

Avaliação de uma Escala para Medir as Dimensões da Qualidade do Ensino Superior pelos Discentes: um Estudo em um Curso de Graduação de uma IES Pública

Lorena Flavia Fernandes Cunha (Universidade Federal de Uberlândia) lorena.cunha@hotmail.com

Marcelo Ruy (Universidade Federal de Uberlândia) marceloruy@ufu.br

Resumo:

O ensino superior é um serviço de extrema importância nas sociedades modernas. A qualidade é um dos fatores mais fundamentais para o serviço educacional. Assim, a criação de instrumentos de avaliação de qualidade educacional é uma questão chave que se apresenta às pesquisas acadêmicas. Há alguns instrumentos existentes para se medir a qualidade de uma Instituição de Ensino Superior (IES) do ponto de vista de seus discentes. Entretanto, tais instrumentos cobrem uma gama muito extensa de aspectos da qualidade dos serviços educacionais. Owlia e Aspinwall (1996, 1998) propuseram uma escala voltada para medir apenas os aspectos acadêmicos e de ensino. Contudo, este instrumento ainda necessita ser submetido a testes empíricos adicionais para ser considerado válido, sendo este o objetivo da presente pesquisa. Esta escala foi aplicada em um curso de graduação de uma universidade federal do interior de Minas Gerais. Por meio de Análise Fatorial Confirmatória concluiu-se que a escala proposta não se ajustou bem aos dados coletados. Uma série de modificações foram executadas na escala, de forma que o ajuste final fosse satisfatório. Porém, testes adicionais em uma nova amostra independente são necessários para validar estas alterações.

Palavras chave: Qualidade em Serviços; Serviços Educacionais; Avaliação da Qualidade; Análise Fatorial.

Assessment of a Scale to Measure Quality Dimensions of Higher Education by Students: a Study in an Undergraduate Course of a Public University

Abstract

Higher Education is an extremely important service in modern societies. Quality is one of the most fundamental factors in educational services. Therefore, the creation of scales to evaluate quality in education is a key question to academic research. There are some existent scales to measure quality in higher education institutions from students' point of view. However, they cover a very large number of quality aspects of the educational services. Owlia and Aspinwall (1996, 1998) proposed a scale to measure academics and teaching aspects only. Nonetheless, this instrument needs further empirical tests to be validated and this is the research objective. This scale was applied in an undergraduate course of a federal university from the interior of Minas Gerais state. By means of Confirmatory Factor Analysis, we concluded that the scale did not fit the sample data well. A series of modifications were proposed so that model fit was adequate. However, more tests in a new independent sample are necessary to validate these modifications.

Key-words: Service Quality; Educational Services; Quality Measurement; Factor Analysis.

1. Introdução

O setor de serviços vem ganhando espaço e importância já há algum tempo nas sociedades contemporâneas. Todos os dias, a maioria das pessoas tem algum tipo de contato com algum tipo de serviço. Pode-se mensurar a importância deste setor por sua participação na mão-de-obra ativa e pela participação na formação do Produto Interno Bruto (PIB) do Brasil.

A tendência moderna é tratar a produção de bens e serviços como operações. Assim, uma empresa pode oferecer ao mercado um pacote de produtos que pode ter ênfase em um ou em outro tipo de operação. Segundo Slack, Chambers e Johnston (2009), as principais características que distinguem as operações de serviços das de manufatura são intangibilidade, simultaneidade entre produção e consumo e presença e a participação do cliente no processo.

De acordo com Jain, Sinha e Sahney (2011) e Milan *et al.* (2015), a educação pode ser considerada um serviço. Para Lourenço e Knop (2011), o serviço educacional é constituído por ações intangíveis e dirigidas às mentes das pessoas. De entrega constante, ele se fundamenta em uma parceria entre prestadores e clientes. Apesar do contato pessoal contínuo, os serviços educacionais apresentam baixa customização devido ao seu aspecto de coletividade, ou seja, o serviço não é prestado a um consumidor de forma individual.

Segundo Eberle, Milan e Lazari (2010) e Mahapatra e Khan (2007), é usual considerar os estudantes como clientes primários dos serviços educacionais e seus pais, empregadores, o governo e a sociedade em geral como clientes secundários.

Galvão, Corrêa e Alves (2011) realçam que o ensino superior é de extrema importância para uma região, por ser indutor do desenvolvimento do conhecimento e da disseminação da informação. Além disso, é ele que irá preparar os futuros profissionais de alto nível que, direta ou indiretamente, são imprescindíveis para o crescimento econômico e para o desenvolvimento tecnológico e científico do país como um todo, principalmente na atual sociedade globalizada e baseada no conhecimento (CALVO-PORRAL; LÉVY-MANGIN; NOVO-CORTI, 2013).

Dada tal importância, Jain, Sinha e Sahney (2011) salientam que, assim como qualquer organização, instituições de ensino superior (IES) necessitam inovar continuamente, diversificar sua estrutura e encontrar novas maneiras de entregar seus serviços de forma mais eficaz a seus clientes. Green (2014) cita, ainda, que instituições com alta reputação e qualidade de ensino normalmente atraem os melhores estudantes, forçando-as a estar sempre atentas às questões da qualidade do serviço. Milan *et al.* (2015) chamam atenção que além da atração de novos alunos, uma alta qualidade aumenta a possibilidade de indicação à terceiros e de “recompra” de serviços por parte de egressos (p. ex., pós-graduação), ao passo que a falta de qualidade leva à desistência ou perda de alunos e enfraquecimento da imagem.

Assim, partindo-se do pressuposto que a qualidade é um dos fatores mais fundamentais para o serviço educacional, Lourenço e Knop (2011) explicam que o estabelecimento de instrumentos de avaliação de qualidade mostra-se, portanto, indispensável para qualquer IES comprometida com a melhoria de processos e com a satisfação de seus discentes.

Dutra, Oliveira e Gouveia (2002) alegam que existem diversas motivações para avaliar a qualidade de instituições de ensino em todos os níveis, que vão desde a melhor utilização dos recursos financeiros, até a oportunidade de subsidiar a tomada de decisões relativas às dimensões administrativa e pedagógica. Além disso, de acordo com os autores, a avaliação da qualidade do ensino superior pode ser utilizada como ferramenta da reforma universitária, fortalecendo um padrão de política educacional e contribuindo para o desenvolvimento da IES. Deste modo, um instrumento adequado de avaliação de qualidade do ensino superior deve também estar direcionado à identificação das falhas da instituição em perceber as reais

necessidades de seus discentes, o que contribui para fortalecer o foco da avaliação, evitando a ênfase em aspectos secundários do processo educacional.

Em suma, a identificação das dimensões que sinalizam a qualidade e a obtenção da excelência no ensino superior é uma questão chave que se apresenta às pesquisas acadêmicas (EBERLE; MILAN; LAZARI, 2010; JAIN; SINHA; SAHNEY, 2011).

Segundo Dutra, Oliveira e Gouveia (2002), há dois modelos de avaliação da qualidade em instituições de ensino: o interno e o externo. Lourenço e Knop (2011) enfatizam que embora processos de avaliação institucional externa sejam conduzidos periodicamente pelo Ministério da Educação, eles não têm como fonte de informação o corpo discente, exceto no que se refere ao conhecimento adquirido. O ENADE representa o único instrumento de avaliação com foco no corpo discente. Entretanto, voltado para o aprendizado, não contempla as condições de ensino e a competência docente. A aferição por amostragem e a análise agregada dos dados, que monitoram o sistema em grandes áreas, esclarecem pouco sobre as atividades de cada IES em particular. Diante desses fatos, a avaliação interna, empreendida pelas próprias IES, por meio da consulta aos graduandos, adquire uma importância fundamental.

Soma-se a isso a enorme expansão ocorrida no ensino superior no Brasil na última década, que pode ser explicada por um conjunto de fatores tais como o aumento do número de matrículas no ensino médio, a flexibilização dos critérios para abertura de novos cursos promovida pelo Ministério da Educação e o reconhecimento contemporâneo da relação entre qualificação e empregabilidade (BORGES; SANTOS; LEAL, 2014; LOURENÇO; KNOP, 2011; MILAN *et al.*, 2015).

Há alguns instrumentos que podem ser aplicados para se medir a qualidade de uma IES do ponto de vista de seus discentes, variando de genéricos, tais como o SERVQUAL (BORGES; SANTOS; LEAL, 2014), até os específicos, como o EduQUAL (MAHAPATRA; KHAN, 2007) e o HEdPERF (ABDULLAH, 2006).

Entretanto, os dois últimos instrumentos citados cobrem uma gama muito extensa de dimensões da qualidade dos serviços educacionais (por exemplo, refeitório, serviços de saúde) e que estão fora do controle dos docentes do curso e de seus gestores (por exemplo, coordenador do curso). Um instrumento mais voltado aos aspectos acadêmicos e de ensino foi proposto inicialmente por Owlia e Aspinwall (1996) e posteriormente refinado por Owlia e Aspinwall (1998). Contudo, este instrumento ainda necessita ser submetido a alguns testes empíricos adicionais para ser considerado válido, de acordo com o paradigma de medição e validação de construtos de MacKenzie, Podsakoff e Podsakoff (2011), o que remete a discussão para a próxima seção, a dos objetivos.

2. Objetivos

Esta pesquisa tem como objetivo geral aplicar e avaliar empiricamente a escala de medição da qualidade em serviços educacionais de Owlia e Aspinwall (1998) por meio de Análise Fatorial Confirmatória em um curso de graduação de uma universidade federal do interior de Minas Gerais. Seguindo o paradigma de medição e validação de construtos de MacKenzie, Podsakoff e Podsakoff (2011), os objetivos específicos necessários para avaliar as características psicométricas de uma escala são: (a) avaliar a qualidade do ajuste do modelo de medição; (b) avaliar a validade e confiabilidade individual de cada indicador; (c) avaliar a validade e a confiabilidade do conjunto de indicadores no nível do construto; (d) eliminar indicadores problemáticos, caso seja necessário.

3. Referencial Teórico

3.1. Qualidade em Serviços

A literatura sobre gestão de negócios permite-nos perceber a importância da gestão de serviços. Dessa forma, a gestão de serviços apresenta diferentes características dependendo do cenário e, conseqüentemente deve ser analisada singularmente em cada ocasião.

De acordo com Severo (2006), o serviço ao cliente está relacionado com um processo que acontece entre um comprador, um vendedor e um terceiro (fornecedor), cujo objetivo é proporcionar benefícios significativos em termos de valor agregado à cadeia de suprimentos a um custo compensador.

A caracterização de um serviço de qualidade não é realizada apenas com a análise do produto final do serviço, pois a qualidade envolve diversos elementos que afetam todo processo, ou seja, inicia-se na solicitação do serviço até a entrega do mesmo.

A satisfação dos clientes com a qualidade dos serviços prestados é resultante da comparação entre às expectativas do serviço desejado com a percepção destes, donde consiste em sensações de prazeres ou desapontamentos, por exemplo.

É evidente que os desejos e as exigências dos clientes passam por constantes modificações. Sendo assim, é de suma importância que os gestores e os fornecedores de tais serviços monitorem constantemente os processos de prestação de serviços e os índices de satisfação de seus clientes para assegurar a oferta de serviços de qualidade e a fidelidade dos seus clientes.

Os modelos de avaliação podem ser associados à duas principais abordagens para mensuração da qualidade em serviços: SERVQUAL e SERVPERF. De acordo com Parasuraman, Zeithaml e Berry (1985), o modelo SERVQUAL possui o objetivo de mensurar a qualidade de serviços através da análise das dimensões confiabilidade, responsabilidade, segurança, empatia e tangibilidade. Estas dimensões são agrupadas em um questionário, onde se compara as expectativas do cliente com a sua percepção sobre o nível do serviço que está sendo ofertado.

Segundo Salomi, Miguel e Abackerli (2005), o modelo SERVPERF foi desenvolvido baseado somente na percepção de desempenho dos serviços. Os autores ressaltam que a qualidade é conceituada mais como uma atitude do cliente com relação às dimensões da qualidade e que não deve ser medida por meio das diferenças entre expectativa e desempenho, mas sim como uma percepção de desempenho.

Os modelos citados anteriormente, SERVQUAL e SERVPERF, são métodos de avaliação da qualidade desenvolvidos para qualquer tipo de organização de serviços.

3.2. Avaliação da Qualidade em Serviços Educacionais

As avaliações nos sistemas de ensino superior são bastante importantes, pois mensuram: a aprendizagem dos estudantes; o desempenho dos docentes, dos cursos, dos currículos e das políticas educacionais; e a eficiência da alocação dos recursos. Ou seja, fazem avaliações com base nas perspectivas políticas, pedagógicas e de apoio à gestão.

A avaliação e a qualidade são termos inerentemente relacionados quando utilizados na educação. De acordo com Burlamaqui (2008), a qualidade é o foco da avaliação, pois o processo avaliativo visa reconhecer a qualidade ou, no mínimo, seus indicadores.

As avaliações institucionais são processos complexos e não há um modelo ideal, único e pronto para ser utilizado. Assim, estudiosos da área se dedicam a montar um modelo no qual realizar a medição da qualidade de ensino, de modo mais simples, seria possível.

Para Balzan e Sobrinho (2000), a avaliação institucional visa o aperfeiçoamento da qualidade da educação, isto é, do ensino, da aprendizagem e da gestão institucional, com a finalidade de transformar a instituição atual em uma que seja comprometida com a aprendizagem de todos e com a transformação da sociedade.

Em suma, a avaliação é considerada mediadora de um processo de tomada de consciência, individual e coletiva, que leva a instituição a uma autocrítica e (re)define seus caminhos com vistas à (trans)formação e melhora contínua da sua realidade associada à sua missão.

Sendo assim, existem diversos instrumentos que são habilitados para se realizar a mensuração da qualidade de uma Instituição de Ensino superior (IES) a partir do ponto de vista de seus discentes, variando de genéricos, tais como o SERVQUAL e, até os específicos, como o EduQUAL e o HEdPERF.

O modelo EduQUAL de Mahapatra e Khan (2007) contém 28 indicadores divididos em 5 dimensões determinadas por meio de Análise Fatorial Exploratória: resultados da aprendizagem, receptividade, instalações físicas, desenvolvimento pessoal do discente e questões acadêmicas.

O modelo HEdPERF de Abdullah (2006) é composto por 41 indicadores divididos em 6 fatores: aspectos não acadêmicos, aspectos acadêmicos, reputação, acesso, programa e compreensão das necessidades dos discentes. Esta escala foi refinada e validada pelo autor utilizando uma combinação de Análise Fatorial Exploratória e Confirmatória.

Apesar destes modelos específicos serem eficientes, os mesmos abrangem muitos aspectos da qualidade de serviços educacionais (p. ex., serviços de saúde, refeitórios). Owlia e Aspinwall (1996) criaram um instrumento voltado para os aspectos acadêmicos e de ensino. Posteriormente Owlia e Aspinwall (1998) refinaram o instrumento, obtendo 19 indicadores medindo 4 dimensões da qualidade, como indica a figura 1. Este modelo foi elaborado com base nas percepções dos clientes sobre as dimensões da qualidade, as suas classificações de importância e a sua avaliação global sobre o prestador de serviços. Entretanto, este instrumento ainda necessita ser submetido a testes empíricos adicionais para ser considerado válido.

Dimensão	Indicadores
Recursos Acadêmicos (Fator 1)	y ₁ : Quantidade suficiente de equipamentos (laboratórios, sala de computadores, etc). y ₂ : Facilidade de acesso aos equipamentos. y ₃ : Grau de modernidade dos equipamentos. y ₄ : Facilidade de acesso a fontes de informação (livros, periódicos, softwares, etc.). y ₅ : Corpo docente em uma quantidade suficiente.
Competência (Fator 2)	y ₆ : Nível de conhecimento teórico relevante do corpo docente. y ₇ : Nível de conhecimento prático relevante do corpo docente. y ₈ : Grau de atualização do corpo docente em suas respectivas disciplinas. y ₉ : Didática do corpo docente.
Atitude (Fator 3)	y ₁₀ : Grau com que o corpo docente entende as necessidades acadêmicas dos estudantes. y ₁₁ : Boa vontade do corpo docente em ajudar e auxiliar os alunos. y ₁₂ : Disponibilidade do corpo docente para orientar e informar os estudantes y ₁₃ : Grau com que o corpo docente dispensa atenção individual aos estudantes.
Conteúdo (Fator 4)	y ₁₄ : Quantidade de disciplinas básicas. y ₁₅ : Quantidade de disciplinas específicas. y ₁₆ : Grau com que os estudantes desenvolvem habilidades de comunicação. y ₁₇ : Grau com que os estudantes aprendem trabalhar em equipe. y ₁₈ : Relevância das ementas e do currículo para o futuro emprego do aluno. y ₁₉ : Interdisciplinaridade dos conteúdos das disciplinas.

Figura 1: Dimensões da Qualidade na Educação Superior. *Fonte: Owlia e Aspinwall (1998), p. 517.*

3.3. Análise Fatorial Confirmatória

De acordo com Katsikatsou *et al.* (2012), a Análise Fatorial (AF) é frequentemente empregada na medição e relacionamento de construtos não observáveis (p. ex. qualidade, satisfação, atitude). A ideia principal é que existem variáveis latentes (ou fatores) que explicam as dependências entre variáveis observadas (indicadores), de forma que se os fatores forem mantidos fixos, os indicadores serão independentes. Do ponto de vista teórico faz-se a diferenciação entre as Análise Fatorial Exploratória (AFE) e Confirmatória (AFC).

Segundo Brown (2015), a AFC é um tipo de modelagem de equações estruturais que trata especificamente de modelos de medição. Uma característica fundamental da AFC é sua natureza voltada ao teste de hipótese. Ao contrário da AFE, é necessário pré-especificar todos os aspectos do modelo. Assim, baseado em teorias anteriores, o pesquisador especifica a quantidade de fatores e fixa quais indicadores se relacionam com cada fator. A principal aplicação da AFC é no processo de construção e validação de escalas de medição de construtos.

Brown (2015) explica que na AFC clássica (indicadores contínuos), as variáveis medidas y são escritas como uma combinação linear de sua variância única ε com os seus fatores comuns η ponderados por suas cargas fatoriais Λ , de forma que $y = \Lambda\eta + \varepsilon$. Os fatores podem ser relacionados, com matriz de covariância Ψ , os erros são independentes dos fatores e possuem uma matriz de covariância $\Theta\varepsilon$. Como resultado, a matriz de covariância $\Sigma(\zeta)$ dos y **implícada** por este modelo é dada por $\Sigma(\zeta) = \Lambda\Psi\Lambda' + \Theta\varepsilon$.

Ao se coletar uma amostra da população, os elementos de ζ são os parâmetros a serem estimados das matrizes Λ , Ψ e $\Theta\varepsilon$. Estes valores formam a matriz reproduzida pelo modelo ajustado $\Sigma(z)$, que é uma estimativa de $\Sigma(\zeta)$. A matriz de covariância S entre os y **calculada** a partir dos dados amostrais é um estimador da matriz de covariância da população Σ . Assim, se o modelo fatorial proposto pelo pesquisador estiver correto, $H_0: \Sigma(\zeta) = \Sigma$. Portanto, se a discrepância entre S e $\Sigma(z)$ for maior do que o esperado apenas por conta de variabilidade amostral, a hipótese nula anterior é rejeitada e o modelo fatorial como postulado não é adequado. Verificadas algumas hipóteses (p. ex. normalidade multivariada dos indicadores), é possível definir uma estatística que segue uma distribuição χ^2 , que é a base do teste de hipótese de ajuste global da ACF.

Entretanto, os métodos e equações da análise fatorial clássica explicados anteriormente são válidos apenas quando a variável resposta y é contínua e medida pelo menos em uma escala intervalar (JÖRESKOG E MOUSTAKI, 2001). Para dados ordinais, métodos de estimação específicos tem que ser empregados. De acordo com Katsikatsou *et al.* (2012), há duas abordagens básicas para a análise fatorial de dados ordinais, cada qual com algumas variantes internas: a abordagem da Variável Resposta Subjacente (VRS) e a abordagem da Teoria de Resposta ao Item (TRI).

Na abordagem VRS, os indicadores são considerados como manifestações de variáveis contínuas subjacentes, parcialmente observadas por meio de seus correspondentes ordinais. Isto é, os indicadores são contínuos, mas os nossos instrumentos não são avançados o suficiente para medi-los e, por isso, os discretizam. A conexão entre a variável ordinal observada y e a variável contínua subjacente y^* é obtida por meio de pontos de corte τ (thresholds), de forma que: $y = c \Leftrightarrow \tau_{c-1} < y^* < \tau_c$, onde c representa a categoria ordinal da resposta e os pontos de corte são ordenados de forma que $-\infty = \tau_0 < \tau_1 < \dots < \tau_{k-1} < \tau_k = +\infty$. Para variáveis ordinais com k categorias, haverá a necessidade de $k - 1$ pontos de corte.

Jöreskog e Moustaki (2001) explicam que uma vez que apenas informações ordinais estão disponíveis a respeito de y^* , o modelo não é identificado. Uma solução é fixar y^* como uma variável aleatória normal com média 0 e variância 1. Nesta parametrização, os λ e τ serão

padronizados. Feito isso, as variáveis subjacentes são relacionadas com os fatores por meio do modelo de ACF clássica delineado anteriormente. Ou seja, este método é semelhante ao caso de indicadores contínuos, exceto pela substituição de y por y^* e pela adição dos τ .

Entretanto, a estimação dos parâmetros é executada de maneira distinta. Segundo Forero e Maydeu-Olivares (2009) o procedimento de estimação mais utilizado se dá em 3 estágios. No primeiro, os τ são estimados para cada variável separadamente, utilizando apenas as frequências das categorias de respostas de y . No segundo estágio, fixados os τ , correlações policóricas (correlações entre pares de variáveis subjacentes y^*) são estimadas separadamente, utilizando a distribuição conjunta de frequências de pares de variáveis y . Na última etapa, os demais parâmetros são estimados por meio de mínimos quadrados ponderados. Diferentes matrizes de ponderação podem ser utilizadas, sendo a escolha mais comum uma matriz diagonal. Normalmente a estatística de teste e os erros-padrão das estimativas são corrigidas, gerando erros-padrão e uma estatística de teste denominados robustos (daí o nome do método, *robust diagonally weighted least squares* – RDWLS). Este método é implementado no software R por meio do pacote *lavaan* de Rosseel (2012).

Forero e Maydeu-Olivares (2009) apontam que a abordagem da TRI modela a probabilidade de cada um dos padrões possíveis de resposta que podem ser observados. Esta probabilidade depende: (a) da probabilidade condicional de se endossar uma categoria de resposta dados os fatores latentes; (b) da distribuição dos fatores latentes (usualmente uma normal multivariada padronizada). Para dados ordinais com 3 ou mais categorias, o modelo mais utilizado é o Modelo de Resposta Gradual de Samejima. Em sua versão multidimensional, a probabilidade de um indivíduo escolher a categoria c do indicador y_i condicionada no vetor de fatores latentes η é: $Pr(y_i = c | \eta) = \Phi(\alpha_{i,c-1} + \beta_i' \eta) - \Phi(\alpha_{i,c} + \beta_i' \eta)$, onde Φ indica a função de distribuição acumulada da curva normal padrão, $\alpha_{i,c-1}$ e $\alpha_{i,c}$ são os dois pontos de corte consecutivos e não padronizados que determinam a categoria c do indicador y_i e β_i' é o vetor (transposto) que contém as cargas fatoriais não padronizadas relacionando o indicador y_i com os fatores latentes. Ambas as abordagens geram parâmetros equivalentes, no sentido que é possível converter τ em α e λ em β e vice-versa, conforme fórmula apresentada em Forero e Maydeu-Olivares (2009, p.278).

A estimação de todos os parâmetros é feita em uma só etapa utilizando o método de máxima verossimilhança marginal (ou de informação plena – *full information maximum likelihood*, FIML). A abordagem multidimensional da TRI é implementada no software R por meio do pacote *mirt* de Chalmers (2012).

Simulações empreendidas por Forero e Maydeu-Olivares (2009) comparando as abordagens VRS e TRI concluíram que na maioria das condições experimentais ambos os métodos geraram estimativas de parâmetros similares, com uma pequena vantagem para a TRI em amostras pequenas. Entretanto, a VRS na sua variante RDWLS tem uma vantagem fundamental em se tratando de validação de modelos: esta abordagem oferece diversas medidas de ajuste, ao passo que a TRI está apenas começando a incorporar este tema.

4. Metodologia

Foi realizado um levantamento com base em um corte transversal com variáveis quantitativas. A população de interesse foram os discentes do 2º ao último período de um curso de graduação de uma universidade federal do interior de Minas Gerais. Os discentes do 1º período não foram incluídos pois (supõe-se que) ainda não tiveram contato suficiente com o curso de forma a avaliar suas características.

O procedimento para a coleta de dados foi o questionário estruturado (GIL, 1999), composto

por 19 questões escritas, fechadas, de múltipla escolha e que foi auto aplicado on-line. Cada questão avaliava o desempenho do curso com relação aos indicadores da figura 1 utilizando uma escala ordinal de 5 pontos (Péssimo, Ruim, Regular, Bom e Excelente). Foi criada uma página na internet e o *link* foi enviado aos discentes dos períodos em questão. Dessa maneira, o método de seleção da amostra seguiu uma amostragem intencional (GIL, 1999), pois participaram os discentes que se auto selecionaram.

Na análise dos dados, foi utilizado o software R versão 3.4.0 (R CORE TEAM, 2017) juntamente com os pacotes *lavaan* versão 0.5-23.1097 (ROSSEEL, 2012) e *mirt* versão 1.24 (CHALMERS, 2012).

5. Resultados

Responderam o questionário 111 discentes de um universo de 245 (taxa de resposta de 45,3%). Desse total, 71 foram do sexo feminino (64,0%). Com relação ao ano de ingresso no curso, a amostra foi razoavelmente uniforme (mínimo 19 e máximo 27).

Para Brown (2015), na ACF o primeiro passo consiste em se determinar o ajuste global do modelo por meio do valor-p da estatística de teste χ^2 . As estimativas dos parâmetros só fazem sentido em soluções que apresentam ajuste adequado. Uma série de índices de ajuste aproximado adicionais também podem ser úteis para auxiliar na decisão. Devido ao fato de o método RDWLS possuir o teste χ^2 (robusto) e os índices, ele será a principal ferramenta desta etapa. Após o ajuste ser considerado adequado, será utilizado o método da TRI na estimação dos parâmetros, uma vez que seus resultados são melhores para pequenas amostras.

Como regra geral, um modelo se ajusta bem aos dados se o valor-p do teste χ^2 exceder 0,05 (devido à natureza de H_0 , ver seção 3.3) e se os seguintes índices de ajuste aproximado estiverem nas faixas indicadas: SRMR < 0,08; RMSEA < 0,06; TLI > 0,95; e CFI > 0,95.

Ao aplicar o método RDWLS aos dados amostrais da presente pesquisa, rejeita-se a hipótese de que a estrutura fatorial proposta se ajusta bem aos dados ($\chi^2 = 246,257$; g.l. = 146; valor-p = 0,000). Os índices de ajuste aproximado também foram ruins (SRMR = 0,087; RMSEA = 0,079; TLI = 0,940; e CFI = 0,949). Inspeccionando as estimativas das cargas fatoriais, seus erros-padrão e os valores-p de seus testes de significância estatística aproximados baseados na distribuição normal ($H_0: \lambda = 0$; $H_1: \lambda \neq 0$) calculados pelo software, verificou-se que todas as cargas fatoriais eram significativamente diferentes de zero exceto a do indicador y_{15} (figura 1). Assim, excluímos y_{15} e refizemos a ACF. O ajuste piorou ligeiramente e a conclusão foi a mesma.

Dessa forma, conclui-se que escala no formato original proposto por Owlia e Aspinwall (1998) não é adequada. Segundo Brown (2015), a falta de ajuste pode ser atribuída ao número errado de fatores (normalmente a causa menos provável), ou mais provavelmente à indicadores problemáticos. Indicadores causam falta de ajuste em CFA devido a: não carregarem em nenhum fator, carregarem no fator errado ou carregarem em mais de um fator (carga cruzada). Em todos esses casos, a melhor solução é eliminar os indicadores problemáticos. Assim, tentaremos modificar o modelo proposto e verificar se é possível obter um ajuste adequado. Entretanto, a partir deste ponto, a análise deixa de ser confirmatória e passa a ser exploratória. Ou seja, os resultados obtidos devem ser validados em uma outra amostra independente. É bem relatado na literatura que modificações baseadas unicamente nos dados coletados podem levar a sobreajuste do modelo. A despeito disso, tais alterações podem ser úteis para guiar pesquisas futuras visando validar o modelo de Owlia e Aspinwall (1998) modificado por nós.

Nesta etapa de exclusão de indicadores e/ou de verificação de número de fatores, é comum se utilizar a AFE. Porém, a AFE tem uma limitação fundamental: não são calculados os erros-

padrão das estimativas e nem o teste de ajuste do modelo. Brown (2015) propõe uma estratégia interessante: executar uma análise fatorial com todas as características restrições de uma AFE, mas utilizando a estrutura conceitual e um software de AFC (AFE/C). Em linhas gerais, fixado a priori o número de fatores, determina-se um item âncora para cada fator (por meio da teoria ou de uma AFE tradicional). O item âncora é fixado para carregar apenas em seu fator. Todos os demais indicadores carregam livremente em todos os fatores (como na AFE) e os fatores podem ser correlacionados. Essa quantidade de restrições é suficiente para identificar o modelo, de forma que este pode ser estimado por meio de AFC. Como subproduto da análise tem-se: (a) os erros-padrão de todas as cargas fatoriais e isso permite identificar quais indicadores carregam no fator errado ou tem carga cruzada significativa; (b) o teste de ajuste global permite determinar se a quantidade de fatores fixadas a priori é uma representação adequada dos dados.

Adotando este procedimento, o número de fatores foi fixado em 4 (conjectura inicial do modelo) e por meio do pacote *mirt* executou-se uma AFE ordinal utilizando o método da TRI. Foram escolhidos como itens âncora os indicadores que atendessem os seguintes critérios na AFE: alta carga fatorial, alta comunalidade e baixa carga cruzada. Os indicadores selecionados para os 4 fatores da figura 1 foram, respectivamente: y_1 , y_6 , y_{13} e y_{16} . Executando a AFE/C no pacote *lavaan*, obteve-se um ajuste excelente ($\chi^2 = 91,513$; g.l. = 87; valor-p = 0,349; SRMR = 0,037; RMSEA = 0,022; TLI = 0,996; e CFI = 0,996).

Uma vez que a análise fatorial busca a parcimônia na representação das variáveis originais e o modelo com 4 fatores se ajustou bem aos dados, não faz sentido testar modelos com mais fatores. Por outro lado, é possível que uma solução com 3 fatores seja também bem-sucedida em reproduzir a associação entre os indicadores. Dessa forma, aplicou-se o procedimento descrito anteriormente fixando-se o número de fatores em 3. O ajuste de modelo piorou, principalmente no teste mais importante que é o de ajuste global ($\chi^2 = 149,036$; g.l. = 102; valor-p = 0,002; SRMR = 0,056; RMSEA = 0,065; TLI = 0,964; e CFI = 0,976), corroborando a hipótese de 4 fatores. Logo, a falta de ajuste na ACF original se deve a problemas nos indicadores.

Dessa forma, passou-se a verificar o resultado da AFE/C com 4 fatores em busca de quais seriam os indicadores que deveriam ser eliminados. Idealmente, caso a AFC original tivesse se ajustado bem aos dados, o padrão observado na AFE/C para cada indicador seria o seguinte. A carga fatorial do indicador no seu fator correto seria elevada (em módulo) e se rejeitaria a hipótese nula $H_0: \lambda = 0$ (o valor-p deste teste seria pequeno, por exemplo, menor que 0,01). Por outro lado, as cargas fatoriais deste mesmo indicador nos outros 3 fatores errados seriam próximas de zero e se falharia em rejeitar essa mesma H_0 (o valor-p deste teste seria alto, por exemplo, maior que 0,05). A exceção é o item âncora que tem suas cargas cruzadas fixada em zero, logo ele só apresentaria o primeiro dos padrões descritos.

Os seguintes indicadores não seguiram esse padrão e foram excluídos: (i) y_4 apresentou uma carga fatorial significativa, porém baixa, em seu fator e no fator 4 (cargas fatoriais padronizadas respectivamente de 0,294 e 0,239); (ii) y_5 carregou menos no seu fator que no fator 2 (0,231 e 0,542); (iii) y_{10} carregou menos no seu fator que no fator 2 (0,338 e 0,423); (iv) y_{14} carregou mais no fator 1 que em seu próprio fator (0,376 e 0,352); (v) y_{19} apresentou uma carga fatorial significativa, porém baixa, em seu fator e nos fatores 2 e 3 (0,332; 0,277; 0,260).

Após essas exclusões, a AFC foi executada novamente pelo método RDWLS. O ajuste foi adequado ($\chi^2 = 70,215$; g.l. = 59; valor-p = 0,151; SRMR = 0,057; RMSEA = 0,042; TLI = 0,990; e CFI = 0,993). Todas as cargas fatoriais foram significativamente diferentes de zero. Assim, obtido um bom ajuste, iremos estimar o modelo novamente, agora utilizando o método da TRI implementado no pacote *mirt*, uma vez que este método é ligeiramente mais preciso em

amostras pequenas. A figura 2 mostra as estimativas obtidas para as cargas fatoriais e para as correlações entre os fatores. Também são apresentadas a variância extraída média dos fatores e as confiabilidades compostas para os grupos de indicadores.

A variância extraída média foi calculada por meio fórmula: $\sum(\lambda^2)/n$, onde n é a quantidade de indicadores do fator. Por exemplo, para o fator 1: $(0,967^2 + 0,848^2 + 0,615^2) / 3 = 0,677$. Já a confiabilidade composta não pode ser computada pelas fórmulas utilizadas para indicadores contínuos (coeficiente alfa de Cronbach ou coeficiente ρ de Raykov). O correto é considerar a natureza ordinal dos indicadores. Para tanto, foi utilizada a fórmula (21) de Green e Yang (2009, p. 160).

		Fator 1	Fator 2	Fator 3	Fator 4	Confiabilidade Composta
Correlações	Fator 1	1,000				
	Fator 2	0,506	1,000			
	Fator 3	0,387	0,709	1,000		
	Fator 4	0,517	0,678	0,589	1,000	
Cargas Fatoriais Padronizadas	y₁	0,967				0,816
	y₂	0,848				
	y₃	0,615				
	y₆		0,881			0,879
	y₇		0,884			
	y₈		0,855			
	y₉		0,728			
	y₁₁			0,855		0,871
	y₁₂			0,856		
	y₁₃			0,899		
	y₁₆				0,775	0,692
	y₁₇				0,582	
	y₁₈				0,709	
Variância Extraída Média		0,677	0,705	0,758	0,481	

Figura 2 – Estimativas dos Parâmetros do Modelo Fatorial

A seguir, são discutidas as conclusões que podem ser extraídas da figura 2. Esta discussão tem por base as recomendações de Hair *et al.* (2014).

Para se calcular a confiabilidade de cada indicador individualmente, eleva-se a carga fatorial padronizada ao quadrado. Recomenda-se uma confiabilidade mínima de 0,50. Apenas os indicadores y_3 e y_{17} não atendem a esse critério. Com relação à confiabilidade dos construtos, recomenda-se que a confiabilidade composta seja no mínimo 0,70. Pela figura 2, verifica-se que os fatores de 1 a 3 atendem ao critério e o fator 4 fica ligeiramente abaixo, devido ao indicador y_{17} .

O mesmo ocorre com a variância extraída média, cujo mínimo recomendado é de 0,50, ficando o fator 4 um pouco abaixo. Esta medida reflete a quantidade de variabilidade dos indicadores explicada pelo fator latente e é uma medida de validade convergente, ou seja, indicadores de um fator particular devem convergir e dividir uma alta proporção de variância conjunta. No caso do fator 4, isto também se deve ao indicador y_{17} .

Por fim, também é recomendado que a variância extraída média de cada fator exceda o quadrado

da maior correlação que este fator está envolvido. Por exemplo, a maior correlação que o fator 4 está envolvida é com o fator 2, que vale 0,678 (figura 2). Este valor ao quadrado ($0,678^2 = 0,460$) é inferior à variância extraída média do fator 4 (0,481). Este critério é atendido por todos os fatores e diz respeito à validade discriminante, ou seja, a extensão que um fator é realmente distinto dos demais.

6. Considerações Finais

Esta pesquisa teve como objetivo avaliar empiricamente as características psicométricas da escala de medição da qualidade em serviços educacionais de Owlia e Aspinwall (1998). Por meio de Análise Fatorial Confirmatória concluiu-se que a escala proposta não se ajustou bem aos dados coletados. Uma série de modificações foram executadas, de forma que o ajuste final fosse satisfatório. Porém, testes adicionais em uma nova amostra independente são necessários para validar estas modificações, uma vez que ajustes guiados exclusivamente pelos dados amostrais podem levar a sobreajuste do modelo.

Mesmo o resultado final não tendo sido a validação da escala, a pesquisa deixa como contribuição uma escala refinada que pode vir a ser confirmada futuramente. De acordo com o paradigma de medição e validação de construtos de MacKenzie, Podsakoff e Podsakoff (2011), este processo envolve muitos passos e normalmente é iterativo, com vários loops e feedbacks, sendo virtualmente impossível umas poucas pesquisas completarem todo o ciclo, desde a etapa inicial de conceituação do construto até a validação e normatização da escala. Assim, convidamos outros pesquisadores a se reunirem a essa tarefa, de forma a avançar o conhecimento em um tema de grande importância acadêmica e prática.

Referências

- ABDULLAH, F.** *The development of HEDPERF: a new measuring instrument of service quality for the higher education sector*. International Journal of Consumer Studies, v. 30, n. 6, p. 569-58, 2006.
- BALZAN, N. C.; SOBRINHO, J. D.** *Avaliação Institucional: teorias e experiências*. 2. Ed. São Paulo: Cortez, 2000.
- BORGES, L. F. M.; SANTOS, C. K. S.; LEAL, E. A.** *Quality in educational service: expectations versus performance in the accounting undergraduate course*. European Scientific Journal, v. 10, n.1, 2014.
- BROWN, A. T.** *Confirmatory factor analysis for applied research*. 2. Ed. New York: The Guilford Press, 2015.
- BURLAMAQUI, M. G. B.** *Avaliação e Qualidade na Educação Superior: tendências na literatura e algumas implicações para o sistema de avaliação brasileiro*. Estudos em Avaliação Educacional, v. 19, n. 39, 2008.
- CALVO-PORRAL, C.; LÉVY-MANGIN, J.-P. NOVO-CORTI, I.** *Perceived quality in higher education: an empirical study*. Marketing Intelligence e Planning, v. 31, n. 6, p. 60-619, 2013.
- CHALMERS, R. P.** *Mirt: A multidimensional Item Response Theory package for the R environment*. Journal of Statistical Software, v. 48, n. 6, p. 1-29, may, 2012.
- DUTRA, H. F. O; OLIVEIRA, P. A. S.; GOUVEIA, T. B.** *Avaliando a qualidade de serviço numa instituição de ensino superior*. In: Encontro Nacional da Associação Nacional dos Programas de Pós-Graduação e Pesquisa em Administração, XXVI, 2002, Salvador. Anais... Salvador: Anpad, 2002.
- EBERLE, L.; MILAN, G. S.; LAZZARI, F.** *Identificação das dimensões da qualidade em serviços: um estudo aplicado em uma Instituição de Ensino Superior*. RAE-eletrônica, v. 9, n. 2, jul/dez. 2010.
- FORERO, C. G.; MAYDEU-OLIVARES, A.** *Estimation of IRT Graded Response Models: limited versus full information methods*. Psychological Methods, v. 14, n. 3, p. 275-299, 2009.
- GALVÃO, H. M.; CORRÊA, H. L.; ALVES, J. L.** *Modelo de avaliação de desempenho global para Instituição de Ensino Superior*. Revista de Administração da UFSM, v.4, n. 3, p. 425-441, set/dez, 2011.
- GIL, A.C.** *Métodos e Técnicas de Pesquisa Social*. 5. ed. São Paulo: Atlas, 1999.

- GREEN, P.** *Measuring service quality in higher education: a South African case study.* Journal of International Education Research, v. 10, n. 2, 2014.
- GREEN, S. B.; YANG, Y.** *Reliability of summed item scores using structural equation modeling: an alternative to coefficient alpha.* Psychometrika, v. 74, n. 1, p. 155–167, 2009.
- HAIR, J. F. Jr.; HULT, G. T.; RINGLE, C. M.; SARSTEDT, M.** *A primer on partial least squares structural equation modeling (PLS-SEM).* Thousand Oaks: Sage, 2014.
- JAIN, R.; SINHA, G.; SAHNEY, S.** *Conceptualizing service quality in higher education.* Asian Journal on Quality, v. 12, n. 3 p. 296-314, 2011.
- JÖRESKOG K. G.; MOUSTAKI, I.** *Factor analysis of ordinal variables: a comparison of three approaches.* Multivariate Behavioral Research, v. 36, n. 3, p. 347-387, 2001.
- KATSIKATSOU M.; MOUSTAKI, I.; YANG-WALLENTIN, F.; JÖRESKOG, K. G.** *Pairwise likelihood estimation for factor analysis models with ordinal data.* Computational Statistics and Data Analysis, v. 56, p. 4243-4258, 2012.
- LOURENÇO, C. D. S.; KNOP, M. F. T.** *Ensino superior em Administração e percepção da qualidade de serviços: uma aplicação da escala SERVQUAL.* Revista Brasileira de Gestão de Negócios, v. 13, n. 39, p. 219-233, 2011.
- MACKENZIE, S. B.; PODSAKOFF, P. M.; PODSAKOFF, N. P.** *Construct measurement and validation procedures in MIS e Behavioral Research: integrating new and existing techniques.* MIS Quarterly, v. 35, n. 2, p. 293-334, 2011.
- MAHAPATRA, S. S.; KHAN, M. S.** *A framework for analysing quality in education settings.* European Journal of Engineering Education, v. 32, n. 2, p. 205-217, 2007.
- MILAN, G. S.; EBERLE, L.; CORSO, A.; De TONI, D.** *A qualidade em serviços e a satisfação de clientes: comparação entre a percepção de alunos de graduação e de pós-graduação de uma IES.* Revista de Administração de Santa Maria, v. 8, n. 3, p. 415-437, jul/set, 2015.
- OWLIA, M. S.; ASPINWALL, E. M.** *A framework for the dimensions of quality in higher education.* Quality Assurance in Education, v. 4, n. 2, p. 12-20, 1996.
- OWLIA, M. S.; ASPINWALL, E. M.** *A framework for measuring quality in engineering education.* Total Quality Management, v. 9, n. 6, p. 501-518, 1998.
- PARASURAMAN, A.; ZEITHAML V. A.; BERRY L. L.** *A conceptual model of services quality and its implication for future research.* Journal of Marketing, v. 49, n. 4, p. 41-50, 1985.
- R DEVELOPMENT CORE TEAM.** *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing: Vienna, Austria. ISBN 3-900051-07-0. Disponível em <<http://www.R-project.org>>, 2017.
- ROSSEEL, Y.** *Lavaan: an R package for structural equation modeling.* Journal of Statistical Software, v. 48, n. 2, 2012.
- SALOMI, G, G, E; MIGUEL, P, A, C; ABACKERLI, A, J.** *SERVQUAL X SERVPERF: comparação entre instrumentos para avaliação da qualidade de serviços internos.* Gestão e Produção, v.12, n.2, 2005.
- SEVERO, F, J.** *Administração de Logística Integrada: Materiais, PCP e Marketing.* 2.ed.Rio de Janeiro: E-papers, 2006.
- SLACK, N.; CHAMBERS, S.; JOHNSTON, R.** *Administração da produção.* 3. ed. São Paulo: Atlas, 2009.