

Avaliação das características psicométricas dos questionários utilizados nos periódicos da área contábil: um estudo longitudinal compreendido no período 2003-2012.

Roseane Patrícia Araújo Silva (UEPB) - rosepat5@hotmail.com

Larissa Cristina Bazilio de Macêdo (UEPB) - larissinha_b_cg@hotmail.com

Izabela Lorena Ribeiro da Silva (uepb) - bela.week@hotmail.com

Resumo:

O desenvolvimento deste trabalho traz como objetivo principal avaliar o rigor metodológico instituído para os instrumentos de medidas psicossociais, utilizados nas pesquisas em contabilidade, no que concerne às suas características psicométricas. Foram avaliados 5 periódicos da área contábil, categorizados como Qualis Capes A, por um período de 10 anos e encontrou-se um total de 37 artigos que utilizaram o questionário como instrumento de coleta de dados para medidas comportamentais. Metodologicamente, este trabalho, caracteriza-se como exploratório, descritivo e de caráter longitudinal. Utilizou-se o protocolo de validação de Kim (2009) para averiguar as etapas de validação utilizadas, quais sejam: validade de conteúdo, pré-teste, confiabilidade e validade de construto. Dos 37 artigos selecionados, nenhum deles apresenta relato descritivo sobre a validade de conteúdo. Entre estes, 16 não fazem nenhuma consideração sobre pré-teste, confiabilidade ou validade de construto. Do restante, 13 artigos apresentam as características de confiabilidade; 10 artigos apresentam a validade de construto e 10 evidenciam o uso do pré-teste. Ressalta-se que algumas evidências das etapas de validação dos artigos apresentaram-se simultaneamente. Observa-se a existência de preocupação e conhecimento por parte dos investigadores no que aduz ao processo de validação dos instrumentos de medidas psicossociais, no entanto preciosas investigações são desenvolvidas com ausências ou relatos parciais de evidências estatísticas que comprovem a legitimidade dos resultados encontrados.

Palavras-chave: *Características psicométricas; Instrumentos de medida; Periódicos da área contábil; Validação de questionários; Análise longitudinal.*

Área temática: *Metodologias de ensino e pesquisa em custos*

Avaliação das características psicométricas dos questionários utilizados nos periódicos da área contábil: um estudo longitudinal compreendido no período 2003-2012.

RESUMO

O desenvolvimento deste trabalho traz como objetivo principal avaliar o rigor metodológico instituído para os instrumentos de medidas psicossociais, utilizados nas pesquisas em contabilidade, no que concerne às suas características psicométricas. Foram avaliados 5 periódicos da área contábil, categorizados como Qualis Capes A, por um período de 10 anos e encontrou-se um total de 37 artigos que utilizaram o questionário como instrumento de coleta de dados para medidas comportamentais. Metodologicamente, este trabalho, caracteriza-se como exploratório, descritivo e de caráter longitudinal. Utilizou-se o protocolo de validação de Kim (2009) para averiguar as etapas de validação utilizadas, quais sejam: validade de conteúdo, pré-teste, confiabilidade e validade de construto. Dos 37 artigos selecionados, nenhum deles apresenta relato descritivo sobre a validade de conteúdo. Entre estes, 16 não fazem nenhuma consideração sobre pré-teste, confiabilidade ou validade de construto. Do restante, 13 artigos apresentam as características de confiabilidade; 10 artigos apresentam a validade de construto e 10 evidenciam o uso do pré-teste. Ressalta-se que algumas evidências das etapas de validação dos artigos apresentam-se simultaneamente. Observa-se a existência de preocupação e conhecimento por parte dos investigadores no que aduz ao processo de validação dos instrumentos de medidas psicossociais, no entanto preciosas investigações são desenvolvidas com ausências ou relatos parciais de evidências estatísticas que comprovem a legitimidade dos resultados encontrados.

Palavras-chave: Características psicométricas; Instrumentos de medida; Periódicos da área contábil; Validação de questionários; Análise longitudinal.

Área Temática : Metodologias de ensino e pesquisa em custos.

1 Introdução

A preocupação com a aferição de medidas psicossociais no âmbito contábil remonta há algum tempo. De maneira mais específica, esta preocupação está centrada nas características psicométricas dos instrumentos de medida, quais sejam: a confiabilidade e a validade. A validação psicométrica de instrumentos de medida em pesquisas acadêmicas constitui-se em um esforço sistemático para assegurar a medição do comportamento, sendo, este, explicado, teoricamente, pelo método científico.

A psicometria tem como base construções teóricas, as quais, denominam-se construtos, que por sua vez têm o objetivo de organizar e atribuir significados ao ambiente. Os construtos ou variáveis latentes não podem ser diretamente observados, no entanto, podem ser mensurados pelos atributos ou indicadores provenientes da clarificação e definição dos construtos (PASQUALI, 2009).

Correntes de autores atestam a fragilidade dos resultados das pesquisas que utilizam medidas comportamentais no âmbito organizacional, especialmente investigações inerentes à pesquisa em educação contábil e contabilidade gerencial. Consideram, no entanto, a existência de complexidades e desafios a serem superados no processo de validação destas

medidas e concordam com a ausência de procedimentos difundidos na literatura que apontem para um conjunto de *best practices* a serem seguidas pelos pesquisadores (SCHREISHEIM ET AL., 1993; HINKIN, 1995; DUFF, 2002; YOUNG, 2006; VAN DER STEDE, YOUNG & CHEN, 2005). Os referidos autores apresentam em suas investigações críticas e sugestões para mitigar os efeitos prejudiciais ocasionados por desenhos incoerentes dos questionários avaliados em suas pesquisas, objetivando maximizar a credibilidade dos dados obtidos através de melhores indícios de confiabilidade e validade.

Diante destes aspectos e alicerçado nas evidências empíricas relatadas pelos estudos que fundamentam o desenvolvimento desta pesquisa, buscou-se, nesta investigação, avaliar as características psicométricas dos instrumentos de medidas psicossociais utilizados nos artigos publicados nos periódicos brasileiros Qualis Capes A. A opção por esta categoria de periódicos parte da premissa de que artigos publicados em revistas de máxima qualificação possuem maior rigor metodológico na construção de seus questionários, atestando maior veracidade aos resultados apresentados. Assim, a seguinte questão problema foi proposta: Os artigos científicos publicados em revistas Qualis Capes A da área contábil apresentam rigor metodológico no que concerne a comprovação das características psicométricas de confiabilidade e validade das escalas psicossociais utilizadas nas investigações?

Neste sentido, o desenvolvimento deste trabalho apresenta os seguintes objetivos fundamentais: 1. Oferecer um aporte teórico sobre as características fundamentais dos testes psicométricos e, 2. Avaliar como os artigos propagados pelos periódicos da área contábil, classificados como Qualis Capes A, estão atuando no processo de validação de medidas das respostas psicossociais.

Quanto aos aspectos metodológicos este trabalho possui características descritivas, exploratórias, de caráter empírico e longitudinal. A amostra consiste nos periódicos, veiculados por instituições acadêmicas nacionais, categorizados como Qualis Capes A e a avaliação compreende um período de dez anos.

Com o intuito de atingir os objetivos propostos, a presente investigação apresenta, sequencialmente, considerações sobre relatos empíricos anteriores que pautam o desenvolvimento desta pesquisa, aportes teóricos sobre a confiabilidade e validade das medidas psicológicas, descrição do desenho da pesquisa, resultados encontrados, considerações finais e por fim as referências bibliográficas.

2 Investigações prévias

Algumas investigações trazem expressamente a dificuldade encontrada pelos pesquisadores no processo de demonstração das características de confiabilidade e validade dos instrumentos psicométricos. Consubstanciadas na observação sobre construção de instrumentos de medidas não observáveis, as pesquisas reportam críticas inerentes aos resultados gerados nos estudos, os quais carecem de cientificidade por não oferecerem suporte sobre a validade das medidas subjetivas geradas (SCHREISHEIM ET AL., 1993; HINKIN, 1995; YOUNG, 1996; ROBERTS, 1999; DUFF, 2002; YOUNG, 2006; VAN DER STEDE, YOUNG & CHEN, 2005; KIM, 2009)

A investigação de Duff (2002) faz menção à ampla utilização de instrumentos psicométricos na pesquisa em educação contábil. Enfatiza que a construção de questionários para avaliar atitudes, percepções ou opiniões de estudantes sobre estilos de aprendizagem, criatividade, ansiedade, entre outros aspectos comportamentais do meio acadêmico, carecem, frequentemente, de indícios que atestem a confiabilidade e validade das escalas produzidas. Afirma que o aspecto mais fraco da pesquisa baseada em questionário diz respeito à precisão da medição dos construtos que estão sendo investigados, e neste sentido, esclarece que muitos estudiosos fizeram importantes previsões em suas pesquisas usando instrumentos que foram invalidados em pesquisas posteriores. Tomando como base a avaliação de dois importantes

periódicos: *Accounting Education: an International Journal e Issues in Accounting Education*, por um período de 5 anos, conclui que as pesquisas em educação contábil apresentam resultados comprometidos por falhas cometidas pelos pesquisadores na determinação da confiabilidade e validade dos dados produzidos pelos testes.

Van der Stede, Young e Chen (2005) avaliaram oito periódicos de contabilidade por um período de 20 anos (1982-2001), focando na qualidade dos questionários utilizados nas pesquisas em contabilidade gerencial e, ao examinarem questões relacionadas à confiabilidade e validade, concluíram que com o passar do tempo a pesquisa *survey* em contabilidade gerencial, através do uso de questionário, apresentou evolução, mas que é necessária uma atenção especial para melhorar ainda mais o caminho no qual o desenvolvimento dos questionários é realizado, sendo este um fator essencial uma vez que há evidências de que a credibilidade dos dados das pesquisas deixa a desejar.

Hinkin (1995) ao propor uma revisão das práticas utilizadas no desenvolvimento de escalas de medidas comportamentais, tendo como amostra seis periódicos representativos da área de comportamento organizacional, entre os quais destaca o *Journal of Management* e o *Academy of Management Journal*, em um período de 5 anos, relata que ao considerar o conjunto de estudos analisados é evidente a existência de problemas no processo e divulgação do desenvolvimento destas medidas. Destaca a existência de excelentes exemplos de desenvolvimento de escalas, mas que estes fatos isolados não devem servir para dar exemplo para outros estudiosos. Defende que as escalas devem ser desenvolvidas para medir com precisão a dinâmica do aspecto investigado e que a pesquisa de qualidade deve começar com a medição de qualidade, atribuindo, neste sentido, responsabilidade tanto para o pesquisador como para o revisor. Acrescenta a estas considerações que se alguém acredita que os problemas decorrem de ignorância ao invés de negligência, a solução é a educação, em vez da advertência.

Roberts (1999) propõe em seu trabalho evidências descritivas e críticas inerentes ao desenvolvimento de uma pesquisa *survey*, por meio de questionário, através da construção de uma escala de medidas para mensurar a satisfação dos usuários do sistema de informação gerencial de uma organização do âmbito industrial australiano. Oferece subsídios para mitigar os problemas decorrentes deste tipo de pesquisa, demonstrando como superá-los. Abarca, principalmente, os seguintes aspectos: validade de conteúdo, pré-teste, seleção e tamanho da amostra e taxas de não resposta. Conclui sua investigação com uma discussão sobre as vantagens e desvantagens do método *survey* nas pesquisas que propõem estudar as características e interrelações entre as variáveis sociológicas e psicológicas.

Schreisheim et al. (1993) defendem que a demonstração do desenvolvimento de práticas de validação de conteúdo é uma propriedade psicométrica essencial da medida de um construto, assegurando que graves deficiências de medidas podem ser mitigadas quando os pesquisadores reportam adequadamente o conteúdo do construto investigado. Ao considerarem que não existem métodos quantitativos bem estabelecidos para a validade de conteúdo, sugerem, a técnica de redução de dados para auxiliar no processo. Através da avaliação de 4 periódicos, quais sejam : *Academy of Management Journal*, *Journal of Management*, *Strategic Management Journal* e *Journal of International Business Studies*, durante o período de um ano, relatam a debilidade inerente aos seguintes aspectos investigados : seleção e modificação dos itens, modificação da escala resposta e falhas para reportar a confiabilidade.

A investigação de Young (1996), centrada nos artigos das principais revistas de contabilidade no período 1985-1994 que utilizaram o questionário como instrumento de coleta de dados, encontrou um pobre padrão na aplicação do uso de questionário na investigação em contabilidade de gestão. O referido autor observou que a maioria dos estudos apresentava falhas em usar métodos de amostras rigorosas, não averiguavam os erros de não resposta e

sofriam de baixas taxas de respostas, o que, conseqüentemente, resultava em amostras não representativas e de baixo poder estatístico, dificultando as inferências dos investigadores e a solidez das conclusões apresentadas nas pesquisas.

O estudo de Kim (2009) estabelece seis etapas a serem colocadas em prática no processo de validação de instrumentos de medidas psicométricas, a estas etapas denomina “protocolo de validação”, são elas: validade de conteúdo, pretest, teste piloto, manipulação de validade, confiabilidade e validade de construto. Ao focar seu estudo na ciência da informação, avalia o periódico *Journal of the American Society for Information Science and Technology* por um período de 25 anos (1982-2007) e, tomando como base o referido protocolo, observa maior ênfase na existência de relatórios de validação nos artigos do último período avaliado na pesquisa (2007), fornecendo indícios para comprovação de avanço promissor para prática de métodos de pesquisa que enfatizam a importância dos testes de validade na área objeto de estudo. As técnicas mais amplamente empregadas foram teste piloto, confiabilidade e validade de construto, enquanto que a validade de conteúdo, pré-teste e manipulação de validade foram pouco utilizadas pelos pesquisadores.

Considerando que não há um único critério compartilhado entre os investigadores a respeito do processo de validação de medidas psicométricas, optou-se, neste estudo, seguir o protocolo de validação proposto por Kim (2009), alicerçado na tipificação de etapas propostas pelo autor, a qual permite detalhar de maneira mais explícita o processo de validade dos instrumentos que aferem atributos subjetivos. Assim, utilizou-se da investigação do referido autor com algumas adaptações, conforme exposto na tabela 1, a qual associa para cada categoria do protocolo de validação as técnicas correspondentes.

Tabela 1. Adaptação do protocolo de validação proposto por Kim (2009).

Validação	Técnicas
Validade de conteúdo	Julgamento de <i>experts</i>
Pretest	Abordagem pessoal (face a face, entrevista por telefone), abordagem impessoal (questionário).
Confiabilidade	Teste-reteste; <i>split-half</i> (Alfa de Cronbach)
Validade de construto	Varia de acordo com os objetivos da pesquisa. Principais técnicas: validade convergente, validade discriminante, análise fatorial exploratória e confirmatória.

FONTE : Adaptado da investigação de Kim (2009).

Abordar-se-á, em seguida, cada uma das etapas descritas na tabela 1 e além destas considerações, far-se-á menção aos aspectos reportados pela literatura no que concerne à validação de instrumentos de medidas subjetivas.

3 Considerações sobre confiabilidade e validade das medidas psicológicas.

A utilização de questionários permite aos investigadores trabalhar com amostras amplas e estabelecer comparações estatísticas, sempre e quando se realize um bom processo de construção, distribuição e análise do questionário (DILLMAN, 2000; IBERT ET AL., 2001).

Robert (1999) e Dillman (2000) afirmam que as desvantagens decorrentes do uso de questionários podem ser minimizadas com um bom planejamento, um cuidadoso desenho, distribuição e detalhada análise psicométrica dos dados coletados. Brownell (1995) considera que quando o instrumento de medida utilizado nas investigações é cuidadosamente desenvolvido e validado antes de ser efetivamente usado, os resultados apresentarão uma grande confiabilidade e validade, sendo estas, as duas qualidades fundamentais dos testes psicométricos.

Neste sentido, Muñiz et al. (2005) esclarecem que a elaboração e análise dos itens de um instrumento de medida devem ser alicerçadas em critérios que permitam satisfazer as características de confiabilidade e validade do mesmo. Para os autores, na prática, estes aspectos são frequentemente esquecidos, convertendo-se a análise de itens em uma acumulação desregrada de indicadores estatísticos, cuja relação com as propriedades globais do teste é normalmente pouco clara ou ignorada.

No que diz respeito à validade, Navas Ara (2007, p. 266) considera que a mesma “é um critério métrico de qualidade que proporciona informação acerca da adequação da interpretação das pontuações do teste em termos do construto psicológico que se está medindo ou de sua utilização para o fim proposto”. Para Cronbach (1955), a validação é um processo; validar é investigar, o que consiste na acumulação de dados e análises, o que vai dar credibilidade a uma interpretação determinada à utilidade de testes e escalas.

De acordo com a revisão da literatura se pode estabelecer que os procedimentos fundamentais para a validação são: validade de conteúdo, validade de construto e a validade referida ao critério (MARTÍNEZ ARIAS, 1995; ABAD ET AL., 2006; BARBERO GARCÍA ET AL., 2003; SÁNCHEZ & SARABIA, 1999; LUQUE, 1997). Aos referidos procedimentos denominam-se “santíssima trindade da validade” (Muñiz et al., 2005) ou modelo trinitário (PASQUALI, 2009).

3.1 Validade de Conteúdo

Segundo Luque (1997) a validade de conteúdo trata de saber se a medida representa os diferentes aspectos ou facetas do fenômeno que se deseja medir. Aconselha que para a obtenção de uma melhor validade de conteúdo, o investigador, deve tentar minimizar as possíveis críticas sobre o aspecto investigado através da revisão da literatura existente, especificando, através de uma correta definição conceitual o que é, ou não é, objeto de medida. Além desses aspectos, deve-se gerar uma ampla relação de itens que representem todas as dimensões relevantes do fenômeno estudado e solicitar uma avaliação de juízes e expertos que informem sobre o âmbito do conteúdo a abarcar. Não há formas de garantir a comprovação da validade de conteúdo, tendo em vista que ela depende do grau de revisão da literatura, de juízos de valor do investigador, além de supor que foram consideradas todas as possíveis dimensões e conteúdos do conceito, fenômeno ou comportamento investigado (NUNNALLY, 1987; SÁNCHEZ, 1999; LUQUE, 1997).

3.2 Pré-teste

Martín Arribas (2004) ao aludir sobre a prova piloto a denomina também de pré-teste, sendo, pois, termos intercambiáveis. Considera, igualmente à concepção de Kim (2009), a prova piloto/pré-teste como a segunda etapa do processo de validação. O pré-teste permite averiguar se o instrumento de medida avaliado comunica aos entrevistados uma parte, ou todos os aspectos que ele tem a intenção de comunicar. Avalia aspectos como: tempo de duração para administrar o questionário, resistências psicológicas, compreensão dos enunciados, correta categorização das respostas, tipos de perguntas mais adequadas, entre outros.

Para Wikman (2006) a comunicação é a chave para garantir a confiabilidade. É importante, neste sentido, destacar que a confiabilidade é uma pré-condição necessária para a validade (Nunnally, 1978). Assim, a definição de um bom teste traz as seguintes características segundo Cohen et al. (1996): 1. Ter instruções claras e ser econômico na sua administração, pontuação e interpretação; 2. Medir o que pretende medir; 3. Os resultados devem proporcionar uma melhor qualidade de vida para os entrevistados e outros.

O pré-teste pode ser conduzido pessoalmente ou impessoalmente, depende do objetivo do pré-teste.

Estabelecendo uma diferença entre o pré-teste e o teste piloto Lewis et al. (2005) afirma que este é precedido de uma série de pré-testes e é realizado tomando como base uma amostra conveniente. Ainda neste contexto, Yin (2003) caracteriza o teste piloto como uma versão em pequena escala da pesquisa final, que permite aos investigadores antecipar resultados do estudo e determina se o instrumento construído mede de forma eficaz o construto da pesquisa.

No desenvolvimento deste trabalho optou-se por averiguar a utilização do pré-teste ou teste piloto como procedimentos equivalentes no âmbito da validação de questionários, mesmo reconhecendo a existência destes distintos momentos no processo de validade de instrumentos, uma vez que o teste piloto, conforme explicitado, é concebido como a primeira versão de uma escala de medidas, com características que permitem inferir resultados em decorrência de satisfatórios atributos de qualidades psicométricas.

3.3 Confiabilidade

A confiabilidade de um teste se define como um critério métrico de qualidade relacionado com a quantidade de erro aleatório contido nas medidas que foram obtidas ao aplicar o teste a um grupo de sujeitos (ALIAGA, 2006).

Uma medida é confiável quando, ao ser aplicada repetidas vezes, se obtém medições semelhantes, ou seja, quando uma escala permite obter medidas similares em diferentes momentos de tempo se pode caracterizar como uma escala confiável (SÁNCHEZ & SARABIA, 1999).

Uma medida pode ser confiável, mas não ser válida, no entanto, o contrário não é possível. A falta de confiabilidade pode ser uma razão para questionar a validade de um instrumento de medida.

As técnicas comumente utilizadas para avaliar a confiabilidade são: provas repetidas ou teste-reteste e o método *split-half* (KIM, 2009; MUÑIZ ET AL., 2005; SÁNCHEZ & SARABIA, 1999; LUQUE, 1997).

A confiabilidade através do teste-reteste é realizada através da administração da escala em dois intervalos de tempo separados. Neste contexto, considera-se que se dois ou mais testes realizados aos mesmos sujeitos em momento diferentes e em circunstâncias similares proporcionam a mesma medida ou medidas muito correlacionadas então o instrumento é estável. Há ressalvas quanto ao uso deste método, tendo em vista que o intervalo de tempo entre a aplicação do primeiro para o segundo teste e as possíveis alterações nas circunstâncias em que os mesmos foram aplicados pode afetar os resultados encontrados, recomendando-se, desta forma, outras formas de avaliação.

O método *split-half* é também conhecido como confiabilidade por consistência interna. A técnica mais evidente nas investigações acadêmicas para obter esta avaliação é através da utilização do coeficiente *alfa de Cronbach*. Para interpretação deste coeficiente, se esclarece que não existe um único critério compartilhado por todos os investigadores a respeito da magnitude real do seu valor (Peterson, 1994; Morales Vallejo et al., 2003). Seus valores variam entre 0 e 1 e, a medida que se incrementa seu valor, maior é a consistência interna e consequentemente menor é a variância. No entanto, um valor mínimo aceitável para fins de investigação, de acordo com Nunnally (1978) é de 0,70. Na investigação básica pode ser suficiente um valor em torno de 0,60 (MORALES VALLEJO ET AL., 2003).

Registre-se que, a construção de escalas mais longas e a existência de um maior número de respostas aos itens formulados, são fatores que aumentam a confiabilidade das escalas. No entanto, deve-se considerar que escalas longas, apesar de gerarem altos valores do coeficiente alfa, ocasionam fadiga aos respondentes e levam mais tempo para serem administradas. Neste sentido, Morales Vallejo et al. (2003) recomenda que aumentar o

número de respostas nos itens pode ser mais prático em algumas ocasiões investigatórias, tendo em vista que este procedimento aumenta a covariância entre os itens.

A confiabilidade composta e a variância extraída são apresentadas por Hair et al. (2005) como coeficientes de consistência interna, sendo estabelecido um valor mínimo aceitável de 0,70 para o primeiro e de 0,50 para o segundo.

3.4 Validade de Construto

A validade de construto, também denominada validade de conceito, é o tipo de validação que dá significado as pontuações dos testes, pois aponta para a obtenção de evidências de que as condutas observáveis que foram escolhidas como indicadores do construto realmente o representam (Barbero García et al., 2003). A validade de construto determina se os itens operacionalizados, ou seja, as perguntas do questionário, medem o conceito pretendido para análise (CAMPBELL & FISKE, 1959).

Ao definir “construto”, Abad et al. (2006, p.62) transcrevem “um conceito elaborado pelos teóricos da psicologia para explicar o comportamento humano”. Acrescentam, além disso, que para o estudo de um construto é indispensável a existência de indicadores observáveis, onde, em muitas ocasiões, estes indicadores são os itens de um teste, os quais se deve comprovar empiricamente se são adequados para representar o construto em questão.

As concepções de Morales Vallejo *et al.* (2003) sobre a validade de construto enfatizam a existência de dois tipos de estratégias complementares para provar as hipóteses instituídas referentes ao construto estudado: a validade convergente e a validade divergente. Na validade convergente se estudará relações esperadas e plausíveis com outras medidas relacionadas a dois tipos de variáveis: 1. Relações estabelecidas com variáveis medidas por outros instrumentos que intencionalmente medem o mesmo construto; 2. Relações com instrumentos que medem outros aspectos com o qual se espera que exista uma relação positiva ou negativa.

A validade divergente, normalmente denominada validade discriminante, consiste no grau em que uma medida não se correlaciona com outras medidas das quais se supõe que deve divergir (SÁNCHEZ, 1999). Morales Vallejo et al. (2003) orientam que deve ser realizado um criterioso planejamento do processo de validação durante a preparação do instrumento para coletar simultaneamente os dados necessários. As correlações entre as medidas envolvidas neste processo podem ser apresentadas através de uma matriz denominada matriz multiconceito-multimétodo ou *multimétodo-multirrasgo*.

Assim, no processo de validação é necessário especificar as hipóteses previstas entre as variáveis envolvidas indicando: 1. O sentido que se espera da relação, seja positivo, negativo, ou ausência de relação e; 2. A magnitude relativa esperada da relação, onde se pode comprovar a existência de relações maiores e mais claras.

A análise fatorial exploratória e confirmatória também são indicadas como técnicas que podem comprovar a validade de construto. Ambas são utilizadas objetivando provar hipóteses anteriormente estabelecidas acerca da estrutura interna do construto investigado e das relações deste com outras variáveis (Barbero García, 2003). Utilizada desde um enfoque exploratório, não se pode estabelecer hipóteses prévias acerca do número de dimensões resultantes do uso da técnica. Quando se concebe hipóteses a priori acerca da estrutura subjacente e do número de dimensões utiliza-se a análise fatorial confirmatória para confirmar tais hipóteses.

A análise fatorial permite comprovar se o instrumento realmente está medindo o que pretende medir ao proporcionar: 1. Nitidez relativa aos aspectos subjacentes a uma série de variáveis; 2. Auxílio na definição de que variáveis ou itens definem cada fator e; 3. Como estes fatores estão correlacionados entre si (MORALLES VALLEJO ET AL., 2003).

Nunnally (1978) defende que a análise fatorial está no centro da medição dos construtos psicológicos. Ressalta-se a indispensável necessidade do conhecimento da teoria subjacente das construções que estão sendo testadas, tendo em vista que a interpretação dos resultados da análise fatorial será menos subjetiva à medida que exista maior domínio sobre este conhecimento.

Uma utilização indicada para técnica de análise fatorial confirmatória é para comprovar se a estrutura de um instrumento adaptado de outro idioma é semelhante à estrutura original, mas normalmente a referida técnica é utilizada para confirmar estruturas fatoriais previamente encontradas na análise fatorial exploratória.

3.5 Validade referida ao critério ou preditiva

A validade referida ao critério se refere ao fato de que uma escala de medida seja capaz de prever outras variáveis, denominadas variáveis critérios, as quais podem ser qualquer tipo de variável obtida com outras escalas e que podem ser estimadas com a escala que está sendo objeto de estudo.

Luque (1997) estabelece que a validade referida ao critério está centrada na habilidade que tem o instrumento para refletir se as relações estabelecidas entre as medidas de uma variável e as de outras estão ou não de acordo com o que estabelece a teoria.

Na validade de critério o investigador correlaciona sua escala com algum critério que esteja correlacionado com a mesma e, ao avaliar o coeficiente de correlação de acordo com hipóteses previamente estabelecidas, observa a existência ou não de validade. As principais técnicas para determinação da validade preditiva são: análise discriminante, análise de correlações e equações de regressão (Sánchez & Sarabia, 1999). Ao apresentar a manipulação de validade como uma das etapas do protocolo de validação, Kim (2009) a define como um teste realizado para determinar relações causais entre variáveis dependentes e independentes, podendo esta etapa da validação estar associada ao processo de validade preditiva.

Este tipo de validade se distingue desde duas perspectivas, de acordo com a temporalidade do critério: validade preditiva simples e validade concorrente.

O desenho de um estudo de validação referida ao critério segue os seguintes passos (Barbero García et al. 2003; Martinez Arias, 1995): 1. Definir um critério relevante e estabelecer um método para sua medição; 2. Selecionar uma amostra de sujeitos representativa da população na qual será usado o teste posteriormente; 3. Aplicar o teste e obter uma pontuação para cada sujeito; 4. Obter uma medida de cada sujeito no mesmo momento da aplicação do teste, o que caracteriza a validade concorrente, ou em um momento posterior, o que caracteriza a validade preditiva; 5. determinar o grau ou a relação entre as pontuações obtidas pelos sujeitos no teste e a medida critério.

4 Metodologia da pesquisa empírica.

4.1 Delineamento da pesquisa

A amostra selecionada para o estudo consiste nos artigos publicados nos periódicos brasileiros qualificados como Qualis Capes A da área contábil na última avaliação da Capes, que trazem em suas investigações empíricas questionários que aferem percepções dos participantes do estudo. Foram identificados 5 periódicos com qualificação A2, são eles : Revista de Administração Contemporânea (Eletrônica), Revista de Administração Contemporânea (Versão impressa), Revista de Administração de Empresas, *Brazilian Business Review-BBR* e Revista Contabilidade e Finanças. A opção por revistas com qualificação Qualis Capes A deu-se por considerar uma maior rigidez no processo de avaliação advinda dos revisores e editores das mesmas. O estudo abarca um período de dez anos correspondente aos anos de 2003 a 2012. Durante o período avaliado foram encontrados

37 artigos, nos cinco periódicos estudados, com uso de questionários que aferiam medidas psicossociais.

Este estudo caracteriza-se, metodologicamente, como um estudo descritivo, exploratório e de caráter longitudinal. Neste sentido, compreende-se como um estudo longitudinal, toda pesquisa direcionada à análise dos processos de mudança e desenvolvimento nos quais o passar do tempo é um fator fundamental (Martínez Arias, 1995). Uma vez selecionados os artigos, buscou-se, através da leitura das características dos procedimentos metodológicos, averiguar a existência de procedimentos de validação dos questionários utilizados, selecionando as principais técnicas utilizadas e comparando-as com o protocolo de validação de Kim (2009) conforme disposto na tabela 1.

5 Resultados encontrados

A etapa referente à validade de conteúdo não é explicitada no corpo dos artigos como uma etapa do processo de validação, mas os investigadores das referidas pesquisas discorrem que a construção dos questionários utilizados foi fundamentada na revisão de amplo arcabouço teórico, o que implica uma ação preliminar do processo, sem, no entanto, oferecer uma real comprovação da validade ou confiabilidade conforme exposto na literatura. Neste sentido, ao averiguar as etapas do processo de validação conforme proposto por Kim (2009), omitiu-se dos resultados expostos no quadro 2 a etapa referente à validade de conteúdo.

Apresenta-se, em seguida, a quantidade de artigos publicados nos periódicos estudados que utilizaram instrumentos psicométricos em suas pesquisas (quadro 1).

É importante frisar que ao avaliar o conteúdo dos artigos, comprova-se que a maioria deles atua na área de contabilidade gerencial e, na sequência, observam-se pesquisas sob o escopo da educação contábil, ética e auditoria.

As Revistas Contabilidade e Finanças e a *Brazilian Business Review-BBR* são as que apresentam o maior número de artigos publicados com uso de questionários para aferir medidas corporativas comportamentais.

Ao avaliar o uso de procedimentos de validação por revista pesquisada, pode-se comprovar que a Revista de Administração Contemporânea (versão impressa) apresenta em 5 dos artigos analisados os processos de confiabilidade e validade de construto, no qual, em apenas um deles apresenta consideração sobre o pré-teste. O artigo publicado em 2004 não faz menção sobre nenhuma característica psicométrica e a publicação de 2007 utiliza uma técnica denominada “Método de análise hierárquica (AHP)” para validação do questionário quanto à amplitude e coerência, eliminando questões com respostas inconsistentes. Quanto ao referido método, segundo Saaty (1991), reflete a forma natural de funcionamento da mente humana, ou seja, de um elevado número de elementos, a mente os agrega em grupos segundo propriedades comuns. Neste sentido, observa-se algo semelhante ao que se pode obter com a análise fatorial exploratória.

Em relação às demais revistas observaram-se os seguintes aspectos: a Revista de Administração Contemporânea (eletrônica) publicou apenas 1 artigo, o qual não apresenta nenhuma consideração sobre o processo de validação; a Revista de Administração de Empresas (RAE) publicou 2 artigos, apenas 1 apresenta confiabilidade e validade de construto; a Revista Contabilidade e Finanças publicou 17 artigos, dos quais 9 não mencionam nenhuma das características psicométricas destacadas no protocolo de Kim (2009), 2 apresentam confiabilidade (tendo um deles usado pré-teste), 5 reportam apenas uso do pré-teste e 1 descreve o processo de validação de construto e confiabilidade; a Revista *Brazilian Business Review-BBR* publicou 10 artigos, dos quais 4 não fazem nenhuma consideração sobre o processo de validação, 3 apresentam validade de construto e confiabilidade, sendo que um deles inclui o pré-teste, 1 inclui apenas consistência interna e 2 descrevem apenas o uso do pré-teste.

Nº de artigos por revista que usaram instrumentos psicométricos.						
Períodos analisados	Revista de Administração Contemporânea (eletrônica)	Rev. Adm. Contemporânea (versão impressa)	Rev. de Adm. de Empresas (RAE)	BBR <i>Brazilian Business Review</i>	Revista Contabilidade Finanças	Total de artigos
2003	-----	1	-----	-----	1	2
2004	-----	1	-----	-----	-----	1
2005	-----	-----	-----	1	-----	1
2006	-----	1	-----	-----	4	5
2007	-----	1	-----	-----	3	4
2008	-----	1	-----	1	4	6
2009	1	1	1	4	-----	7
2010	-----	-----	1	1	2	4
2011	-----	-----	-----	1	3	4
2012	-----	1	-----	2	-----	3
Total de artigos por revista	1	7	2	10	17	37

Fonte : Elaboração própria a partir de dados da pesquisa.

Quadro 1 Nº de artigos publicados nas revistas Qualis Capes A que utilizaram instrumentos psicométricos no período de 2003 a 2012.

O quadro 2 apresenta as etapas do processo de validação utilizadas pelos artigos investigados de acordo com o protocolo de validação utilizado nesta pesquisa, excetuando-se a etapa de validação de conteúdo com justificativa anteriormente reportada.

Seguindo o roteiro traçado para pesquisa, exposto na tabela 1, constatou-se que 27% (10) dos artigos da amostra descreveram o uso do pré-teste, denominado similarmente de teste piloto, como uma das etapas do processo de validação. Ao tratar da confiabilidade 35% (13) dos artigos investigados demonstraram o uso da técnica *split half* para determinação da consistência interna, tendo como coeficientes utilizados o alfa de *Cronbach* e a confiabilidade composta. Há predominância quanto ao uso do alfa de *Cronbach* em detrimento da determinação da confiabilidade composta, sendo encontrados 9 artigos com o uso do primeiro coeficiente e 4 com o uso do segundo.

ANO	Quantidade de artigos por etapas do processo de validação		
	Pré-teste/teste piloto	Confiabilidade	Validade de Construto
2003	1	1	1
2004	-----	-----	-----
2005	1	-----	-----
2006	2	1	1
2007	-----	-----	-----
2008	-----	4	2
2009	2	2	2
2010	2	3	2
2011	0	0	0
2012	0	2	2
Quantificação das etapas do processo de validação por amostra total.	10/37	13/37	10/37
% de artigos por etapa avaliada em relação à amostra total.	27%	35%	27%

Fonte : Elaboração própria a partir de dados da pesquisa.

Quadro 2 Evidência quantitativa dos artigos que apresentaram etapas do protocolo de validação proposto na pesquisa.

No processo de validação de construto as principais técnicas destacadas foram: análise fatorial exploratória (AFE) e análise fatorial confirmatória (AFC) e, com o uso destas técnicas, destacaram-se a validade convergente e divergente. Ao total, 27% dos artigos apresentaram a validação de construto de forma explícita, sendo a AFE a técnica mais evidente estando presente em 6 trabalhos, os 4 artigos restantes usaram a AFC.

A validade de construto obtida através da AFC determinou a validade convergente através da variância média extraída e da análise das cargas fatoriais, esta última, objetivando determinar a unidimensionalidade, a qual representa o grau em que um conjunto de itens representa apenas um construto. Hair et al. (2005) apontam que o contraste de unidimensionalidade consiste em que os itens apresentem valores elevados em um único fator. Esta evidência pode ser obtida através da AFC ou AFE. Sobre a variância média extraída Hair et al. (1998) destaca que ela mede a variância da variável relacionada com a variância devida a um erro de medição e que o nível mínimo recomendado é de 0,5.

Acrescenta-se que a validade convergente mede a extensão em que a escala se correlaciona positivamente com outras medidas do mesmo construto (Malhora, 2001). Assim, correlações significativas e positivas provenientes da matriz de correlação das dimensões (medidas) de um questionário garantem a existência da validade convergente.

Atendendo ao critério de Bagozzi (1994) a validade divergente pode ser verificada através da existência de correlações com valores inferiores a 0,8 entre as referidas medidas.

Alguns dos artigos avaliados propõem hipóteses entre variáveis dependentes e independentes no escopo do construto teórico investigado e, neste sentido, utilizam esta caracterização das variáveis para estabelecer relações que remetem à averiguação de validade convergente e divergente. Neste sentido, os principais procedimentos estatísticos utilizados foram: o teste qui-quadrado e os testes de regressão. Outras duas evidências estatísticas encontradas na amostra total investigada (37 artigos) foram: o teste exato de *Fischer* e a regressão logística, objetivando confirmar hipóteses entre variáveis dependentes e independentes, no entanto, na análise descritiva dos mesmos não há nenhuma consideração sobre a utilização destas informações estatísticas como procedimentos de validação.

6 Considerações finais

A presente pesquisa objetivou oferecer um aporte teórico sobre as principais características dos testes psicométricos, quais sejam: a confiabilidade e a validade, assim como, avaliar como os artigos científicos publicados em revistas brasileiras Qualis Capes A do âmbito contábil estão atuando no processo de validação de instrumentos que aferem medidas psicossociais, tomando como base o protocolo de validação de Kim (2009).

Durante o seu desenvolvimento fez um relato das principais etapas e técnicas utilizadas no processo de validação dos testes psicométricos. Assim, descreveu características indispensáveis para atestar a cientificidade dos resultados das pesquisas baseadas no uso de questionários que aferem medidas subjetivas no meio corporativo, enfatizando: a validade de conteúdo, o uso do pré-teste, a confiabilidade, a validade de construto e a validade preditiva.

Ao averiguar a utilização do protocolo de validação de Kim (2009) nos artigos delimitados na amostra, encontrou que, do universo dos 37 artigos investigados, nenhum deles fazia menção sobre a validade de conteúdo conforme explicitado na literatura, 16 não faziam nenhuma consideração sobre pré-teste, confiabilidade ou validade de construto. Dos 21 restantes, houve maior evidência do processo de confiabilidade através do uso do alfa de *Cronbach* e da confiabilidade composta, sendo 9 incidências para o primeiro coeficiente e 4 para o segundo. O processo inerente à validade de construto se fez presente em 10 artigos, tendo 6 deles utilizado o uso da AFE e 4 a AFC. A etapa do pré-teste esteve evidente em 10

dos artigos avaliados. De maneira simultânea apenas 10 artigos apresentaram evidências de confiabilidade e validade, e destes, 5 também incluíram a etapa do pré-teste.

Observa-se a existência de preocupação e conhecimento por parte dos investigadores no que concerne ao processo de validação dos instrumentos de medidas psicométricas, no entanto preciosas investigações são desenvolvidas com ausências ou relatos parciais de evidências estatísticas que comprovem a legitimidade dos resultados encontrados. Algumas razões podem estar atreladas aos resultados encontrados, como: 1. O desenvolvimento de escalas é um processo difícil e demorado, e o desejo de concluir a pesquisa e submetê-la à publicação pode não ser compatível com o tempo disponível pelo investigador para realizar tal processo (Schimit & Klinoski, 1991); 2. Parece não existir estrutura bem definida para orientar os pesquisadores nos vários estágios do processo de validação (Hinkin, 1995); 3. A profissão pode colocar uma ênfase muito grande na análise estatística (SCHIMIT, 1989).

Compactuando com as considerações de Cronbach (1955), o qual assegura que a validação é um processo e que consiste na acumulação de dados e análises, sugere-se, diante do relato exposto no marco teórico deste trabalho sobre as características dos testes psicométricos, assim como dos resultados encontrados na investigação empírica, que as pesquisas baseadas em medidas psicossociais no âmbito corporativo descrevam qual o processo de validação que legitimou os resultados encontrados, conscientes de que, quanto maior a incorporação de inferências que comprovem a validade dos questionários, maior a legitimidade associada à pesquisa.

Este trabalho oferece como contribuição principal um indicador concernente à avaliação do nível de comprometimento por parte dos investigadores quanto à descrição das características psicométricas dos instrumentos de medidas nos artigos Qualis Capes A, traçando um perfil de como os resultados destas pesquisas apresentam conclusões sólidas, inferências pertinentes e poder estatístico relevante.

A amostra compreendida em um período de dez anos é uma das suas principais limitações. Para futuras investigações sugere-se confrontar estes resultados com periódicos internacionais da mesma área e com a mesma qualificação Qualis Capes, assim como sugere-se ampliar a amostra aumentando o espaço de tempo para coleta de dados.

7 Referência Bibliográfica

ABAD, F. J.; GARRIDO, J.; OLEA, J.; POSANDA, V. *Introducción a la psicometría*. Teoría clásica de los tests y teoría de la respuesta al ítem, Universidad Autónoma de Madrid (UAM), 2006.

ALIAGA, J. *Psicología: Tópicos de actualidad*. Lima: UNMSM. Recuperado em 21 de abril, 2006, obtido de <http://www.unmsm.edu.pe/psicología/documentos>.

BARBERO GARCÍA, M^a. I.; ABAD, E. V. *Psicometría*. Madrid: Universidad Nacional de Educación a Distancia (UNED), 2003.

BROWNELL, P. *Research methods in management accounting*. Austrália: Coopers & Lybrand, 1995.

CAMPBELL, D.; FISKE, D. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, v. 56, 1959, p. 81-105.

COHEN, R.J.; SWERDLIK, M. E.; PHILLIPS, S. M. *Psychological Testing and Assessment*. Mountain View, CA: Mayfield Publishing.

CRONBACH, L. J. Construct validity in psychological test. *Psychological Bulletin*, v. 52, p. 281-302, 1955.

DILLMAN, D. *Constructing the questionnaire*. Mail and internet survey. New York: John Wiley & Sons, 2000.

- DUFF, A. Psychometric methods in accounting education: a review, some comments and implications for accounting education researchers. *Accounting Education*, v. 10 (4), 2001, p.383-401.
- HALL, J.R.; ANDERSON, R.E.; TATHAM, R.L.; BLACK, W.C. *Análisis Multivariante*. Madrid: Prentice-Hall Iberia, 2005.
- HINKIN, T. R. A review of scale development practices in the study of organizations. *Journal of Management*, v. 21(5), 1995, p. 967-988.
- IBERT, J.; BAUMARD, P.; DONADA, C., XUEREB, J. *Data Collection and Managing the Data Source*. In Alain Thiart et al., *Doing Management Research: a comprehensive guide*. Sage Publications: Londres, 2001.
- KIM, Y. Validation of psychometric research instruments: the case of information science. *Journal of the American Society for Information Science and Technology*, v.60 (6), 2009, p.1.178-1.191.
- LEWIS, B.; TEMPLETON, B.; BYRD, T. A methodology for construct development in MIS research. *European Journal of Information Systems*, v. 14, 2002, p. 388-400.
- LUQUE, T. *Investigación de Marketing: fundamentos*. Barcelona: Editorial Ariel S.A., 1997.
- MARTÍNEZ ARIAS, R. *Psicometria: teoría de los test psicológicos y educativos*. Madrid: Editorial Síntesis S.A., 1995.
- MARTÍN ARRIBA, M. C. Diseño y Validación de Cuestionarios. *Matronas Profesión*, v. 5 (7), 2004.
- MORALES VALLEJO, P. M.; SANZ, B. U.; BLANCO, A. *Cuadernos de Estadística*, Construcción de escalas de actitudes tipo Likert. Madrid: Editorial La Muralla S.A, 2003.
- MUÑIZ, J.; FIDALGO, A. M.; CUETO, E. G.; MATÍNEZ, R.; MORENO, R. *Cuadernos de Estadística*, Análise de los ítems. Madrid: Editorial La Muralla S.A, 2005.
- NAVAS ARA, M^a. J. *Métodos, diseños y técnicas de investigación psicológica*. Madrid: Universidad Nacional de Educación a Distancia (UNED), 2007.
- NUNNALLY, S. W. *Construction methods and management*. Englewood Cliffs: Prentice-Hall, 1987.
- NUNNALLY, J. C. *Psychometric Theory*. New York: McGraw-Hill, 1978.
- PASQUALI, L. Psicometria. *Revista Escola Enfermagem USP*, v. 43, p.992-999, 2009.
- PETERSON, R. A. A meta-analysis of Cronbach's coefficient alpha. *Journal of Consumer Research*, v. 21, 1994, p. 381-391.
- ROBERT, E. S. In defence of the survey method : an illustration from a study of user information satisfaction. *Accounting and Finance*, v. 39 (9), 1999, p. 53-77.
- SAATY, T. L. *Método de Análise Hierárquica*. Rio de Janeiro: Makrom Books, 2Ed., 1991.
- SÁNCHEZ, M.; SARABIA, F. J. Validez y fiabilidad de escalas. In SARABIA, F. J. (Coordinador), *Metodología para la investigación en marketing y dirección de empresas*, Madrid, Editorial Pirámide, 1999.
- SCHREISHEIM, C. A.; POWERS, K; SCANDURA, T. A.; GARDINER, C. C.; LANKAU, M. Improving construct measurement in management research: comments and a quantitative approach for assessing the theoretical content adequacy of paper-and-pencil survey-type instruments. *Journal of Management*, v. 19 (2), 1993, p. 385-417.
- SCHMITT, N. W. Editorial. *Journal of Applied Psychology*, v. 74, p. 843-845, 1989.
- SCHMITT, N. W.; KLIMOSKI, R. J. *Research methods in human resources management*. Cincinnati: South-Western Publishing, 1991.
- VAN DER STEDE, W. A.; YOUNG, S. M.; CHEN, C. X. Assessing the quality of evidence in empirical accounting research: the case of survey studies. *Accounting, organization and society*, v. 30 (7), 2005, p. 655-684.
- WIKMAN, A. Reliability, validity, and true values in surveys. *Social Indicators Research*, v. 78, p. 85-110.

YOUNG, G. M. *Survey research in management accounting: a critical assessment*. In *Research Methods in Accounting*. Vancouver : Ed. A. Richardson, 1996, p. 55-68.

YIN, R. K. *Case-study research : Design and methods*. Thousand Oaks, C. A.: Sage, 2003.