - TCOA

O TCoA foi desenvolvido em 2004 por Brown, na Nova Zelândia, e investiga as concepções de professores sobre a avaliação. O primeiro questionário desenvolvido foi o *Teachers Conceptions of Assessment* (TCoA-I), constituído de 115 itens e respondido por 84 professores. Os itens do TCoA-I estão agrupados em três grandes concepções: responsabilização (das escolas e dos alunos), melhora e irrelevância. A Análise Fatorial reteve 65 itens e levou a 10 fatores razoavelmente consistentes: 2 fatores de responsabilização, 5 fatores de melhora e 3 fatores de irrelevância. Esse estudo não foi capaz de correlacionar vários fatores (devido ao tamanho pequeno da amostra), mas conseguiu demonstrar que as principais concepções descritas pela literatura poderiam ser facilmente identificadas no pensamento dos professores. Segundo Brown (2008), quase metade dos itens do TCoA-I não se adaptaram ao modelo. Assim, mais itens foram elaborados com base em uma análise mais aprofundada da literatura sobre as concepções de avaliação de professores (Brown, 2008).

A segunda versão do questionário (TCoA-II) teve 105 itens e foi aplicada em 188 participantes. Os itens desse foram agrupados em três concepções: responsabilização (das escolas e dos alunos), melhora e irrelevância. Baseado em 46 itens retidos pela Análise Fatorial, três fatores foram obtidos: a concepção de responsabilização tinha 12 itens; a concepção de melhora tinha 24 itens; e a concepção de irrelevância tinha 10 itens (Brown, 2008). Para a terceira versão (TCoA-III), adicionou-se 19 itens. Além disso, a concepção de responsabilização foi dividida em duas, ficando assim quatro grupos de concepções de avaliação de professores: melhora, responsabilização da escola, responsabilização do aluno e irrelevância. Foi analisada uma amostra de 525 docentes. Através da Análise Fatorial, foram removidos 15 itens. Os itens foram retirados do questionário por diversos motivos,

como: variância de erro negativa ou por apresentarem cargas fatoriais baixas nos fatores. A versão final ficou com 50 itens (BROWN, 2008).

O TCoA-III é a versão mais atual do instrumento. O formato de respostas é uma escala *Likert* de seis pontos: discordo fortemente, discordo na maior parte, concordo ligeiramente, concordo moderadamente, concordo na maior parte, concordo fortemente. O TCoA III apresenta uma estrutura fatorial complexa, com fatores de primeira e segunda ordem (FIGURA 1).

Figura 1 - Estrutura fatorial original do Teachers Conceptions of Assessment - TCoA III



Fonte: Adaptado de BROWN, 2004.

O fator de segunda ordem Melhora (*Improvement*) possui quatro fatores de primeira ordem: descrição (*describe* – a avaliação descreve habilidades do aluno, conhecimento e pensamento); aprendizado do aluno (*student learning* – a avaliação melhora o aprendizado do aluno); válida (*valid* – a informação da avaliação é válida devido à sua confiabilidade) e ensino (*teaching* – a avaliação melhora o ensino) (Brown, 2004). O fator de segunda ordem Irrelevância (*Irrelevance*) possui três fatores de primeira ordem: ruim (*bad* – a avaliação é ruim para o ensino); ignorada (*ignore* – os professores ignoram a avaliação) e imprecisa (*inaccurate* – a avaliação é imprecisa) (Brown, 2004). Os fatores Responsabilização da escola e Responsabilização do aluno são apenas fatores de primeira ordem (BROWN, 2004).

Existe, também, uma versão abreviada do questionário. O TCoA-III (Brown, 2008) possui 27 itens e avalia as mesmas quatro grandes concepções. O TCoA III ainda averigua as definições de avaliação, que são representadas por uma lista de práticas avaliativas. Essa parte não foi analisada no presente trabalho.

2.2 ESTUDOS ANTERIORES COM O TEACHERS CONCEPTIONS OF ASSESSMENT - TCOA

O *Teachers Conceptions of Assessment* foi utilizado em mais de 25 estudos nos seguintes países: Nova Zelândia, Hong Kong, Grécia, Espanha, Egito, China e Estados Unidos. Alguns desses estudos serão apresentados abaixo. Nos casos em que existe mais de uma pesquisa realizada no país, apresentamos um trabalho como exemplo. No total, oito trabalhos foram selecionados.

Brown (2004), por meio do questionário *Teachers Conceptions of Assessment* – TCoAIII, estudou as concepções de avaliação dos professores da Nova Zelândia e a relação dessas concepções com o ensino, o currículo e a eficácia do professor. O questionário utilizado possui 50 itens. Participaram 81 professores. O modelo de Análise Fatorial mostrou um ajuste bom de um modelo hierárquico multidimensional para os dados. Nos resultados, percebeu-se que os professores concordavam moderadamente com as concepções de melhora e de responsabilização da escola e discordavam que a avaliação é irrelevante. As concepções de melhora, responsabilização da escola e responsabilização do aluno foram correlacionadas positivamente. Já a concepção de irrelevância foi inversamente associada com a

concepção de melhora e não associada com a concepção de responsabilização da escola¹.

Brown *et al.* (2009) pesquisaram as concepções e as práticas de avaliação de 374 docentes de Hong Kong por meio do TCoA-III A, com 27 itens. Os docentes de Hong Kong concordam mais fortemente com a concepção de melhora e rejeitam a concepção de irrelevância. A correlação entre a concepção de responsabilização dos estudantes e a de melhora foi alta ($r = 0,91$). Isso sugere que os docentes de Hong Kong veem a responsabilização do aluno associada à melhora da aprendizagem.

Brown *et al.* (2011), para desenvolver estudos comparativos entre Hong Kong e a província chinesa de Guangdong, desenvolveram um outro questionário: o C-TCoA, com 31 itens. Assim, o C-TCoA passou a ter duas concepções a mais do que o TCoA: desenvolvimento (a avaliação cultiva qualidades morais, éticas e valores positivos em estudantes que contribuem para a sua aprendizagem ao longo da vida) e controle (a avaliação controla o comportamento e as ações dentro e fora da sala de aula. A avaliação é utilizada para melhorar e manter o controle e o domínio da opinião do professor sobre o aluno). As outras quatro concepções são as mesmas para os dois instrumentos. Além disso, outros itens foram acrescentados. Participaram 1014 professores de Hong Kong e 898 da China. Esta análise apontou a existência de invariância fatorial entre os dois grupos de professores. Os professores da China concordaram de forma moderada para forte com a concepção de irrelevância.

Brown e Michaelides (2011) fizeram um estudo comparativo por meio do TCoA-III A (27 itens) entre os professores cipriotas-gregos e os neozelandeses (249 cipriotas-gregos e 929 da Nova Zelândia). No processo de adaptação, o questionário passou a ter 26 itens. Ao contrário dos neozelandeses, os professores gregos concordam mais com as concepções de responsabilização da escola e irrelevância, mas essas duas concepções apresentam uma correlação negativa para esse grupo de docentes.

Deneen e Brown (2011) examinaram as concepções de avaliação de seis professores em Nova Iorque nos Estados Unidos por meio de entrevistas semiestruturadas. Essas entrevistas foram construídas em torno das quatro concepções de avaliação do TCoA, versão abreviada (27 itens). A concepção dominante foi de irrelevância (os participantes viam a avaliação como algo negativo,

¹ Encontramos outras 14 pesquisas, na Nova Zelândia, que utilizaram o TCoA. Elas não serão incluídas aqui por limite de espaço.

como se ela não tivesse um lugar legítimo dentro do processo de ensino e aprendizagem).

Brown e Remesal (2012) estudaram as concepções de avaliação de licenciandos da Nova Zelândia e da Espanha. Participaram 627 espanhóis e 324 neozelandeses (versão abreviada, TCoA-III A). Os neozelandeses concordaram com as concepções de melhora, responsabilização da escola e responsabilização dos alunos. Os espanhóis concordaram com a concepção de irrelevância.

Gebril e Brown (2014) estudaram as concepções de avaliação de professores do Egito (versão abreviada do TCoA). Participaram 507 professores. As respostas obtidas foram separadas em nove fatores de primeira ordem, que estão correlacionados com as quatro grandes concepções (segunda ordem). Os professores egípcios concordaram mais com as concepções de melhora e responsabilização dos alunos. Essas duas concepções tiveram uma correlação forte e positiva.

Portanto, evidencia-se que o *Teachers Conceptions of Assessment* (TCoA) é instrumento adaptado e utilizado em diversos países. Trata-se de uma escala já bem fundamentada e descrita na bibliografia. Assim, utilizar o TCoA possibilita a análise e comparação dos resultados brasileiros com outros países (pesquisa transcultural).

3 MÉTODO

Para a adaptação e validação do instrumento, seguiu-se o modelo de Pasquali (1999), que indica três grandes polos: teórico, empírico (experimental) e analítico (estatístico). Nos procedimentos teóricos, inicialmente, realizamos a tradução da versão original em inglês do questionário TCoA-III para o português. Nessa primeira tradução em comitê, participaram os autores deste trabalho e uma tradutora. Depois, realizou-se o procedimento de *back translation* (tradução inversa), no qual três tradutores independentes, que não se envolveram na tradução original, fizeram a tradução da versão em português para o inglês. A partir da análise dessas traduções, foi possível chegar a uma versão final da tradução do questionário. Participaram da etapa final da tradução em comitê os autores deste trabalho. Em seguida, fez-se a análise semântica, que tem por objetivo confirmar se os itens do instrumento estão claros para a população a que se destina. Uma das maneiras mais eficazes de testar a compreensão dos itens é realizar uma entrevista com pequenos grupos (PASQUALI,

1999). Dez docentes participaram de entrevistas semiestruturadas. Os professores foram separados em dois grupos: um com 6 docentes e outro com 4. Todos os participantes eram de escolas públicas. A partir das entrevistas, algumas alterações no TCoA-III foram feitas.

O polo empírico (experimental) é a etapa de aplicação do questionário, que envolve a definição da amostra, as instruções de aplicação do instrumento e a coleta das informações. Efetuou-se um estudo piloto com 10 professores de escolas públicas. Após os docentes responderem, eles não relataram nenhum problema ou dúvida. Assim, depois da aplicação do teste piloto, não houve necessidade de mudanças no questionário. Posteriormente, aplicou-se a versão final do questionário na mostra total, disponibilizou-se, também, uma versão on-line do instrumento.

O polo analítico envolve as análises estatísticas (Pasquali, 1999). Essa etapa é complementar a todas as outras. Utilizou-se métodos quantitativos ligados à adaptação e validação de instrumentos: Análise Fatorial Exploratória (AFE), consistência interna (coeficiente Alfa de Cronbach) e evidências de validade do instrumento. Segundo Souza, Alexandre e Guirardello (2017) “a validade refere-se ao fato de um instrumento medir exatamente o que se propõe a medir” (p.652). Pesquisadores e desenvolvedores de teste apontam que a tradução do instrumento não é suficiente para estabelecer sua validade em uma segunda língua (Junior *et al.*, 2016). A Análise Fatorial, a consistência interna e as correlações entre os fatores são fontes de evidências sobre padrões de convergência e divergência do instrumento (um dos aspectos da validade do construto). Evitou-se utilizar os termos validade convergente/discriminante, pois na bibliografia eles são mais empregados na comparação com testes diferentes (variáveis externas). A validação convergente deve apresentar uma correlação alta com outro teste que mede um traço teoricamente relacionado ao que o teste mede. Em contraste, a validação discriminante deve mostrar uma correlação nula com outro teste com o qual foi destinado a diferir (PASQUALI, 1999). No entanto, produziu-se uma matriz de correlação entre os fatores como mais uma forma de coletar fontes de evidência de validade.

O *software* MPLUS versão 7.2 foi utilizado para realizar a AFE, que foi conduzida usando um estimador robusto para variáveis categóricas chamado *mean and variance-adjusted weighted least squares* (WLSMV). Usou-se o método de rotação oblíqua *oblimin*, cujos fatores extraídos são correlacionados (Matos;

Rodrigues, 2019). O tipo de correlação utilizada para produzir os resultados foi a correlação policórica, empregada quando as variáveis são ordinais. O MPLUS apresenta os resultados da AFE de uma maneira diferente de outros pacotes estatísticos. Na AFE, são fornecidos os mesmos índices de ajuste de uma Análise Fatorial Confirmatória (AFC).² Os índices de ajuste do modelo utilizados foram: o índice de ajuste comparativo (CFI), a raiz do erro quadrático médio de aproximação (RMSEA) e a raiz do resíduo quadrático médio padronizado (SRMR). Um bom ajuste dos dados ocorre quando o CFI é $\geq 0,95$, o RMSEA $\leq 0,05$ e o SRMR $\leq 0,08$. No CFI, os valores entre 0,90 e 0,95 sugerem um ajuste de dados aceitável, bem como os valores do RMSEA entre 0,05 e 0,08. Valores fora desses limites sugerem que o modelo não deve ser aceito (Fan; Sivo, 2007). Portanto, o critério de corte de CFI $\geq 0,95$, RMSEA $\leq 0,05$ e SRMR $\leq 0,08$ foi usado no presente trabalho para estabelecer um ajuste adequado. Assim, no MPLUS, é necessário estabelecer um número mínimo e máximo de fatores a serem extraídos, com o objetivo de comparar os índices de ajuste de soluções com número de fatores diferentes.

3.1 PARTICIPANTES

Participaram da pesquisa professores da educação básica [N=179; 110 mulheres (61,5%) e 69 homens (38,5%); idade M=37,6 anos, DP=10,5] de cinco estados do Brasil (Minas Geras 151; São Paulo 16; Espírito Santo 6; Rio Grande do Sul 1 e Paraná 1)³. Dentre os respondentes, 109 possuem pós-graduação; 144 dão aulas em instituições públicas, 22 em escolas privadas e 12 em ambas; 84 são professores iniciantes e 95 professores experientes (aqueles que possuem oito anos ou mais de prática docente); 48 são docentes do ensino fundamental, 58 do ensino médio e 60 de ambos os segmentos (Tabela 1).

Tabela 1 – Caracterização da amostra

| Categoria | N |
|-----------|--------------|
| Sexo | |
| Feminino | 110 (61.45%) |

² Vale lembrar que a análise continua sendo uma AFE, onde o pesquisador deixa os dados observados determinarem o modelo fatorial *a posteriori*. Não existe uma estrutura fatorial previamente determinada como no caso da AFC.

³ Pesquisa aprovada no Comitê de ética da UFOP, processo número 39686214.0.0000.5150.

| | |
|----------------|--------------|
| Masculino | 69 (38.55%) |
| Experiência | |
| Iniciantes | 84 (46.92%) |
| Experientes | 95 (53.08%) |
| Pós-graduação | |
| Sim | 109 (60.89%) |
| Não | 60 (33.51%) |
| Dados ausentes | 10 (5.6%) |
| Instituição | |
| Pública | 144 (80.45%) |
| Privada | 22 (12.29%) |
| Ambas | 12 (6.7%) |
| Dados ausentes | 1 (0.56%) |
| Segmento | |
| Fundamental | 48 (26.82%) |
| Médio | 58 (32.40%) |
| Ambos | 60 (31.52%) |
| Dados ausentes | 13 (7.26%) |

Fonte: Elaboração própria (2016).

Portanto, a maioria dos docentes participantes são do sexo feminino, de Minas Gerais, possuem pós-graduação e trabalham em escolas públicas. Na experiência, a amostra é equilibrada, com uma pequena porcentagem maior de professores experientes (isso também acontece na média de idade). Os segmentos da educação básica estão equilibrados, com pequeno número maior de docentes do ensino médio. A amostra é não-probabilística. No entanto, tentou-se garantir a maior heterogeneidade possível.

4 RESULTADOS

Conforme indicado, no software MPLUS, é preciso estabelecer um número mínimo e máximo de fatores a serem extraídos, com o objetivo de comparar os índices de ajuste de soluções com número de fatores diferentes. Indicou-se ao programa soluções entre 1 e 12 fatores. A solução com 12 fatores não convergiu. A Tabela 2 apresenta os índices de ajuste dos 11 modelos.

Tabela 2 – Comparação de índices de ajuste da AFE: Concepções de avaliação

| Modelos | χ^2 | gl | CFI | RMSEA | SRMR |
|-----------|----------|------|-------|-------|-------|
| 1 fator | 2228.104 | 1175 | 0.882 | 0.071 | 0.094 |
| 2 fatores | 1878.624 | 1126 | 0.916 | 0.061 | 0.076 |
| 3 fatores | 1607.503 | 1078 | 0.941 | 0.052 | 0.062 |

| | | | | | |
|------------|----------|------|-------|-------|-------|
| 4 fatores | 1435.146 | 1031 | 0.955 | 0.047 | 0.055 |
| 5 fatores | 1355.154 | 985 | 0.959 | 0.046 | 0.051 |
| 6 fatores | 1275.294 | 940 | 0.963 | 0.045 | 0.048 |
| 7 fatores | 1210.875 | 896 | 0.965 | 0.044 | 0.045 |
| 8 fatores | 1131.523 | 853 | 0.969 | 0.043 | 0.042 |
| 9 fatores | 1051.890 | 811 | 0.973 | 0.041 | 0.039 |
| 10 fatores | 986.788 | 770 | 0.976 | 0.040 | 0.036 |
| 11 fatores | 923.658 | 730 | 0.978 | 0.038 | 0.033 |

Fonte: Elaboração própria (2016).

Nota: Todos os valores χ^2 foram estatisticamente significantes ($p < .0000$). χ^2 : qui-quadrado, gl: graus de liberdade.

Na Tabela 2, fica evidenciado que, a partir do modelo de 4 fatores, todos os índices de ajuste atendem o critério de corte proposto. Sendo assim, qualquer escolha feita entre 4 e 11 fatores está justificada empiricamente. A partir desses resultados, fica claro que a decisão não deve ser apenas empírica, mas também teórica. Nesse sentido, optamos por analisar os modelos que mais se aproximassem da estrutura fatorial original do instrumento (FIG. 1). Escolheu-se dois modelos: o com 4 fatores (representando as quatro grandes concepções de avaliação de professores: melhora, responsabilização do aluno, responsabilização da escola e irrelevância) e o com 9 fatores (representando os 9 fatores de primeira ordem da estrutura fatorial original do TCoA: responsabilização do aluno, responsabilização da escola, melhora, descrição, aprendizado do aluno, válida, ensino, irrelevância, ruim, ignorada e imprecisa).

Para o modelo de quatro fatores, não se obteve o resultado esperado teoricamente, pois a hipótese inicial seria encontrar uma estrutura semelhante às quatro grandes concepções de avaliação de professores. Assim, esse resultado sugere uma estrutura fatorial diferente para o TCoA brasileiro. Da mesma forma, a partir dos resultados da AFE, fica evidente a impossibilidade de nomear (interpretar teoricamente) de maneira satisfatória o modelo de 9 fatores. Esse não foi o resultado esperado teoricamente, pois a hipótese era encontrar uma organização semelhante aos 9 fatores de primeira ordem da estrutura fatorial original do TCoA (FIG. 1). Vale ainda destacar que, no caso do modelo de 4 fatores, apesar de algumas diferenças, isso foi possível de ser feito em certa medida.

Logo, tomados em conjunto, os resultados da AFE sugerem uma estrutura fatorial diferente para o TCoA brasileiro. Dessa forma, análises adicionais foram realizadas e excluiu-se itens que se mostraram problemáticos. Utilizou-se uma combinação de critérios simultaneamente: carga fatorial baixa, comunalidade baixa e indeterminação fatorial. Os itens foram removidos em três análises subsequentes. Na primeira análise, foram excluídos 23 itens: 1, 5, 7, 11, 12, 16, 17, 19, 22, 28, 29, 30, 32, 34, 36, 37, 38, 39, 40,

44, 46, 48 e 49; na segunda, quatro itens: 6, 8, 21 e 47, e na terceira análise o item 35. Assim, foram excluídos 28 itens da versão original do TCoA (Quadro 1).

Quadro 1 – Versão aplicada do *Teachers' Conceptions of Assessment* – TCoA

| | |
|--|---|
| 1. Professores prestam atenção às avaliações somente quando as consequências são maiores | 26. A avaliação estabelece o que os alunos aprenderam |
| 2. Avaliações selecionam alunos para o futuro educacional ou oportunidades de emprego | 27. A avaliação separa os alunos em categorias |
| 3. Os resultados da avaliação são confiáveis o suficiente para que eu possa basear minhas decisões | 28. A avaliação averigua o progresso em comparação com os objetivos de aprendizagem |
| 4. A avaliação é um fator positivo para melhorar o clima na sala de aula | 29. Os professores deveriam levar em conta o erro e a imprecisão em todas as avaliações |
| 5. A avaliação prejudica o ensino | 30. A avaliação é um indicador preciso da qualidade da escola |
| 6. As informações da avaliação são coletadas e usadas durante o ensino | 31. Os resultados da avaliação são confiáveis |
| 7. A avaliação permite que estudantes diferentes recebam instruções diferentes | 32. A avaliação mede as habilidades de pensamento complexas dos alunos |
| 8. Os professores avaliam exageradamente | 33. A avaliação modifica a maneira que os professores ensinam |
| 9. A avaliação faz com que os alunos deem o seu melhor | 34. A avaliação mostra o quanto as escolas agregam ao aprendizado do aluno |
| 10. A avaliação mostra como os alunos pensam | 35. A avaliação está integrada com a prática de ensino |
| 11. A avaliação é injusta com os alunos | 36. A avaliação é um processo impreciso |
| 12. A avaliação dá um retorno para os alunos sobre suas necessidades de aprendizagem | 37. A avaliação é um bom modo de avaliar a escola |
| 13. A avaliação é objetiva | 38. A avaliação é apropriada e benéfica para os estudantes |

| | |
|--|---|
| <p>14. Os resultados das avaliações são consistentes</p> <p>15. Avaliação é uma experiência envolvente e agradável para os alunos</p> <p>16. Avaliação força os professores a ensinar de uma maneira contrária às suas crenças</p> <p>17. A Avaliação mede a qualidade das escolas</p> <p>18. As informações da avaliação modificam o andamento do ensino dos alunos</p> <p>19. Avaliação é comparar o trabalho do estudante com base em critérios estabelecidos</p> <p>20. A avaliação determina se os alunos atingem padrões de qualidade</p> <p>21. A avaliação fornece feedback para os alunos sobre o seu desempenho</p> <p>22. A avaliação identifica os pontos fortes e fracos dos alunos</p> <p>23. Os professores ignoram as informações da avaliação mesmo quando eles as coletaram</p> <p>24. Os resultados das avaliações predizem o desempenho futuro dos estudantes</p> <p>25. Avaliação é dar uma nota ou conceito para o trabalho do aluno</p> | <p>39. As respostas das avaliações mostram o que se passa na cabeça dos alunos</p> <p>40. A avaliação tem um impacto pequeno no ensino</p> <p>41. A avaliação influencia a maneira que os professores pensam</p> <p>42. A avaliação fornece informação sobre como as escolas estão indo</p> <p>43. Os professores conduzem avaliações mas usam pouco seus resultados</p> <p>44. A avaliação mantém as escolas em um padrão esperado e satisfatório</p> <p>45. A avaliação ajuda os alunos a melhorar seu aprendizado</p> <p>46. A avaliação é sem valor</p> <p>47. A avaliação é completar checklists (uma lista que indica para o estudante quais elementos precisam estar presentes na avaliação)</p> <p>48. A avaliação é uma forma de determinar o quanto os alunos aprenderam do ensino</p> <p>49. Os resultados das avaliações deveriam ser tratados cautelosamente por causa de erros de medida</p> <p>50. Os resultados das avaliações são arquivados e ignorados</p> |
|--|---|

O modelo final apresentou o seguinte ajuste: $\chi^2(149, N= 179) = 269.134, p < 0.000$, CFI= .960; RMSEA= .067; SRMR = .045. Dois índices indicaram bom ajuste do modelo (CFI e SRMR) e um índice indicou um ajuste dos dados considerado aceitável (RMSEA). As cargas fatoriais e as comunalidades dos itens do modelo final do TCoA brasileiro são apresentados na Tabela 3.

Tabela 3 – Análise final TCoA brasileiro - 4 fatores

| Item | 1 | 2 | 3 | 4 | Comunalidade |
|------|--------------|--------------|--------------|--------------|--------------|
| 2 | .533* | -.018 | .098 | .226* | .345 |
| 3 | .734* | -.132* | .074 | .208* | .605 |
| 4 | .608* | .040 | .071 | .068 | .381 |
| 9 | .686* | .091 | -.020 | -.122* | .494 |
| 10 | .639* | .223* | .096 | -.254* | .532 |
| 13 | .633* | -.053 | .051 | -.141 | .426 |
| 14 | .791* | -.044 | -.070 | -.170* | .661 |
| 15 | .548* | .030 | -.300* | .099 | .401 |
| 18 | .087 | .586* | -.086 | -.096 | .367 |
| 20 | .605* | .112* | .000 | .103 | .389 |
| 23 | 0.038 | -.038 | .667* | .076 | .453 |
| 24 | .574* | -.096 | -.007 | .103 | .349 |
| 25 | .071 | -.053 | -.056 | .652* | .436 |
| 26 | .545* | .165* | -.127* | .211* | .385 |
| 27 | -.028 | .150* | .155* | .691* | .528 |
| 31 | .711* | .088 | -.208* | .032 | .557 |
| 33 | -.111* | .790* | -.069 | .116* | .655 |
| 41 | -.015 | .614* | .220* | .041 | .427 |
| 42 | .324* | .526* | -.073 | -.056 | .390 |
| 43 | .043 | .071 | .674* | -.077 | .467 |
| 45 | .351* | .505* | -.141* | -.049 | .400 |
| 50 | -.084 | -.074 | .731* | .073 | .552 |

Fonte: Elaboração própria (2016).

Nota: as cargas fatoriais mais altas estão em negrito. Método de rotação: *oblimin*. *estatisticamente significativo, $p < .05$

A estrutura original do TCoA é (Quadro 1): Melhora (descrição): 10, 22, 26, 32, 39, 48; Melhora (aprendizado do aluno): 4, 9, 12, 15, 21, 38, 45; Melhora (válida): 3, 13, 14, 24, 31; Melhora (ensino): 6, 7, 18, 33, 35, 41. Responsabilização da escola: 17, 30, 34, 37, 42, 44; Responsabilização do aluno: 2, 19, 20, 25, 27, 28, 47; Irrelevância (ruim): 1, 5, 8, 11, 16; Irrelevância (ignorada): 23, 43, 46, 49, 50; Irrelevância (imprecisa): 36, 29, 49.

Os resultados da AFE foram: Fator 1 itens: 2, 3, 4, 9, 10, 13, 14, 15, 20, 24, 26 e 31 (todos os 5 itens melhora-válida, 3 itens melhora-aprendizado do aluno, 2 itens melhora-descrição e 2 itens responsabilização do aluno); Fator 2 itens: 18, 33, 41, 42 e 45 (3 itens melhora-ensino, 1 melhora-aprendizado do aluno, 1 responsabilização

da escola); Fator 3 itens: 23, 43 e 50 (todos irrelevância-ignorada); Fator 4 itens: 25 e 27 (responsabilização do aluno) (TAB. 3).

A partir da Tabela 3, é possível observar que o fator 1 representa prioritariamente a concepção de melhora. Nesse fator estão presentes 10 itens pertencentes a essa concepção. Os dois itens de responsabilização do aluno (2 e 20) possuem carga fatorial positiva (.533 e .605, respectivamente), o que faz sentido teoricamente, pois segundo Brown et al. (2009) as concepções de melhora e responsabilização do aluno possuem uma relação positiva com a autorregulação da aprendizagem. Além disso, nesse fator, encontram-se todos os itens do fator de primeira ordem “válida” e todos os itens não removidos do fator de primeira ordem “descrição”. O fator 2, também, representa prioritariamente a concepção de melhora. Nesse fator estão todos os itens não removidos do fator de primeira ordem “ensino” (18, 33 e 41). O item de responsabilização da escola (42) possui carga fatorial positiva (.526). Já o fator 3 representa a concepção de irrelevância “ignorada”. Por fim, o fator 4 representa a concepção responsabilização do aluno.

Portanto, para fins interpretativos da AFE, considerando o modelo final, nomeia-se os fatores da seguinte forma: fator 1: melhora; fator 2: melhora-ensino; fator 3: irrelevância-ignorada e fator 4: responsabilização do aluno. Quanto ao resultado esperado teoricamente, a maior diferença é que na versão brasileira não apareceu a concepção de responsabilização da escola.

Também fica evidente na Tabela 3, que no modelo final não existem mais itens com cargas fatoriais baixas ou com indeterminação fatorial. Apesar de alguns itens apresentarem comunalidades abaixo do valor de 0,5, destaca-se que as decisões do pesquisador não devem se basear em apenas um critério. Portanto, inclusive em função dos índices de ajuste do modelo, foram mantidos alguns itens com comunalidades consideradas mais baixas. A consistência interna (confiabilidade) dos fatores é apresentada na Tabela. 4.

Tabela 4 – Consistência interna dos fatores do TCoA brasileiro

| Fator | Alfa de Cronbach |
|--------------------------------|------------------|
| 1 (melhora) | .88 |
| 2 (melhora-ensino) | .77 |
| 3 (irrelevância-ignorada) | .70 |
| 4 (responsabilização do aluno) | .61 |

Fonte: Elaboração própria (2016).

Como fica evidente a partir da Tabela 4, todos os fatores apresentaram confiabilidade satisfatória, com exceção do fator 4. Os valores nos quais o alfa de Cronbach $> 0,70$ são satisfatórios, mas para pesquisas exploratórias esse valor pode diminuir para $0,60$ (Matos & Rodrigues, 2019). Provavelmente, o valor mais baixo do coeficiente no fator 4 se deve ao fato dele possuir apenas dois itens. Tecnicamente, é pouco adequado um fator com somente dois itens. Adicionalmente, produziu-se uma matriz de correlação entre os fatores como mais uma forma de coletar fontes de evidência de validade do questionário (Tabela 5).

Tabela 5 – Correlação dos fatores

| | 1) Melhora | 2) Melhora-ensino | 3) Irrelevância | 4) Responsabilização do aluno |
|---|------------|-------------------|-----------------|-------------------------------|
| 1 | 1.000 | | | |
| 2 | 0.405* | 1.000 | | |
| 3 | -0.350* | -0.232* | 1.000 | |
| 4 | -0.039 | 0.008 | 0.181* | 1.000 |

Fonte: Elaboração própria (2016).

Nota: *estatisticamente significativa, $p < .05$.

A partir da Tabela 5, quatro correlações estatisticamente significativas são perceptíveis: correlações positivas entre os fatores 1 e 2 ($r = 0,405$) (maior valor encontrado) isso era esperado teoricamente, uma vez que os dois fatores representam a concepção de melhora (fator 1: melhora; fator 2: melhora-ensino); e fatores 3 e 4 ($r = 0,181$) a correlação significativa, nesse caso, não era esperada, visto que o primeiro representa a concepção de irrelevância e o segundo a concepção de responsabilização do aluno. No entanto, o valor da correlação é baixo. Além disso, existem correlações negativas entre os fatores 1 e 3 ($r = -0,350$), e os fatores 2 e 3 ($r = -0,232$), ambas esperadas teoricamente, já que os fatores representam as concepções de melhora e irrelevância, respectivamente, nos dois casos.

5 DISCUSSÃO

Nesta pesquisa, abordou-se um tema que é pouco pesquisado no Brasil: as concepções de avaliação de professores. Como mencionado, estudos com esse foco são escassos no Brasil (Ribeiro et al., 2010) e este trabalho avança nessa discussão. Entender como os professores estão concebendo a avaliação pode nos apontar caminhos para uma melhor compreensão das práticas avaliativas que têm sido

utilizadas tanto no ensino básico quanto no superior. Nesse sentido, estudos sobre as concepções de avaliação podem contribuir para a reflexão da prática pedagógica. A eficácia das práticas avaliativas está condicionada pela forma como os professores entendem e, conseqüentemente, respondem à avaliação. Destaca-se que conhecer as concepções de avaliação de alunos e gestores é igualmente importante.

A análise teve com foco a adaptação e validação do questionário *Teachers' Conceptions of Assessment* (TCoA-III), instrumento desenvolvido originalmente na Nova Zelândia (2004). A versão do questionário utilizada possui 50 itens, mas existe, também, uma versão abreviada. O TCoA-IIIA (Brown, 2008) possui 27 itens e foi desenvolvido para fornecer uma medição mais sintética das mesmas quatro grandes concepções presentes no TCoA-III (melhora, responsabilização da escola, responsabilização do aluno e irrelevância). Escolheu-se usar a versão completa do questionário por ser a primeira adaptação desse instrumento para a realidade brasileira. Os resultados, no entanto, aproximam a versão brasileira da resumida (no que tange ao número de itens), visto que ficou com 22 itens. Dessa forma, a primeira adaptação e validação para o contexto brasileiro indica caminhos de como melhorar o questionário em estudos futuros.

Acredita-se que a utilização do TCoA-III em pesquisas futuras é uma alternativa viável para pesquisadores interessados na temática das concepções de avaliação de professores em diversos sentidos: aplicação parcial ou integral do questionário propriamente dito; utilização dos itens para elaborar um roteiro de entrevista semiestruturada, como já realizado no estudo feito nos EUA (Deneen; Brown, 2011). Além disso, estudar as concepções de avaliação dos professores poderá ajudar a suprir lacunas – no que se refere à avaliação educacional – nos cursos de formação de professores. Dessa forma, é importante que nos cursos de formação de professores sejam conhecidas as concepções de avaliação dos docentes.

Uma limitação do trabalho diz respeito ao tamanho da amostra. Assim, são necessários trabalhos adicionais no contexto brasileiro que utilizem amostras maiores, no sentido de confirmar os resultados que encontramos. Novas pesquisas são necessárias para investigar de maneira mais aprofundada as concepções de avaliações dos professores, inclusive por meio de outros métodos.

TEACHERS' CONCEPTIONS OF ASSESSMENT QUESTIONNAIRE: ADAPTATION AND VALIDATION FOR THE BRAZILIAN CONTEXT

The aims of this work were: 1) to carry out a literature review about the Teachers' Conceptions of Assessment (TCoA-III) questionnaire and 2) to adapt and collect evidence sources of validity of TCoA-III for Brazilian context. As for review of literature, we found four major groups of teachers' assessment conceptions: 1) improvement; 2) school accountability; 3) student's accountability; 4) irrelevance. We analyzed a total of 25 international studies that used TCoA-III. Regarding the empirical study, elementary school teachers were the participants. We performed an Exploratory Factor Analysis, and the results suggest a factorial structure different for the Brazilian TCoA. We developed several analyses, excluding 28 items that proved problematic. The final model showed a good fit (CFI= .960; RMSEA= .067; SRMR = .045) and four factors: 1) improvement; 2) improvement-teaching; 3) ignore-irrelevance; and 4) student accountability. Thus, this result suggests a different factorial structure for the Brazilian TCoA, which does not include the conception of 'school accountability.' It is believed that the use of TCoA in future research is a viable alternative for researchers interested in the theme of teachers' conceptions of assessment in various senses.

Keywords: Conceptions. Assessment. Teacher. Education. Factor Analysis.

REFERÊNCIAS

ARAUJO, K.; LEITE, R. ; PASSONE, E. F. K. Política accountability educacional no estado do Ceará: repercussões nas dinâmicas pedagógicas. **Revista Teias**, 19 (54), 2018.

BROWN, G. T. L. **Teachers' conceptions of assessment:** implications for policy and professional development. *Assessment in Education*, 11(3), 301–318, 2004.

BROWN, G. T. L. **Conceptions of Assessment:** understanding what assessment means to teachers and students. New York: Nova Science Publishers. 2008.

BROWN, G. T. L. *et al.* Assessment for student improvement: understanding Hong Kong teachers' conceptions and practices of assessment. **Assessment in Education: Principles, Policy & Practice**, 16(3), 347-363, 2009.

- BROWN, G. T. L. *et al.* Teachers' conceptions of assessment in Chinese contexts: a tripartite model of accountability, improvement, and irrelevance. **International Journal of Educational Research**, 50(5-6), 307–320, 2011.
- BROWN, G. T. L.; MICHAELIDES, M. Ecological rationality in teachers' Conceptions of from Cyprus and New Zealand. **European Journal of Psychology of Education**. 26(3), 319-337, 2011.
- BROWN, G. T. L.; REMESAL, A. Prospective Teachers' Conceptions of Assessment: a Cross-Cultural Comparison. **The Spanish Journal of Psychology**, 15(1), 75–89, 2012.
- DENEEN, C.; BROWN, G. T. L. **The persistence of vision**: an analysis of continuity and change in conceptions of assessment within a teacher education program. In: 37th Annual Meeting of the International Association of Education Assessment. pp. 1-10, 2011.
- FAN, X.; SIVO, S. A. Sensitivity of fit indices to model misspecification and model types. **Multivariate Behavioral Research**, 42(3), 509–529, 2007.
- GEBRIL, A.; BROWN, G. T. L. The effect of high-stakes examination systems on teacher beliefs: Egyptian teachers' conceptions of assessment. **Assessment in Education: Principles, Policy & Practice**. 21(1), 16-33, 2014.
- LAMY, B.; GALIETA, T. Funções da avaliação da aprendizagem na formação inicial de professores de Ciências. **Revista Insignare Scientia - RIS**, v. 2, n. 2, p. 166-184, 2019.
- JUNIOR; S. D. D., *et al.* Adaptação transcultural e validação de questionários na área da saúde. **Braz J Allergy Immunol**; 4(1):26-30, 2016.
- MATOS, D. A. S.; RODRIGUES, E. C. **Análise Fatorial** – Brasília: Enap. 2019.
- MIQUELANTE; M. A., *et al.* (2017). As modalidades da avaliação e as etapas da sequência didática: articulações possíveis. **Trabalhos em Linguística Aplicada**, 56(1), 259-299, 2017.
- PASQUALI, L. (ed.) **Instrumentos Psicológicos**: manual prático de elaboração. Brasília: LabPAM, 1999.
- RIBEIRO, V. M. *et al.* Crenças de professores sobre reprovação escolar. **Educação em Revista**, 34, e173086. Epub January 18, 2018.
- SANTOS, L. C. dos *et al.* **Concepções de professores sobre a avaliação**. Anais VII CONEDU - Edição Online... Campina Grande: Realize Editora. 2020. Disponível em: <<https://editorarealize.com.br/index.php/artigo/visualizar/68190>>. Acesso em: 09 out. 2021.
- SILVA, F. C. *et al.* Proposta para Implementar Avaliação Formativa no Ensino Médio. **Ciência & Educação (Bauru)**, 26, e20026. Epub 13 de julho de 2020.

SOUZA, A. C. DE, ALEXANDRE, N. M. C.; GUIRARDELLO., E. DE B. **Propriedades psicométricas na avaliação de instrumentos:** avaliação da confiabilidade e da validade. *Epidemiologia e Serviços de Saúde*, 26(3), 649-659, 2017.