

Avaliação da formação superior pelo discente: proposta de um instrumento

Evaluation of the Higher Education by the student: an instrument proposal

Francisco José Costa¹ - ¹Universidade Federal da Paraíba | Departamento de Administração | João Pessoa | PB | Brasil. Contato: franze.mq@gmail.com. ORCID: <http://orcid.org/0000-0002-4090-5619>

José Jorge Lima Dias Júnior² - ²Universidade Federal da Paraíba | Departamento de Administração | João Pessoa | PB | Brasil. Contato: jorge@dcx.ufpb.br. ORCID: <http://orcid.org/0000-0003-4381-1930>

Resumo: Este artigo apresenta uma proposta de instrumento de ‘avaliação pelo discente’ enquanto componente da avaliação institucional de organizações de educação superior. Desenvolvemos um relato do processo de construção do instrumento estruturado em três fases: (1) análise da experiência prévia de uma instituição em avaliação do docente pelo discente; (2) apresentação dos procedimentos de construção do novo instrumento, redimensionado para contemplar construtos relativos às dimensões de docentes, disciplinas, curso, além da autoavaliação do aluno; (3) avaliação do novo instrumento e fechamento da proposta. Nesse processo foram utilizados procedimentos quantitativos e qualitativos, segundo orientações da teoria psicométrica da mensuração. Os resultados das análises do novo instrumento demonstraram que as dimensões definidas foram mensuradas com evidências de validade e confiabilidade, e com resultados potencialmente úteis para docentes e lideranças de gestão acadêmica. O estudo, além de oferecer um instrumento com métricas válidas, pode servir de referência para outras instituições interessadas em aprimorar seu instrumento de avaliação pelo discente.

Palavras-chave: Avaliação. Instrumento. Ensino superior. Discente. Teoria da mensuração.

Abstract: This paper presents a proposal for an instrument for ‘assessment by the student’ as a component of the institutional evaluation of higher education organizations. We developed a report of the process of the instrument construction in three phases: (1) analysis of the previous experience of an institution regarding its professor’s evaluation by the students; (2) presentation of the development procedures for the new instrument, which included assessment of the dimensions relating to professors, disciplines, course, in addition to the student’s self-assessment; (3) evaluation of the new instrument and closing of the proposal. In this process, quantitative and qualitative procedures were used according to the guidelines of the psychometric theory of measurement. The results of the analysis of the new instrument showed that the defined dimensions were measured with evidence of validity and reliability, and with potentially useful results for professors and institutional managers and leaders. The study, in addition to offering an instrument with valid metrics, can serve as a reference for other institutions interested in improving their instrument for assessment by the student.

Keywords: Assessment. Instrument. Higher Education. Student. Measurement theory.

• Recebido em: 8 de setembro de 2018 • Aprovado em: 5 de maio de 2020

DOI: <http://dx.doi.org/10.1590/S1414-4077/S1414-40772020000200003>

Este é um artigo publicado em acesso aberto sob uma licença Creative Commons
<https://creativecommons.org/licenses/by-nc/4.0/>

1 Introdução

A avaliação das atividades acadêmicas é prática recorrente e está na própria essência das instituições de formação. Enquanto objeto de estudo, a diversidade de temas de interesse dos estudos de avaliação, em conjunto com os variados fundamentos teóricos e epistemológicos, está na base de um dos temas mais dinâmicos e, por vezes, mais polêmicos, da gestão de organizações de ensino superior. As questões teóricas e aplicadas da avaliação emergem quando falamos desde as avaliações de aprendizagem escolares, passando pelas provas de concursos públicos ou dos exames gerais (como o Exame Nacional do Ensino Médio), chegando até os testes psicológicos, as práticas de avaliação de políticas públicas, ou a análise de qualidade de serviços. Todas essas práticas estão no dia a dia das pessoas, sejam estudantes, docentes ou gestores, sejam os profissionais que estudam o tema e desenvolvem instrumentos.

Neste artigo nos posicionamos na condição de estudiosos ou propositores de um instrumento relativo à ‘avaliação institucional’ das instituições de educação superior, com foco específico no que chamaremos de ‘avaliações pelo discente’ para nos referir à dimensão de avaliação em que os discentes são convidados a se posicionar sobre os serviços de formação recebidos, como, por exemplo, sua satisfação com os docentes, com o currículo, além de sua autoavaliação etc. Avaliações assim são recorrentes, como, por exemplo nas avaliações do docente pelo discente ou na avaliação da satisfação do aluno com os serviços recebidos.

As práticas desse tipo de avaliação envolvem duas definições centrais e associadas: a primeira é relativa ao que será avaliado; e a segunda concerne aos instrumentos a serem utilizados. A complexidade dessas questões está no fato de que os processos de avaliação do ‘posicionamento do discente’ (digamos, sobre a qualidade das aulas ministradas), envolvem, quase sempre, uma aferição indireta, da modalidade do que a teoria psicométrica chama de ‘mensuração de construtos latentes’, ou seja, daquelas variáveis que, embora se suponha uma intensidade mensurável, tal medição não pode ser feita de forma direta, sendo realizada por manifestações desta intensidade de forma indireta (COSTA, 2011; DEVELLIS, 2016).

Esse tipo de aferição está associada a diversos indutores de ‘erro de mensuração’, sendo dois os principais: a condição de subjetividade envolvida (cada aluno tem uma noção própria do que seja o nível ‘bom’ de um determinado critério, de modo que, para um aluno, na escala de 0 a 10, bom seja 6 e, para outro, seja 8); e a posição de discente em relação a critérios que, eventualmente, tenham dificuldades de avaliar (por exemplo, há casos em que ‘atualização do

docente' é critério que o aluno é convidado a opinar, mas sobre o qual ele pode não ter informações ou discernimento para emitir a avaliação).

Com efeito, desenvolver instrumentos adequados é um desafio recorrente e que requer a devida análise técnica, acadêmica e com foco na aplicação. E esse é um desafio que tratamos e relatamos neste estudo, que tem como objetivo apresentar um instrumento de 'avaliação pelo discente' para instituições de educação superior, a partir de uma análise baseada no desempenho do instrumento proposto à luz da teoria da mensuração. Como referencial de teste e aplicação concreta, desenvolvemos nossa proposta a partir da experiência vivenciada na análise de resultados do esforço de mudança de instrumento de avaliação pelo discente da Universidade Federal da Paraíba (UFPB) entre os anos de 2016 e 2017. O instrumento produzido vem sendo aplicado aos discentes da instituição desde então e pelo menos até a data de publicação deste artigo. Em nosso entendimento, o instrumento desenvolvido e aqui apresentado poderá ser útil para uso em outras instituições ou, ao menos, poderá ser referência para elaboração e aprimoramento de instrumentos adaptados a cada organização. Além dessa contribuição de caráter mais prática, também buscamos apontar uma contribuição teórica a partir da análise dos desempenhos de estratégias de mensuração baseadas em múltiplos itens, como sugerem Elliott e Shin (2002), ou em único item, que entendemos ser a mais apropriada.

Para alcançar esse objetivo, além dessa introdução, temos, ordenadamente, a exposição de algumas considerações sobre a avaliação na educação superior e dos principais fundamentos teóricos e metodológicos do processo de mensuração de construtos latentes; o processo completo de elaboração e teste do instrumento proposto; e, finalizando, as considerações finais do trabalho. O instrumento proposto está no anexo do artigo.

2 Avaliação pelo discente na educação superior

Nesse item tecemos breves considerações teóricas sobre a avaliação na educação superior e sobre alguns aspectos técnicos do processo de mensuração, considerando que a avaliação envolve, necessariamente, um esforço de medição.

2.1 Avaliação na educação superior

No Brasil, as instituições de ensino superior devem estar alinhadas às diretrizes do SINAES (Sistema Nacional de Avaliação da Educação Superior), criado em 2004 sob a Lei nº 10.861, que tem como objetivo maior avaliar os aspectos relacionados ao ensino, à pesquisa e à extensão, além do corpo docente, da gestão institucional e da infraestrutura. Essas avaliações

podem ser internas, quando a instituição cria mecanismos próprios de avaliação, e externa, que são aquelas executadas geralmente pelo Ministério da Educação (MEC) para todos os cursos do ensino superior do país. Em relação a essa última, um dos mais conhecidos é o ENADE (Exame Nacional do Desempenho dos Estudantes), que avalia o desempenho dos alunos formandos, em geral a cada três anos em cada curso.

A avaliação externa, apesar de relevante, possui uma ênfase na homogeneização e na padronização de critérios avaliativos centrados na regulação do governo (FERREIRA; FREITAS, 2017). Por isso, há a recomendação no SINAES para que as instituições de ensino superior implementem uma política de avaliação interna, focada nas suas particularidades institucionais. Dessa forma, a avaliação pelo discente, em diferentes dimensões, torna-se uma ação relevante, pois considera a percepção dos principais atores que recebem os serviços educacionais da instituição. Além disso, diferentemente de uma avaliação externa como o ENADE, a avaliação interna pode ter uma periodicidade maior (semestral, por exemplo), permitindo subsidiar ações de gestão de forma mais efetiva, uma vez que há um processo de *feedback* mais breve.

Independentemente da visão de estado, a gestão pública, seja no campo educacional ou não, deve buscar a melhoria da qualidade dos seus processos organizacionais. E em qualquer ciclo de melhoria de processos há a premissa fundamental de que a medição é ferramenta de apoio avaliativo que deve estar alinhada aos objetivos e metas organizacionais. As discussões, as intervenções e o planejamento na gestão de cursos do ensino superior, tanto por parte dos agentes públicos quanto privados, devem ser subsidiados por indicadores que permitam diagnósticos dos cursos ofertados. Nesse sentido, parece clara a necessidade de as instituições de ensino superior avaliarem, em algum nível, a satisfação dos seus alunos, em diferentes dimensões, tais como corpo docente, infraestrutura, gestão do curso etc.

O diagnóstico por meio de indicadores pode, inclusive, ser uma ferramenta útil para lidar com duas das preocupações históricas e proeminentes no ensino superior, que são a evasão e a retenção dos alunos. Esses fenômenos têm atraído cada vez mais a atenção de pesquisadores e de gestores acadêmicos no intuito de compreender as causas e possíveis soluções. O que se percebe é que as causas são multifatoriais, incluindo, por exemplo, aspectos como a baixa qualidade do ensino e o baixo desempenho dos alunos (VELOSO; ALMEIDA, 2002); variáveis sociodemográficas, como gênero e idade; número de semestres cursados e notas (COSTA; BISPO; PEREIRA, 2018); o processo de transição do ensino médio para o ensino superior (DIAS; THEÓPHILO; LOPES, 2010); indecisão profissional e expectativas não atendidas em relação ao curso (SLHESSARENKO *et al.*, 2014). Portanto, a avaliação periódica pelo discente

deve considerar também aspectos que permitam que os gestores diagnostiquem e prevejam possíveis retenções e evasões.

Diante dessas considerações sobre a importância em se estabelecer um processo avaliativo periódico e específico para instituições dessa natureza, é necessário considerar aspectos teóricos envolvidos no desenvolvimento de medidas. A próxima subseção, portanto, trata desse assunto.

2.2 A construção de instrumentos de avaliação à luz da teoria da mensuração

Como indicado na introdução, propomos aqui um instrumento para analisar as avaliações dos discente de graduação. Nosso esforço é, portanto, desenvolver um instrumento com métricas para aferição de intensidades de variáveis específicas e bem delimitadas que atendam a interesses acadêmicos e de gestão. Na teoria da mensuração utilizada em ciências sociais e comportamentais (cf. NUNNALLY; BERNSTEIN, 1978; URBINA, 2009), essas variáveis são comumente chamadas ‘construtos’, que se referem às características de um objeto ou sujeito que variam e cujos estados são passíveis ou de classificação ou de quantificação. As métricas são em geral indicadas como ‘escalas de mensuração’, e se referem ao instrumento utilizado para classificação ou quantificação de um construto. Dado que nosso propósito é de desenvolver métricas, duas decisões precisam ser tomadas: a primeira concerne à definição do construto; e a segunda refere-se às definições operacionais do processo de desenvolvimento da escala.

Relativo ao construto, e se a referência de avaliação é um discente, diversos são os aspectos que podem ser definidos, desde que se suponha que o discente tenha condições de emitir posição. Trata-se, portanto, de uma definição de gestão, guiada pelo interesse institucional. Métricas comuns em que o discente costuma ser a origem/fonte da medidas são: a satisfação (com o curso, com o docente etc.), as predisposições à evasão ou à mudança de curso, o nível de apego ou de vínculo ao curso, as intenções de recomendar o curso etc.; genericamente, esses construtos, baseados em opiniões e posições das pessoas, são chamados atitudes (DEVELLIS, 2016).

Bastante recorrentes são as avaliações de satisfação do discente, seja com docentes, disciplinas, aspectos específicos da instituição (estrutura, currículo...), como foi o trabalho de Wojahn, Ramos e Carvalho (2018), por exemplo, que apresentou uma proposta de métrica específica para análise de satisfação de discentes. Geralmente, satisfação é medida ou de forma global (por um único item ou variável), em que se avalia em nível geral a satisfação de um indivíduo em relação a um produto ou serviço; ou a nível de atributo (com múltiplos itens ou

variáveis), em que se analisam diferentes dimensões de um produto ou serviço (SZYMANSKI; HENARD, 2001).

Referente às definições operacionais, temos variações de modelos e etapas, mas, considerando os diferentes autores, podemos apontar como decisões principais as seguintes (cf. ROSSITER, 2011; COSTA, 2011): 1 - identificação e especificação do construto, ou seja, o que deve ser mensurado; 2 - decisão quanto à forma de acesso, que envolve definir a forma (direta ou indireta), o suporte (impresso ou virtual), a forma (questionamento aplicação presencial, correio, telefone...), o enunciado dos itens, e a forma de verificação métrica (escala de concordância de Likert, escala binária...); 3 - implementação e teste preliminar; 4 - análise de validade e confiabilidade; 5 - desenvolvimento de normas (recomendações e formas de interpretação).

Embora todos os passos sejam relevantes, sem dúvidas o que mais interessa é o passo 4, que consiste em analisar a validade (ou seja, se a métrica de fato mensura o que pretende mensurar) e a confiabilidade (ou seja, supondo que seja inevitável erro de medição, se o nível de erro é pequeno o suficiente para garantir uma boa medição). As alternativas de avaliação para esses aspectos são as mais diversas, indo desde modelos mais qualitativos e exploratórios, até procedimentos sofisticados de análise de estatística (como análise fatorial, teoria da resposta do item...). Analisando as práticas recorrentes e a literatura, os autores da área de teoria da mensuração costumam seguir métodos mistos (qualitativos e quantitativos), ora enfatizando os procedimentos mais qualitativos, ora os mais quantitativos (COSTA, 2011).

Aqui seguimos próximos da proposta de Rossiter (2011), que sugere os passos acima indicados, mas destaca que o esforço de proposição de escalas deve enfatizar, muito mais que resultados estatísticos, a validade qualitativa, ou seja, aquela que se define a partir da discussão de especialistas as decisões sobre os itens ou variáveis, as formas de aferição e a potencial utilidade do resultado gerado. Por essa perspectiva, os resultados estatísticos não são indispensáveis e determinantes por si só da qualidade da métrica, embora possam trazer evidências importantes para essa análise. Seguimos esse encaminhamento metodológico na análise empírica realizada, cujos detalhe estão indicados no item seguinte.

3 Construção de um instrumento

Para a proposição do instrumento foram realizadas 3 etapas principais, sobre as quais discorreremos neste item. Na primeira etapa apresentamos o resultado da análise do instrumento pré-existente na organização analisada (Universidade Federal da Paraíba). Em seguida,

detalhamos uma série de procedimentos que foram realizados para a construção do novo instrumento. Na terceira etapa apresentamos uma avaliação da performance do instrumento proposto.

3.1 Análise do instrumento anterior

Como etapa preliminar de proposição do instrumento, entendemos que era necessário avaliar as práticas pré-existentes. No caso da instituição de referência neste trabalho, havia um instrumento que fora elaborado entre os anos de 2012 e 2013, e que foi utilizado por cinco semestres (de 2014.1 até 2016.1). O foco específico desse instrumento era a avaliação do docente pelo discente (ou seja, o instrumento não continha outros construtos que seriam de potencial avaliação pelo discente). Tal instrumento foi desenvolvido por uma comissão acadêmica, levando em conta dois elementos centrais: as indicações legais e teóricas da avaliação do docente pelo discente, e um exaustivo processo de debate com a comunidade acadêmica. Nesse último aspecto, representantes da comissão buscaram desenvolver reuniões em cada uma das 16 unidades acadêmicas da instituição, sempre com convites para participação enviado a todos os docentes.

- Critérios, coleta de dados e divulgação de resultados

Como construtos de referência da avaliação do docente, foram então definidos 10 itens e para cada um foi definida uma pequena definição que era exposta ao discente quando da aplicação. O Quadro 1 apresenta cada item e respectiva definição.

Cada um desses itens era avaliado utilizando uma escala de verificação de 11 pontos (0 a 10), com os seguintes indicativos de referência para os discentes: Excelente (equivalente a notas de 9,0 e 10,0); Bom (equivalente a notas de 7,0 a 8,9); Regular (equivalente a notas de 5,0 e 6,9); Insuficiente (equivalente a notas de 3,0 a 4,9); e Péssimo (equivalente a notas de 0,0 a 2,9). Para extração de uma medida de síntese da avaliação do docente pelos alunos do semestre, era obtida a média das notas atribuídas. Na divulgação, era ainda utilizada uma medida de ‘avaliação global’ do professor a partir da média aritmética das médias finais desses dez itens.

Quadro 1 – Itens e definição

ITEM	DEFINIÇÃO
Cumprimento do plano de curso	Correspondência entre os conteúdos apresentados e desenvolvidos pelo professor, na sala de aula, e os conteúdos programados e descritos no plano de curso
Relacionamento com a turma	O uso, pelo professor, da cordialidade, respeito, consideração e cortesia, no trato com os discentes
Assiduidade	A frequência com que o professor compareceu às aulas
Pontualidade	O cumprimento, pelo professor, dos horários estabelecidos para o início e término das aulas
Motivação do discente	As ações empreendidas, de forma geral, pelo professor para estimular a participação dos discentes nas aulas
Domínio dos conteúdos	O conhecimento teórico demonstrado pelo professor, em conjunto com os recursos didáticos que ele utilizou para promover a aprendizagem dos discentes
Clareza na exposição dos conteúdos	A habilidade demonstrada pelo professor em expor conteúdos e em esclarecer as dúvidas surgidas durante a exposição de um tema, explicando o mesmo conteúdo com diferentes raciocínios e com riqueza de detalhes, por exemplo
Atendimento individual em sala de aula	A sensibilidade do professor com relação às dificuldades individuais de aprendizagem, disponibilizando tempo, em sala de aula, para tirar dúvidas, rever conceitos e dar exemplos, sempre que solicitado;
Atendimento fora de sala de aula	O tempo dedicado pelo professor, fora dos horários de suas aulas (no ambiente de trabalho, por e-mail, entre outros), para, de forma individualizada ou em grupo, auxiliar os discentes na superação de eventuais dificuldades de aprendizagem
Utilização das referências bibliográficas	A efetiva utilização, pelo professor, durante o curso, dos livros, artigos e periódicos relacionados no plano de curso, incluindo as referências bibliográficas básica e complementar.

Fonte: Instrumento de avaliação docente pelo discente da PRG

O instrumento era aplicado sempre no período de matrícula online, quando o discente era condicionado a fazer primeiro a avaliação do semestre anterior antes de acessar o campo de matrícula do sistema de informação de gestão acadêmica da instituição. Dessa forma, era garantida uma boa aproximação do censo (a exceção ficava por conta dos discentes concluintes, que não avaliavam o semestre anterior por não ter que proceder mais a qualquer matrícula), e uma proximidade temporal que permitia uma avaliação mais consistente, devido ao fato de a matrícula ocorrer no máximo um mês após a finalização do semestre anterior, de modo que o discente não esqueceria facilmente as experiências relativas ao semestre avaliado. Na distribuição de resultados, a Pró-Reitoria de Graduação da instituição utilizou duas formas: primeiro, com a disponibilidade dos resultados na área docente do sistema de informação de gestão acadêmica; segundo, com a extração das notas e envio em um documento digital para o e-mail de cada docente. Em ambas as formas, o docente tinha acesso às suas médias globais e segmentadas por disciplinas, e no relatório digital eram ainda acrescentadas as médias dos demais docentes da unidade (departamento).

- Problemas e avaliação para mudança

Durante o uso desse instrumento e após os primeiros esforços de divulgação dos resultados, algumas reclamações e sugestões foram formalizadas e dirigidas à Pró-reitoria de Graduação, com questionamentos os mais diversos, indo desde a contestação dos critérios, passando pela suposição de incapacidade de avaliação pelos discentes de alguns dos critérios, e chegando até mesmo a hipóteses de condicionantes de resultados (por exemplo, professores que tendem a emitir maiores notas aos alunos seriam melhor avaliados) e indutores de erros no instrumento, nas variáveis etc.

Apesar dessas críticas, por decisão da Pró-reitoria, o instrumento seguiu sendo implementado segundo o procedimento acima indicado por ao menos seis semestres letivos, como forma de gerar uma série inicial e também conteúdo para avaliação psicométrica dos resultados. Em 2016 foi então instituída uma equipe técnica para avaliação e proposição de aperfeiçoamentos do instrumento. Esta equipe definiu duas etapas para diagnosticar a qualidade do instrumento de avaliação então vigente. A primeira foi buscar envolver a comunidade acadêmica da instituição para identificar problemas e gerar ideias de melhoria para um novo instrumento; e a segunda foi aplicar um conjunto de técnicas quantitativas para analisar as características das medidas.

Para a primeira ação, foi elaborado um questionário online que foi enviado a todos os docentes da instituição, solicitando opiniões e sugestões para mudança do instrumento que estava em uso. O questionário ficou disponível por 20 dias entre os meses de outubro e novembro de 2016, e foram recebidas ao todo sugestões de 70 docentes de 12 diferentes unidades acadêmicas. As contribuições foram analisadas de forma qualitativa e, em geral, as observações foram mais relacionadas à maior eficiência na divulgação e no uso dos dados (e.g. muitos docentes questionaram sobre o que o resultado poderiam implicar para os casos de docentes pior avaliados); à organização do modelo de apresentação (e.g. alguns docentes sugeriram a apresentação de médias separadas de alunos aprovados e reprovados); e à possibilidade de melhorar os critérios da avaliação. Em conjunto com as observações desta pesquisa, consideramos ainda os relatos e reclamações e sugestões que foram diretamente à em Pró-reitoria momentos anteriores. Dois aspectos se destacaram na orientação dos procedimentos seguintes:

- primeiro, as reclamações frequentes sobre a avaliação de itens como "Assiduidade" e "Pontualidade" (alguns docentes que nunca faltaram ou se atrasaram em um semestre recebiam nota de assiduidade menor que o máximo (10,0)). Levando em conta que a qualidade de uma métrica, em uma perspectiva de teoria da mensuração, está

relacionada a sua capacidade de definir valores que representem satisfatoriamente os objetos ou eventos reais, de fato temos nesses casos indicativos de que esses itens não estavam medindo de forma válida aquilo ao qual ele se propunha. Dessa forma, as observações dos docentes geraram a desconfiança de que as dimensões propostas no instrumento em uso não estavam sendo aferidas de forma adequada;

- segundo, foram formuladas reclamações de vários docentes de que as notas que eles atribuíam aos alunos impactavam de forma substantiva na avaliação recebida por eles, motivo pelo qual teriam potencialmente médias baixas não por sua qualidade docente, mas por seu rigor na avaliação. Essas reclamações estavam associadas à sugestão de separar as notas por aprovados de reprovados, pela suposição de os alunos reprovados atribuírem notas baixas aos docentes em retaliação pela reprovação sofrida.

O primeiro aspecto acima apresentado sugeriu à equipe de análise que a provável razão das avaliações inapropriadas (de itens como assiduidade e pontualidade) residiria no fato de os alunos não avaliarem o item em si, mas que cada item refletiria, na verdade, a avaliação geral que o aluno fazia do professor. Esse é, na verdade, o conceito de ‘mensuração refletiva’, que está na base de grande parte das métricas desenvolvida com métodos psicométricos. Essa forma de mensuração parte do pressuposto de que o construto a ser mensurado, por ser abstrato, deve ser acessado de forma indireta, a partir, por exemplo, da indicação do respondente de sua nota em um item numa escala especificada (no nosso caso, de 0 a 10), sob o entendimento de que a magnitude dos escores ‘refletem’ a magnitude da intensidade do construto latente (COSTA, 2011). Tivemos, por essas considerações, o desafio de, primeiro, verificar se há sentido em supor que os 10 itens do Quadro 1 seriam de fato refletivos em relação a um construto latente, de modo que os discentes, ao avaliarem os docentes, o fariam a partir de uma única dimensão geral, presumivelmente a satisfação do discente com o docente; segundo, tivemos que analisar as consequências dessa condição, se procedente, no nosso intento de melhorar o instrumento. Para fixarmos um referencial típico de verificação quantitativa, enunciamos esse entendimento na forma da proposição indicada a seguir:

- **Proposição 1:** as 10 dimensões associadas aos docentes que são avaliadas pelos discentes são indicadores de único construto latente e mantêm uma variação refletiva em relação à variação desse construto.

Relativo ao segundo aspecto, uma averiguação foi possível a partir da análise bivariada realizada considerando as médias dos docentes para os conjuntos de alunos de um período e as médias globais que o docente recebeu do conjunto de alunos nesse mesmo período. Também aqui, caso isso se verifique, é necessário analisar a repercussão no próprio processo e na finalidade de avaliação, posto que não é, de forma alguma, desejável que o docente seja avaliado

por outra razão que não seja seu desempenho acadêmico na condução das disciplinas. Também aqui, para procedermos à avaliação, enunciamos o entendimento na forma de proposição, como indicado abaixo:

- **Proposição 2:** as notas atribuídas aos alunos pelo docente influenciam na sua avaliação geral recebida.

Os resultados da avaliação dessas duas proposições tanto permitiram analisar a performance do instrumento anterior quanto sinalizaram os melhores ajustes a serem realizados no novo instrumento. Os detalhes e resultados serão exibidos a seguir.

- Preparação dos dados e análise das proposições

Na preparação da base de dados, optamos por utilizar os dois períodos de avaliação (semestres 2015.2 e 2016.1) imediatamente anteriores à mudança do instrumento. Nesses dados constavam, em cada semestre, aproximadamente 125 mil avaliações para cerca de 2250 professores ativos da instituição, ou seja, em torno de 55 avaliações por docente. Na base de dados constavam, além das avaliações de cada docente, dados sobre qual o curso, o departamento, a disciplina, a média final do aluno, e se houve aprovação ou reprovação. A partir dessa base, derivamos uma outra tendo como unidade de referência o professor e em que constasse o total de alunos que o avaliaram, a média das notas desses alunos, as médias dos escores recebidos em cada um dos 10 critérios e a média global do docente (média das médias dos critérios). Como estamos operando com valores de médias, e considerando que a média oscila bastante quando o número de observações é pequeno, excluimos da análise todos os docentes que tiveram menos de 10 avaliações em cada período e, após esse procedimento, cada período ficou com cerca de 2000 entradas de dados (2068 em 2015.2, e 1977 em 2016.1).

Para verificar a Proposição 1 foi aplicada a técnica de Análise Fatorial Exploratória (AFE) sobre os dados dos 10 critérios. Em ambas as avaliações, as variáveis mostraram-se adequadas para se proceder a análise fatorial (2015.2 - KMO 0,937; Bartlett: qui-quadrado 37.860,2, com 45 gl, $p < 0,001$; 2016.1 - KMO 0,941; Bartlett: qui-quadrado 35.735,1, com 45 gl, $p < 0,001$) e em ambas as extrações apenas um fator com autovalor maior que 1 foi gerado, e nos dois casos os escores fatoriais foram sempre moderados ou elevados (o escore fatorial mínimo de 2015.2 foi de 0,843, e de 2016.1 foi de 0,852). Ou seja, em ambos os casos verificamos que, de fato, parece haver apenas um fator subjacente ao conjunto de variáveis, o que converge com a Proposição 1 e incide que há pleno sentido em supor que os discentes

avaliam cada critério, em média, a partir de uma motivação geral latente, que presumimos ser a 'satisfação geral do discente com o trabalho do docente'.

Uma consequência imediata deste resultado é que, em lugar de usar 10 itens de avaliação do formato anterior, e seguindo a recomendação de alguns psicometristas (BERGKVIST; ROSSITER, 2007; FISHER; MATTHEWS; GIBBONS, 2016), seria possível a utilização de apenas um item global. Esse procedimento tem a desvantagem de aumentar, potencialmente, o erro de mensuração (dados que múltiplos itens facilitam a dissolução dos erros; cf. COSTA, 2011), mas tem a vantagem de tornar a aferição do construto (a satisfação geral) mais direta, além de deixar mais espaço disponível no instrumento para potencial acréscimo de outras avaliações relevantes. Em suma, o resultado de confirmação da Proposição 1 sinalizou a possibilidade de simplificação da averiguação e o ganho de espaço no instrumento, e gerou ainda o desafio de não se perder o detalhamento referente aos 10 critérios anteriormente avaliados.

Já na avaliação da Proposição 2, optamos por usar o coeficiente de correlação de Pearson entre as variáveis de média das notas atribuídas aos alunos pelo docente, e de média global das avaliações (considerando os 10 itens) recebidas pelo docente. A correlação encontrada foi positiva e estatisticamente não nula, mas em um nível baixo (2015.2 – $r=0,221$, $p<0,001$; 2016.1 – $r=0,195$, $p<0,001$). Ou seja, embora haja associação positiva entre as variáveis, o que confirmaria a Proposição 2, a intensidade dessa associação mostra que, se as notas dadas pelo professor ajudam a melhorar sua avaliação, isso tem uma influência diminuta.

Para explorar ainda mais esse resultado, decidimos realizar uma segunda análise considerando a situação final dos discentes nas disciplinas (aprovado ou reprovado) e as notas atribuídas aos docentes. Nesse caso, computamos para os docentes duas médias globais, gerando duas variáveis adicionais na base de dados (uma com as médias dos alunos aprovados e outra dos alunos reprovados), para posterior comparação dos resultados médios dessas duas novas variáveis. Como resultado, verificamos que os alunos reprovados tendem a avaliar pior, em média, o professor do que os alunos aprovados, sendo essa uma das prováveis explicações da correlação verificada. A partir desses resultados, entendemos que há indicativos de que a Proposição 2 tem confirmação parcial, e concluimos que as avaliações dos professores são pouco impactadas pelas notas absolutas atribuídas aos alunos, com esse baixo impacto associado à variação de satisfação dos alunos que tiveram sucesso ou insucesso na disciplina.

O impacto desse resultado para a o novo instrumento está relacionado não ao conteúdo em si, mas à apresentação dos resultados, e ficaria justificada a implementação, na divulgação separadamente dos resultados das médias dos alunos aprovados e dos alunos reprovados. Esse

resultado também nos motivou a refletir sobre alternativas de minimizar ainda mais qualquer efeito da nota final recebida sobre a avaliação emitida pelo discente e, no debate dos resultados, a equipe de análise ponderou a opção de inclusão de um item de autoavaliação do discente por disciplina, pelo entendimento de que a autoavaliação poderia motivar uma avaliação mais consciente da parte do aluno.

Autoavaliação do discente é, em verdade, uma recomendação recorrente e até prevista em alguns modelos de avaliação institucional. Sobre esse aspecto, Novaes e Musse (2008, p. 18) relatam, no caso que estudaram, que “se o aluno avalia negativamente o professor, atribui também a si um mau desempenho”. Se isso procede, então podemos supor que essas duas avaliações se complementam e sinalizam potencialmente coerência das medidas, o que pode ser útil para gestores e docentes. Com essas constatações, avançamos para a etapa de construção do novo instrumento.

3.2 Construção e análise do novo instrumento

Para a construção do novo instrumento foram considerados alguns apontamentos levantados na etapa de avaliação do instrumento usado anteriormente e outros requisitos definidos pela equipe responsável a partir da seguinte série de ações: (1) fechamento da análise das avaliações anteriores; (2) avaliação de indicações legais, da literatura, de especialistas e do próprio julgamento da equipe de trabalho; (3) organização do conjunto de ideias e um primeiro esboço para exposição à equipe de gestão da Pró-reitoria de Graduação; (4) desenvolvimento do primeiro instrumento para implementação computacional; (5) articulação com a equipe técnica de tecnologia da informação para implementação do instrumento junto ao sistema integrado de gestão acadêmica; (6) debate de detalhes técnicos, acadêmicos e gerenciais envolvendo a equipe de análise, a equipe de gestão da Pró-Reitoria e os técnicos de tecnologia; (7) consolidação do instrumento e aplicação a partir de 2016.2; (8) avaliação da consistência estatística dos resultados e consolidação do instrumento.

Esse conjunto de ações foi desenvolvido nessa sequência enquanto etapas lógicas de construção do instrumento, porém, durante a execução, a articulação entre as pessoas envolvidas era contínua com trocas de mensagens e tomadas de posições por meio presencial ou digital. O pressuposto foi de que a proposição e o teste do instrumento envolveriam, necessariamente: a equipe de elaboração, com professores com melhor domínio em teoria da mensuração; a equipe de gestão, que, efetivamente é a que toma as decisões institucionais; e a equipe de tecnologia da informação, que teve que implementar o instrumento definido de tal

modo que cada aluno recebesse um questionário personalizado referente às disciplinas que cursou.

Como premissa de elaboração por parte das pessoas envolvidas foi definido que o instrumento deveria ser aprimorado em usabilidade referente ao sistema computacional (ou seja, teria que ser menos cansativo para o respondente e assim melhorar a aferição do que se pretende medir). Nesses termos, e considerando o espaço ganho em relação ao instrumento anterior, foi decidido que outras questões seriam incluídas desde que não tornasse o instrumento longo. Concernente às questões e variáveis do instrumento, tomamos por decisão:

- a inclusão de um item para avaliar a satisfação geral do aluno em relação ao professor, com uma escala de mensuração de 11 pontos (de 0 a 10) em substituição dos 10 itens anteriores. A referência aqui foi de que a avaliação da satisfação é baseada em um processo cognitivo em que os alunos comparam suas expectativas com a realidade percebida sobre aquilo que ele está analisando (SZYMANSKI; HENARD, 2001), que neste caso, é o professor.
- para não perder a referência e a opinião dos discentes relativo aos critérios de avaliação anteriores, procedemos à alteração na estratégia de medição de nove dos 10 itens correspondentes ao Quadro 1 (optamos por excluir o item "Atendimento individual em sala de aula", que foi entendido como de pouco valor informativo), de modo que, ao invés de uma escala de mensuração de 0 a 10, passamos a usar uma escala binária, permitindo que o aluno marcasse apenas os itens que ele considerava que o professor necessitaria melhorar. Essa alteração diminuiu substancialmente o tamanho do instrumento e a quantidade de cliques do mouse por parte do aluno, sem perder a avaliação dos critérios. Consideramos também aqui que a quantidade de cliques é tida como um relevante indicador de qualidade na usabilidade de sistemas de informação, de modo que quanto menor for essa quantidade, melhor a sua usabilidade, já que o usuário alcança seu objetivo mais rapidamente;
- Incluímos três dimensões a serem avaliadas: autoavaliação, disciplina e curso. Isso foi possível pela diminuição do esforço do aluno decorrente do aprimoramento da usabilidade do instrumento. Abaixo temos os detalhes:
 - a) **Autoavaliação:** o discente era convidado a emitir uma nota para o seu desempenho pessoal na disciplina. O seguinte enunciado foi incluído: *“Por favor, dê uma nota (de 0 - muito ruim, a 10 - muito bom) para SEU desempenho pessoal na disciplina em termos de comprometimento e motivação”*; na tela do sistema, ao lado dessa questão, haveria campos com os nomes das disciplinas cursadas, de modo que o discente apontava sua autoavaliação por disciplina;
 - b) **Disciplina:** o discente era convidado a indicar sua percepção de dificuldade e de importância da disciplina para o curso, com os seguintes enunciados: importância - *“Na sua percepção, qual o nível de importância (de 0 - sem importância, a 10 - extremamente importante) das disciplinas cursadas para o seu curso?”*; dificuldade - *“Na sua percepção, qual o nível de dificuldade DOS CONTEÚDOS das disciplinas cursadas (de 0 - muito fácil, a 10 - muito difícil)?”*; na tela do sistema, ao lado de cada questão, havia campos com os nomes das disciplinas cursadas;

- c) **Curso:** a nível de curso, dois indicadores foram criados para permitir que a instituição e os gestores responsáveis tivessem um diagnóstico atual e ao longo do tempo sobre os cursos. O primeiro estava relacionado à medida de ‘vinculação do aluno ao curso’, na qual usamos o NPS (*Net Promoter Score*) definido por Reichheld (2006). O seguinte enunciado foi, portanto, definido: “*Considerando a experiência com seu curso até esse último período, a probabilidade de você recomendar esse curso para um amigo ou parente próximo é (de 0 - muito improvável, a 10 - muito provável)*”. O NPS é um indicador de lealdade amplamente utilizado em virtude da sua simplicidade e confiabilidade, e por ser um importante preditor de variáveis relacionados ao desempenho organizacional; seu uso nesta avaliação foi uma inovação em relação ao contexto universitário. Já o segundo indicador questionava a predisposição do aluno a sair do curso, e usamos o seguinte enunciado: “*Seu interesse em sair de curso (mudar de curso na UFPB ou para outra instituição, parar de estudar etc.) no momento atual é (de 0 - muito baixo, a 10 - muito alto)*”. A motivação para a inclusão desse item repousou principalmente na necessidade de se obter um indicador associado à intenção de evasão do aluno. Essas duas questões eram gerais e não tinham vinculação a disciplina ou docente.

Além desses itens de aferição objetiva, no final do instrumento foi deixado um campo para comentários escritos de até 600 caracteres, em resposta à seguinte questão: “Deseja comentar sobre quais disciplinas/professores?”.

O novo instrumento, que pode ser observado em anexo, foi aplicado em dois semestres (2016.2 e 2017.1) e os dados gerados foram apropriados para análise estatística. Além dos dados desse novo instrumento, utilizamos ainda os dados do instrumento anterior para avaliação de compatibilidade de resultados da ‘satisfação global’. Nas subseções seguintes discorreremos sobre cada uma das dimensões do novo instrumento proposto.

- A variável de satisfação geral do aluno em relação ao professor

Avaliamos o desempenho e a adequação da variável ‘satisfação geral do discente com o docente’ de duas formas: primeiro, pela associação das médias das notas dos docentes nessa avaliação com as notas globais agregadas do instrumento anterior; segundo, pela relação dessa variável com os nove indicadores que mantivemos oriundos do instrumento anterior, com a mudança na escala de verificação (que passou a ser binária, conforme indicado).

Referente ao primeiro procedimento, optamos então por comparar a satisfação geral de cada professor considerando os períodos 2016.2 e 2017.1 com a média global gerada considerando as 10 notas referente às dimensões avaliadas nos períodos 2015.2 e 2016.1. Tivemos então avaliações nesses quatro períodos de aproximadamente 2000 docentes, que foram reunidos em uma base de dados com variáveis pareadas de notas, uma para cada período,

em seguida foram verificadas as correlações de Pearson das notas dos docentes entre cada um dos quatro semestres considerados.

Pelos resultados, que estão mostrados na Tabela 1, observamos que todas as correlações ficaram em torno de 0,60, tanto na comparação dos semestres em que foram aplicados o mesmo instrumento, quanto na comparação dos semestres em que foram aplicados os instrumentos diferentes. O resultado sinaliza que há uma relação evidente entre a ‘satisfação geral’ do novo instrumento e a média das 10 dimensões do instrumento anterior. Há evidências, portanto, de que não houve perda na aferição de um indicador global para o docente, reforçando a proposição 1.

Tabela 1 – Correlações extraídas entre as médias das dimensões (no caso dos semestres 2015.2 e 2016.2) e a variável satisfação geral (no caso dos semestres 2016.2 e 2017.1)

Semestre	2015.2 (antigo)	2016.1 (antigo)	2016.2 (novo)
2016.1 (antigo)	0,64	-	
2016.2 (novo)	0,56	0,58	-
2017.1 (novo)	0,55	0,54	0,63

Fonte: Dados da pesquisa (2018)

Como informado, nove dos 10 critérios que definiam as variáveis do instrumento anterior foram mantidos, mas com a conversão da escala para o formato binário. Esse formato possibilita a extração do percentual de indicações atribuídas pelos alunos para cada professor, servindo como medida complementar da medida de satisfação geral. Ou seja, cada docente passou a ter a indicação de uma média geral de satisfação mais a indicação dos percentuais de alunos que sinalizaram necessidade de melhoria em cada critério.

Com esses dados computados para cada um dos docentes, decidimos avaliar em que medida a satisfação geral do aluno poderia ser explicada pelos percentuais dos critérios (dimensões de avaliação). Fixamos então o entendimento de que, se a variável de ‘satisfação geral’ mensurasse bem o construto, então esta seria bem explicada por suas dimensões de avaliação. Aplicamos então a técnica de análise de regressão linear múltipla, considerando como variável explicada a média de satisfação geral por docente, e como variáveis explicativas os percentuais de respostas dos itens apontados pelos alunos relativos ao que o professor precisava melhorar. Também aqui foram consideradas apenas docentes com o mínimo de 10 avaliações no total, independente de serem de uma ou de mais turmas; o total foi de 2088 docentes para 2016.2 e de 2324 docentes para 2017.1.

Os resultados da avaliação global (pelo teste F) e do nível de explicação dos modelos de regressão linear (pelo R²) para ambos os semestres podem ser verificados na Tabela 2, e

mostram, pelos valores dos coeficientes R^2 , que foi bastante significativo o nível de explicação da variável predita pelas variáveis predictoras ($R^2=0,72$ para 2016.2 e $R^2=0,74$ para 2017.1). Ou seja, por esses resultados, temos evidências de que, independente do semestre, os percentuais de indicações de melhoria dos 9 critérios avaliados pelos discentes explicam mais de 70% da variação do nível de satisfação geral dos discentes quanto aos professores.

Tabela 2 – Resultados dos testes dos modelos de regressão linear referentes aos semestres 2016.2 e 2017.1

Semestre	R^2	F (gl1; gl2)	p-valor
2016.2	0,72	601,8 (9; 2078)	p < 0,00
2017.2	0,74	729,7 (9; 2314)	p < 0,00

Fonte: Dados da pesquisa (2018)

Tendo como base todos esses procedimentos, pudemos concluir que havia consistente evidência de validade estatística da medida proposta, ficando então fixada a nova ‘regra’ de mensuração nas duas formas: a medida geral com um único item e com escala de verificação de 11 pontos (0 a 10); e as medidas pelos 9 critérios, com escala de verificação binária.

- Dimensões de autoavaliação e disciplina

Concernente à autoavaliação, a questão tinha conteúdo genérico e sua implementação tornou possível identificar, por disciplina, qual o nível declarado de comprometimento e motivação dos alunos a partir de um escore médio ou mediano pontual, na escala de 0 a 10. Além disso, essa medida foi colocada como sendo a primeira questão no instrumento, como forma de reduzir potencial efeito de ‘vontade de retaliação’ do discente em relação ao professor. Na extração de medidas, que foi por professor e por disciplina, tivemos resultados bastante variados, o que era esperado supondo que o nível de envolvimento e motivação dos alunos varia por disciplina.

Nas duas medidas relativas à disciplina (de dificuldade e importância para o curso) procedemos de forma semelhante ao que fizemos na análise de autoavaliação, com verificação das medidas de posição e dispersão, além da análise, exploratória e qualitativa, de algumas disciplinas específicas. Os resultados mostraram bom potencial de uso de ambas as medidas, em alguns casos com evidente necessidade de melhor explicação do docente sobre a relevância de seu componente no currículo, ou dando a indicação da necessidade de melhoria dos métodos de ensino dos docentes para, se fosse o caso, melhorar a percepção de dificuldade das disciplinas. As variações por disciplina, ou mesmo em cada disciplina (aferida pelo desvio

padrão) mostraram que os padrões de resposta foram adequados e bem explicados nos casos analisados.

Nas primeiras extrações nosso foco foi de natureza mais exploratória e qualitativa e analisamos o desempenho da escala em termos de variação das medidas extraídas e de potencial utilidade de resultados para docentes e gestores de graduação. Entendemos, portanto, que os resultados extraídos foram adequados, porém outros tratamentos de performance dos itens podem ser ainda avaliados, principalmente considerando a evolução de medidas ao longo do tempo.

- Dimensão de avaliação do curso

Referente à questão de vínculo do aluno ao curso, utilizamos a métrica já consolidada na literatura especializada em mensuração de lealdade – *Net Promoter Score* (NPS). Segundo a metodologia de Reichheld (2006), para extrair essa medida é necessário calcular a diferença (saldo) entre o percentual de promotores e o percentual de detratores do curso, sendo os ‘promotores’ aqueles que marcam escores 9 ou 10 na escala (esses estudantes, possivelmente, irão promover o curso de alguma forma, opinando publicamente em seu meio social em favor do curso), e os ‘detratores’ aqueles que marcam escores de 0 a 6 na escala (esses alunos, possivelmente, são aqueles que não emitem posição sobre o curso, e se o fazem é com opiniões desfavoráveis ao curso; segundo a metodologia de Reichheld, as pessoas que marcam 7 e 8 são considerados ‘neutras’ ou ‘passivas’ em relação ao curso).

O escore de NPS gera um saldo líquido de promotores em relação a detratores, e a suposição é de que, quanto maior o NPS, maior é a intensidade de vínculo do aluno com seu curso. Valores negativos ou positivos próximos de zero indicam a necessidade de ações da liderança institucional no sentido de motivar mais o corpo discente e gerar mais proximidade, maior envolvimento, maior divulgação positiva do curso etc.

A extração experimental do NPS em 98 cursos da instituição indicou uma significativa variação de resultado entre os cursos (desvio padrão de 24,1 em 2016.2; e 24,9 em 2017.1). Como apresentado na Tabela 3, poucos cursos demonstraram NPS negativos nos dois semestres avaliados, indicando haver maioria de alunos dispostos a falar bem do curso (92% em 2016.2 e 93% em 2017.1). A correlação entre o NPS dos cursos nos dois períodos analisados foi de 0,88, indicando que não há mudanças significativas entre os semestres. Além disso, houve o entendimento da liderança institucional e da equipe de proposição do instrumento de que a medida gera resultados adequados para embasar a tomada de decisão de diversos gestores

vinculados aos cursos (coordenadores, chefes de departamento, diretores de unidades etc.). Esse entendimento é obtido a partir de análises de cursos específicos, por meio da triangulação da medida objetiva (NPS) com evidências do conhecimento prévio que se tem a cerca do contexto desses cursos.

Tabela 3 – Medidas sobre o indicador NPS dos cursos

Semestre	NPS				Interesse em evadir do curso	
	% NPS positivo	%NPS negativo	Média	Desvio padrão	Média	Correlação com NPS
2016.2	92%	6%	36	24,1	3,10	-0,68
2017.1	93%	7%	37	24,9	3,22	-0,59

Fonte: Dados da pesquisa (2018)

Ainda relativo à dimensão ‘curso’, extraímos a medida de predisposição à evasão dos cursos, na escala de 0 a 10. Para a análise, extraímos a média, mediana e medidas de desvio padrão por curso. Os resultados foram variados e, na leitura da equipe de avaliação, mostravam de forma clara quais os cursos tinham maiores problemas de risco de evasão. Na análise exploratória, houve boa aproximação entre as medidas por curso e os níveis de evasão por curso calculados em outros estudos da Pró-reitoria.

Em suma, na dimensão ‘curso’, as duas variáveis usadas mostraram-se adequadas por fornecerem resultados úteis para os gestores e, principalmente, por terem captado variações por curso que eram bem justificadas e coerentes com outras medidas de avaliação utilizadas. Ademais, essas duas variáveis apresentaram uma correlação negativa e entre moderada para forte ($r=-0,68$ para 2016.2 e $r=-0,59$ para 2017.1), conforme exposto na Tabela 3. Ou seja, quanto maior fosse o NPS de um curso, menor era a média do interesse dos alunos em abandonar o curso. Ressalta-se que a interpretação sobre os indicadores deve ser realizada pelos atores e gestores institucionais envolvidos no curso, pois há particularidades contextuais que devem ser consideradas. Nesse sentido, a recomendação foi de que os gestores, juntamente com o NDE do curso, empreendessem sempre pesquisas complementares para um melhor entendimento de sua realidade, possibilitando a definição de ações e estratégias mais efetivas.

4 Conclusões

Considerando o esforço empreendido na proposta do novo instrumento, tivemos uma proposta que se tornou mais simples e mais completa do que o anterior, principalmente pelo esforço de redução da complexidade e pela diminuição na interação do aluno com a interface

do sistema computacional de avaliação, o que impacta na qualidade de aferição dos itens que estão sendo medidos. Pela adoção da medição baseada em um só item para o construto de satisfação com o professor pudemos incluir os novos itens para avaliar outras dimensões além do docente, como a autoavaliação, a dificuldade do conteúdo e importância das disciplinas no curso, o NPS do curso e o interesse em abandoná-lo. Ou seja, em um mesmo instrumento temos um conjunto de outras informações relevantes para docentes e gestores acadêmicos.

Com as análises qualitativas e quantitativas desenvolvidas pudemos concluir que as métricas do novo instrumento apresentaram boas evidências de validade e confiabilidade. Além dos aspectos psicométricos, o conjunto de variáveis propostas foi submetido à avaliação de gestores com a intenção de verificar sua pertinência e utilidade. Ao final, pudemos apresentar um instrumento em que: o primeiro item, referente à autoavaliação, analisa a performance do aluno e ainda o conscientiza no sentido de gerar autorreflexão sobre o seu papel no processo formativo; as variáveis seguintes, sobre dificuldade e importância da disciplina, permitem pensar quais dessas merecem um esforço maior para serem compreendidas pelos alunos em relação à sua relevância e sua condição dentro de cada curso; as variáveis referentes à performance docente vêm em seguida, com as nove dimensões e a medida de 'satisfação geral' em relação ao professor, e servem de referência tanto para um nível individual, permitindo que cada docente reflita sobre suas práticas pedagógicas, quanto para um nível agregado, possibilitando que os gestores acadêmicos compreendam o nível de satisfação dos alunos em relação aos seus docentes; por fim, temos as variáveis da dimensão curso, que oportunizam diagnosticar a intensidade do vínculo dos alunos ao curso e sua intenção em abandoná-lo, sendo importantes indicadores para enfrentar o problema da evasão e retenção.

A construção do instrumento aqui relatada foi resultado de um dos esforços empreendidos na Universidade Federal da Paraíba, direcionados a consolidar suas práticas de gestão da informação e de inteligência institucional. Entendemos, por outro lado, que os resultados podem ser úteis para outras instituições e deixamos como recomendações de outros estudos o relato de experiências semelhantes, que, no conjunto, vão dando maior profundidade aos estudos e maior qualidade às práticas de avaliação institucional.

Referências

BERGKVIST, Lars; ROSSITER, John R. The predictive validity of multiple-item versus single-item measures of the same constructs. **Journal of Marketing Research**, Los Angeles, v. 44, n. 2, p. 175-184, 2007.

COSTA, Francisco José. **Mensuração e desenvolvimento de escalas:** aplicações em Administração. Rio de Janeiro: Ciência Moderna, 2011.

COSTA, Francisco José; BISPO, Marcelo Sousa; PEREIRA, Rita de Cássia de Faria. Dropout and retention of undergraduate students in management: a study at a Brazilian Federal University. **RAUSP Management Journal**, São Paulo, v. 53, n. 1, p. 74-85, 2018. Disponível em: https://www.scielo.br/scielo.php?pid=S2531-04882018000100074&script=sci_abstract. Acesso em: 20 ago. 2018.

DEVELLIS, Robert F. **Scale development:** theory and applications. Los Angeles: Sage publications, 2016.

DIAS, Ellen C. M.; THEÓPHILO, Carlos R.; LOPES, Maria A. S. Evasão no ensino superior: estudo dos fatores causadores da evasão no curso de Ciências Contábeis da Universidade Estadual de Montes Claros–Unimontes–MG. In: CONGRESSO USP DE INICIAÇÃO CIENTÍFICA EM CONTABILIDADE. 7., 2010, São Paulo. **Anais [...]**. São Paulo: Êxito, 2010.

ELLIOTT, Kevin M.; SHIN, Dooyoung. Student satisfaction: an alternative approach to assessing this important concept. **Journal of Higher Education Policy and Management**, Abingdon, v. 24, n. 2, p. 197-209, 2002.

FERREIRA, Maurício da Silva; FREITAS, Antônio Alberto da Silva Monteiro. Implicações da avaliação institucional na gestão universitária: a experiência da Universidade Católica do Salvador (UCSAL). **Avaliação**, Campinas; Sorocaba, v. 22, n. 1, p. 201-221, 2017. Disponível em: https://www.scielo.br/scielo.php?pid=S1414-40772017000100201&script=sci_abstract&tlng=pt. Acesso em: 20 ago. 2018.

FISHER, Gwenith G.; MATTHEWS, Russell A.; GIBBONS, Alyssa Mitchell. Developing and investigating the use of single-item measures in organizational research. **Journal of Occupational Health Psychology**, Washington, v. 21, n. 1, p. 3-23, 2016. Disponível em: https://www.researchgate.net/publication/275216825_Developing_and_Investigating_the_Use_of_Single-Item_Measures_in_Organizational_Research. Acesso em: 20 ago. 2018.

NOVAES, Adelina de Oliveira; MUSSE, Luciana Barbosa. Avaliação Institucional: a busca pela integração dos resultados. **Estudos em Avaliação Educacional**, São Paulo, v. 19, n. 39, p. 9-28, 2008.

NUNNALLY, Jum C.; BERNSTEIN, Ira H. **Psychometric theory**. New York: McGraw-Hill, 1978.

REICHHELD, Frederick. **The Ultimate Question:** driving good profits and true growth. Boston, MA: Harvard Business Press, 2006.

ROSSITER, John R. **Measurement for the social sciences:** the C-OAR-SE method and why it must replace psychometrics. New York: Springer Science & Business Media, 2011.

SLHESSARENKO, Michelli *et al.* Evasão na educação superior para o curso de bacharelado em Sistema de Informação. **Revista Gestão Universitária na América Latina (GUAL)**,

Florianópolis, v. 7, n. 1, p. 128-147, jan. 2014. Disponível em: <https://periodicos.ufsc.br/index.php/gual/article/view/1983-4535.2014v7n1p128>. Acesso em: 20 ago. 2018.

SZYMANSKI, David M.; HENARD, David H. Customer satisfaction: a meta-analysis of the empirical evidence. **Journal of the Academy of Marketing Science**, New York, v. 29, n. 1, p. 16-35, 2001.

URBINA, Susana. **Fundamentos da testagem psicológica**. Porto Alegre: Artmed Editora, 2009.

VELOSO, Tereza Christina M. A.; ALMEIDA, Edson Pacheco de. Evasão nos cursos de graduação da Universidade Federal de Mato Grosso, campus universitário de Cuiabá: um processo de exclusão. **Série Estudos**, Campo Grande, MS, n.13, p. 133-148, 2002.

WOJAHN, Rafaela Matte; RAMOS, Sheila Patrícia; CARVALHO, Luciano Castro de. Proposta de modelo para avaliação da satisfação com a qualidade do ensino. **Revista Gestão Universitária na América Latina - GUAL**, Florianópolis, v. 11, n. 1, p. 1-23, jan. 2018. Disponível em: <https://periodicos.ufsc.br/index.php/gual/article/view/1983-4535.2018v11n1p1>. Acesso em: 20 ago. 2018.

ANEXO A – O instrumento proposto

Dimensão	Conteúdo	Escala	Métricas
Autoavaliação	Por favor, dê uma nota (de 0 - muito ruim, a 10 - muito bom) para SEU desempenho pessoal na disciplina em termos de comprometimento e motivação	11 pontos, de 0 a 10	Média, mediana e desvio padrão
Disciplina - importância	Na sua percepção, qual o nível de importância (de 0 - sem importância, a 10 - extremamente importante) das disciplinas cursadas para o seu curso?	11 pontos, de 0 a 10	Média, mediana e desvio padrão
Disciplina - dificuldade	Na sua percepção, qual o nível de dificuldade DOS CONTEÚDOS das disciplinas cursadas (de 0 - muito fácil, a 10 - muito difícil)?	11 pontos, de 0 a 10	Média, mediana e desvio padrão
Docente - critérios	Cumprimento do plano de curso	Binária	Percentual
	Relacionamento com a turma	Binária	Percentual
	Comparecimento às aulas	Binária	Percentual
	Cumprimento do horário de início e de término das aulas	Binária	Percentual
	Atualização dos conteúdos	Binária	Percentual
	Clareza na exposição dos conteúdos	Binária	Percentual
	Disponibilidade para atendimento fora da sala de aula	Binária	Percentual
	Qualidade da bibliografia	Binária	Percentual
Qualidade das avaliações	Binária	Percentual	
Docente - satisfação geral	Por favor, aponte sua satisfação geral (de 0 - totalmente insatisfeito, a 10 - totalmente satisfeito) com o desempenho de cada professor	11 pontos, de 0 a 10	Média, mediana e desvio padrão
Curso - vínculo	Considerando a experiência com seu curso até esse último período, a probabilidade de você recomendar esse curso para um amigo ou parente próximo é (de 0 - muito improvável, a 10 - muito provável)	11 pontos, de 0 a 10	NPS
Curso - intenção de evasão	Seu interesse em sair de curso (mudar de curso na UFPB ou para outra instituição, parar de estudar etc.) no momento atual é (de 0 - muito baixo, a 10 - muito alto)	11 pontos, de 0 a 10	Média, mediana e desvio padrão
Geral	Deseja comentar sobre disciplinas/professores?	-	