



doi: <http://dx.doi.org/10.7213//psicolargum.38.100.AO02>

## **Feedback fundamentado no Teacher Behavior Checklist para professores universitários: Uma pesquisa exploratória**

*Feedback based on the Teacher Behavior Checklist for university teachers: an exploratory research*

*Feedback basados en lo Teacher Behavior Checklist para docentes universitarios: una investigacion exploratória*

---

Marcelo Henrique Oliveira Henklain

Doutor em Psicologia pela Universidade Federal de São Carlos, professor da Universidade Federal de Roraima, E-mail: [marcelo.henklain@ufr.br](mailto:marcelo.henklain@ufr.br) Orcid: <http://orcid.org/0000-0001-9884-8592>

Vicente Cícero Gerônimo Júnior

Psicólogo formado pela Universidade Federal de Roraima, E-mail: [psivicentejr@gmail.com](mailto:psivicentejr@gmail.com) Orcid: <http://orcid.org/0000-0002-6713-288X>

Juliana Rosa Lira

Psicólogo formado pela Universidade Federal de Roraima, E-mail: [rlirajuliana@gmail.com](mailto:rlirajuliana@gmail.com) Orcid: <http://orcid.org/0000-0001-5563-3022>

Rafael Vilas Boas Garcia

Doutor em Educação Especial pela Universidade Federal de São Carlos e Professor no Centro de Educação da Universidade Federal de Roraima  
E-mail: [rafaelgarcia2@gmail.com](mailto:rafaelgarcia2@gmail.com) Orcid: <http://orcid.org/0000-0001-8246-6813>

---

### Resumo

Esta pesquisa exploratória investigou efeitos da avaliação com o Teacher Behavior Checklist (TBC), respondido por estudantes, em condições com ou sem feedback, sobre o desempenho de professores universitários brasileiros. Participaram seis professores, de diferentes disciplinas, avaliados duas vezes, no meio do semestre por 105 alunos e ao final por 93. Após a primeira avaliação, o procedimento dependeu do grupo no qual o professor estava: G1 - sem feedback; G2 - apenas feedback; G3 - feedback e plano sobre o que fazer para melhorar resultados no TBC. Os três grupos aperfeiçoaram o desempenho, sendo que G1 obteve os maiores escores de melhora, contrariando as hipóteses da pesquisa. Verificou-se que o TBC pode ser útil para favorecer mudanças de desempenho docente, que o delineamento deve ser modificado para garantir a formação de grupos homogêneos entre si, e que o modelo de feedback adotado pode ser insuficiente para promover mudanças em curto prazo.

**Palavras-chave:** Avaliação do desempenho do professor; Ensino Superior; Psicometria.

### Abstract

*This exploratory research investigated the effects of the evaluation with the Teacher Behavior Checklist (TBC), answered by students, under conditions with or without feedback, on the performance of Brazilian college teachers. Six teachers from different areas participated, assessed twice, in the middle of the semester by 105 students, and at the end by 93. After the first assessment, the procedure depended on the group in which the teacher was: G1 - no feedback; G2 - feedback only; G3 - feedback and plan on what to do to improve TBC results. The three groups improved their performance, with G1 obtaining the highest scores, contrary to the research hypotheses. We found that the TBC can be useful to favor changes in teacher performance, that the study design must be modified to guarantee the formation of homogeneous groups among themselves, and that the adopted feedback model may be insufficient to promote changes in the short term.*

**Keywords:** Evaluation of teacher performance; University education; Psychometry.

### Resumen

*Esta investigación exploratoria investigó los efectos de la evaluación con lo Teacher Behavior Checklist (TBC), respondida por los estudiantes, en condiciones con o sin feedback, sobre el desempeño de profesores universitarios brasileños. Participaron seis maestros de diferentes disciplinas, evaluados dos veces, en mediados del semestre por 105 estudiantes y al final por 93. Después de la primera evaluación, el procedimiento dependía del grupo en el que el maestro estaba: G1 - sin feedback; G2: solo feedback; G3: feedback y plan sobre qué hacer para mejorar los resultados en TBC. Los tres grupos mejoraron su rendimiento, con G1 obteniendo los puntajes de mejora más altos, contrario a las hipótesis de investigación. Se descubrió que el TBC puede ser útil para favorecer cambios en el desempeño de los docentes, que el diseño de estudio debe modificarse para garantizar la formación de grupos homogéneos entre ellos y que el modelo de feedback adoptado puede ser insuficiente para promover cambios a corto plazo.*

**Palabras clave:** Evaluación del desempeño docente; Enseñanza superior; Psicometría.

---

## Introdução

A avaliação de desempenho é um processo necessário ao aprimoramento profissional. No Brasil, a avaliação de professores de nível superior é uma realidade que pode ser facilmente demonstrada com base em leis federais. Contudo, deve-se questionar se a avaliação existente de fato auxilia o professor a aprimorar o seu trabalho. De acordo

com Cassettari (2014), existem dois tipos básicos de avaliação de desempenho: somativa e formativa. A primeira se refere a avaliações cujo objetivo é orientar a tomada de decisões organizacionais, tais como contratação, promoção, demissão, dentre outras. Já as avaliações formativas cumprem o papel primordial de auxiliar o profissional a aperfeiçoar o seu desempenho, não tendo impacto sobre a sua condição funcional na organização. Pergunta-se, então, quais são algumas das principais leis sobre a avaliação de professores e qual o tipo de avaliação que contemplam.

A Lei n. 8.112/1990, por exemplo, estabelece o sucesso em certos critérios avaliativos (assiduidade, disciplina, capacidade de iniciativa, produtividade e responsabilidade) como condição para a aprovação do professor de universidade pública em seu estágio probatório. A Lei n. 12.772/2012, também vinculada às instituições públicas, define requisitos para a progressão docente na carreira, que incluem a avaliação de seu desempenho. A Lei n. 10.861/2004, por sua vez, institui o Sistema Nacional de Avaliação da Educação Superior (SINAES) e abrange todas as instituições de ensino superior brasileiras. O SINAES inclui nos seus critérios a avaliação do ensino oferecido, o que é feito em parte com base no perfil e no desempenho em sala de aula dos professores. Assim, não há dúvidas de que professores são constantemente avaliados. Será que se beneficiam de tais avaliações?

Ribeiro (2012) sustenta que, atualmente, as avaliações de professores de nível superior são frequentemente de natureza somativa, o que fica claro nas Leis n. 8.112/1990 e 12.772/2012. Segundo Ribeiro, pode-se supor que os professores, provavelmente, pouco se beneficiam dos resultados das avaliações a que são submetidos, pois as avaliações somativas não têm a finalidade primordial de promover o aperfeiçoamento desses profissionais.

Esse cenário é problemático porque a classe profissional “professor do magistério superior” é quem fica em contato direto com os estudantes no dia a dia, e cujo trabalho é crucial para facilitar o aprendizado, garantindo que os estudantes adquiram comportamentos profissionais necessários à sua atuação na sociedade em diversos setores fundamentais (Henklain, Carmo, & Haydu, 2018). Além disso, sabe-se que o problema com a adoção de avaliações somativas é ainda maior. Dados recentes de uma metanálise indicaram que é fraca a associação entre escores de professores em instrumentos para avaliar desempenho do professor, com base na percepção discente, e o grau de aprendizado dos estudantes (ver Uttl, White, & Gonzalez, 2017). Tais resultados sugerem

que pode ser um equívoco decidir sobre a qualidade da atuação profissional do professor com base em avaliações somativas, sendo pior quando decisões são tomadas com base unicamente nessa avaliação.

No entanto, os problemas com as avaliações somativas não alteram o fato de que seja necessário obedecer às leis federais existentes. Esses problemas apenas apontam para a necessidade de aperfeiçoamento das leis e de que as instituições de ensino superior implantem avaliações formativas, o que não é impedido por qualquer legislação. Essa seria uma alternativa promissora para que o professor receba informações sobre o seu desempenho e identifique como melhorar (ver Ribeiro, 2012; Cassettari, 2014).

Entendendo a importância da avaliação de desempenho do professor do magistério superior, Vieira-Santos e Henklain (no prelo) identificaram em uma revisão de literatura 25 instrumentos brasileiros com evidências psicométricas para avaliar desempenho de professores, sendo que poucos estavam sendo formalmente utilizados por universidades brasileiras. Isso sugere que recursos de apoio à avaliação de professores têm sido construídos, mas ficam subutilizados. Tais recursos, muitas vezes criados para viabilizar avaliações formativas, precisam, portanto, de divulgação e teste empírico sobre a sua eficácia para auxiliar professores.

### **Teacher Behavior Checklist**

Uma das alternativas disponíveis no Brasil para a realização de avaliações formativas é o *Teacher Behavior Checklist* (TBC). Trata-se de um instrumento de 28 itens criado por Buskist et al. (2002) para medir o desempenho do professor em relação ao ensino e, a partir dos dados que produz, fornecer feedbacks para o professor.

Uma das vantagens do TBC é que a versão desse instrumento utilizada para avaliar o desempenho do professor pelo discente possui evidências psicométricas favoráveis nos Estados Unidos (ver Keeley, Smith, & Buskist, 2006) e no Brasil (ver Henklain, Carmo, Haydu, Muniz, Buskist, & Keeley, *no prelo*). Isso facilita o exame de convergências e divergências na avaliação do professor entre países, permitindo a identificação de práticas cruciais para a promoção de aprendizado.

Adicionalmente, as evidências indicam que o TBC é uma medida que descreve comportamentos profissionais do professor que são fundamentais para o ensino de qualidade. Em revisão da literatura conduzida por Buskist e Keeley (2018), foi verificado que o TBC possui evidências de validade de conteúdo dos seus itens em culturas tão

distintas quanto a de países como Estados Unidos, Brasil, Colômbia, Alemanha, China e Arábia Saudita. Em cada país, os dados foram obtidos ao solicitar que professores e estudantes avaliassem o grau de relevância dos itens do TBC para o ensino de excelência.

Esse contexto científico favorável confere mais segurança no uso do TBC em contextos aplicados, pois trata-se de um instrumento que está sob constante escrutínio. Por resistir aos testes científicos reiterados, o TBC demonstra o seu valor enquanto instrumento indicador de condutas docentes observáveis associadas à promoção de aprendizado e do prazer pelos estudos, de acordo com avaliação de alunos em diferentes culturas. Portanto, é de fato possível adotar o TBC como parte de processos de avaliação formativa.

Importa destacar que, de um ponto de vista científico, o termo *feedback*, central para a avaliação formativa, requer uma definição precisa. Para Mangiapanello e Hemmes (2015), o *feedback* pode ser interpretado como um estímulo produzido pelo avaliador cuja natureza é verbal e cujos parâmetros variam em função do responder que o antecedeu, apresentado pela pessoa avaliada. Esse estímulo deve ter uma dupla função: (a) indicar à pessoa avaliada em que medida os seus comportamentos se aproximaram ou diferiram dos comportamentos-alvo; (b) aumentar as chances de que o avaliando se comporte de modo a atenuar a diferença entre seu comportamento e o comportamento-alvo ou, se possível, extinguir essa diferença. Quando comportamento apresentado e comportamento-alvo coincidem, o *feedback* deve ter a função de manter o padrão comportamental que já está adequado.

Ao considerar a definição de *feedback*, é possível notar que o TBC informa ao professor com que frequência ele apresenta qualidades importantes para o ensino, permitindo que identifique o que precisa manter e o que deve mudar, pelo menos, segundo a percepção de uma determinada turma de alunos. Adicionalmente, além de apontar qualidades que requerem aperfeiçoamento, o TBC exibe um rol de condutas que o professor pode adotar para melhorar em determinada qualidade. Com isso, o TBC atende aos critérios da definição de feedback, pois permite comparar o desempenho esperado com aquele apresentado e facilita para o professor a identificação do que pode ser feito para melhorar o seu desempenho.

Em resumo, é viável usar o TBC para administrar *feedbacks*. Nesse processo, deve-se identificar qual a melhor estratégia para informar ao professor sobre os seus comportamentos em sala de aula a partir do TBC – considerando a percepção de quem o

avaliou –, e como descrever o que é esperado de modo que seja possível traçar um plano de ação que promova uma mudança efetiva de comportamento. É preciso também propor uma forma de organizar os resultados decorrentes do TBC que permita ao professor participar ativamente da interpretação dos mesmos, bem como projetar qual ou quais comportamentos ele entende que deva modificar.

### **Consequências sociais da avaliação do desempenho docente**

Como mencionado anteriormente, de acordo com os dados de Vieira-Santos e Henklain (no prelo), instrumentos que avaliem o desempenho de professores em instituições de ensino superior e que possuam evidências psicométricas são escassos, e são ainda menos frequentes estudos que examinem evidências de validade baseadas nas consequências sociais do uso de medidas de avaliação do desempenho docente. No exame de Vieira-Santos e Henklain, nenhum dos 25 instrumentos reportou esse tipo de evidência. Nos estudos com o TBC, quando usado para avaliação docente pelo discente, também não existem estudos nacionais ou internacionais dessa natureza.

O que seria uma consequência social esperada no uso de instrumentos? Segundo Primi, Muniz e Nunes (2009), as evidências baseadas nas consequências de uso de um instrumento consistem em estudos que examinam os seus efeitos sociais (intencionais ou não). Isso significa investigar se o instrumento, ao ser utilizado, produz os efeitos desejados de acordo com o propósito para o qual foi construído. No caso de um instrumento para avaliação do desempenho docente, uma investigação dessa natureza envolve examinar se de fato as informações fornecidas por ele ajudam o professor a aperfeiçoar o seu desempenho.

Silva et al. (2017), em um dos estudos brasileiros mais recentes que apresentam uma medida de desempenho do professor com evidências psicométricas, ilustram o que são consequências sociais do uso de um instrumento. Os autores esclarecem o que esperam do uso da Escala de Avaliação Docente (EADoc): “espera-se que a sua utilização possa gerar impacto em quatro níveis: individual, institucional, político e teórico-científico” (Silva et al., 2017, p. 701). Cada um desses níveis é explicado. De todos eles, o que está vinculado à avaliação formativa do professor é o nível individual. Nesse nível, espera-se que o uso da EADoc auxilie o docente a “[...] avaliar as suas habilidades e competências, refletindo sobre aspectos a serem melhorados na sua atuação” (Silva et al.,

2017, p. 701-702). No limite, espera-se que o uso do instrumento ajude o professor a aprimorar a sua prática.

Esse tipo de investigação é especialmente importante após o estudo de Uttl et al. (2017), que colocou em relativo descrédito medidas de desempenho do professor. Não obstante, os mesmos autores sugerem que há consenso na literatura em torno do uso de tais instrumentos para auxiliar os professores a aperfeiçoar seu desempenho. Nesse sentido, é preciso investigar em que medida e como instrumentos da natureza do TBC podem ser utilizados para auxiliar os professores no seu trabalho.

Diante desse cenário apresentado e de quanto os professores precisam de auxílio em sua atividade profissional, foi definido como objetivo deste estudo exploratório investigar evidências de mudanças no comportamento de professores universitários ao receber feedback sobre o seu desempenho com base em protocolo elaborado a partir de uma avaliação de desempenho do professor pelo discente realizada com o uso do TBC.

Este estudo tem natureza exploratória porque as características dos participantes de cada grupo estudado, conforme será explicado no método, não foram mantidas constantes. Assim, eventuais diferenças nos resultados podem estar relacionadas às diferenças entre os participantes e não ao uso do feedback. Logo, não seria adequado compará-los. Os resultados que serão apresentados devem ser examinados com cautela e na qualidade de sugestão sobre como o TBC poderia ser usado e estudado em contexto natural. Ficará claro que tipo de dado o professor poderia ter sobre o seu desempenho, bem como informações que os gestores de instituições de ensino poderiam monitorar para acompanhar o desenvolvimento dos professores.

## **Método**

### **Participantes**

Participaram nove professores de uma mesma instituição pública da Região Norte. A amostra foi definida por conveniência. A Tabela 1 exhibe a caracterização dos professores, contendo na primeira coluna a identificação, seguida do sexo, curso, disciplina na qual a coleta foi realizada, módulo da disciplina, idade do professor e grupo no qual foi alocado.

Os professores tinham média de idade de 44,30 ( $DP = 5,38$ ). Eles foram distribuídos em três condições: Grupo Controle (G1), Grupo de Feedback (G2) e Grupo

de Feedback com Plano de Ação (G3). Os professores P1 e P4 foram designados para o G1, P2 e P5 para o G2 e P3 e P6 para o G3. Esses professores não possuíam histórico de acesso a feedbacks sobre desempenho ou a elaboração de planos de ação. A universidade na qual trabalham adota apenas avaliação somativa, cujo sistema em vigor na época do estudo era composto por avaliação do superior imediato, autoavaliação e avaliação discente, sendo que o único resultado a que os professores tinham acesso era a nota final.

**Tabela 1.**

Caracterização dos professores participantes do estudo.

Professor	Sexo	Curso	Disciplina	Módulo	Idade	Grupo
P1	♀	LELH	Prática como Componente Curricular IV (PCC IV)	5	47,93	1
P2	♀	LCBL	Zoologia de invertebrados inferiores	3	36,97	2
P3	♂	LCBL	Zoologia dos vertebrados	5	49,17	3
P4	♂	TADS	Inglês Instrumental	1	49,69	1
P5	♂	TADS	Introdução à computação	1	39,62	2
P6	♀	TADS	Estrutura de Dados II	3	42,44	3

*Nota.* ♀ = Feminino; ♂ = Masculino; LELH = Letras, Espanhol e Literatura Hispânica; LCBL = Licenciatura em Ciências Biológicas; TADS = Tecnólogo em Análise e Desenvolvimento de Sistemas.

Considerando que a pesquisa envolve a avaliação do professor pelo discente, além dos professores, participaram também os alunos das disciplinas citadas na Tabela 1. A média de idade os discentes foi de 24,51 ( $DP = 6,78$ ). Essa participação foi dividida em dois momentos: 1ª e 2ª avaliação. Na primeira e segunda aplicação participaram 105 e 93 estudantes, respectivamente. Considerando os casos de estudantes que faltaram em uma das aplicações ou cujos dados não estavam completos porque o discente não respondeu a algum item do protocolo, foi possível identificar para 74 discentes a sua avaliação inicial e final. Para o restante da amostra, não houve como ter certeza de que o participante realizou as duas avaliações. Isso significa que é possível comparar médias do professor antes e depois, considerando a amostra total, 105 e 93. Ou, ainda, comparar resultados com base nos 74 participantes com dados das duas avaliações.

Da amostra de alunos, 10 declararam possuir algum tipo de deficiência, 50,48% declararam pertencer a classe média, 46,67% à classe baixa, 0,95% à classe alta e 1,90% não responderam. A Tabela 2 exibe os principais dados dos estudantes em função do professor avaliado e de ter sido a primeira ou segunda avaliação, desta forma, possibilitando análises mais aprofundadas dos resultados obtidos.

**Tabela 2.**

Caracterização dos estudantes que participaram do estudo.

Professor	Avaliação	Alunos (N)	%	Sexo	Média Idade (DP)
P1	1	7	100,00	♀: 6 ♂: 1	27,89 (9,28)
	2	7	100,00	♀: 6 ♂: 1	27,99 (9,27)
P2	1	20	83,33	♀: 14 ♂: 6	25,56 (8,23)
	2	21	87,50	♀: 13 ♂: 8	26,16 (9,36)
P3	1	13	100,00	♀: 8 ♂: 5	28,48 (8,92)
	2	10	76,92	♀: 6 ♂: 4	28,92 (7,76)
P4	1	18	100,00	♀: 3 ♂: 15	22,46 (6,95)
	2	16	88,90	♀: 2 ♂: 14	21,41 (6,82)
P5	1	21	100,00	♀: 0 ♂: 21	24,53 (4,53)
	2	17	80,90	♀: 0 ♂: 17	24,66 (4,08)
P6	1	26	100,00	♀: 1 ♂: 25	22,31 (3,62)
	2	22	84,60	♀: 2 ♂: 20	23,49 (5,90)

*Nota.* DP = Desvio-padrão; % = Porcentagem da turma completa; ♀ = Feminino; ♂ = Masculino.

### Cuidados éticos

Este estudo foi aprovado pelo Comitê de Ética em Pesquisa com Humanos da Universidade Federal de Roraima, CAAE n. 54448416.6.0000.5302. O pesquisador contatou a universidade, que autorizou o estudo mediante Termo de Anuência. Com os professores que, comunicados pela coordenação de curso, aceitaram participar, foi agendada reunião com o pesquisador. Nessa oportunidade, os objetivos do estudo foram apresentados sem explicar sobre o procedimento de feedback. Foi esclarecido que os resultados das avaliações não seriam compartilhados e que o professor também deveria manter sigilo.

### Instrumentos

O protocolo de pesquisa foi composto por: (a) TBC adaptado por Henklain (2017) para ser usado por universitários brasileiros na avaliação de professores. Esse instrumento possui 28 itens com uma escala Likert de frequência (“1 = Nunca apresenta” a “5 = Sempre apresenta”). O alpha de Cronbach do TBC na primeira aplicação foi de 0,94 (IC = 0,92 – 0,95), e na segunda foi de 0,95 (IC = 0,93 – 0,96); (b) Questionário de Caracterização do Estudante cujo intuito foi coletar informações sociodemográficas para descrever a amostra com relação à idade, gênero, escolaridade e se o participante é ou não pessoa com deficiência; (c) Protocolo de *Feedback* composto por três tabelas: a primeira com os resultados de relevância dos itens do TBC (ver Henklain, 2017), a segunda com os resultados da avaliação do professor feita pelos discentes neste estudo, e

a terceira com os resultados da autoavaliação obtidos neste estudo. Havia, ainda, uma questão sobre se o professor concordava com os resultados e, em caso positivo, o que considerava que poderia aperfeiçoar em sua prática; e para os participantes do Grupo 3, (d) Plano de ação, que consistiu em dois itens que pediam ao professor para que, considerando os resultados do feedback, selecionasse uma qualidade que considerava mais fácil de ser aperfeiçoada e, então, escrevesse comportamentos que poderia desempenhar em sala de aula correspondentes a essa qualidade. A Figura 1 exibe o Protocolo de Feedback (PF) e os itens do Plano de Ação (PA).

RELATÓRIO DE FEEDBACK BASEADO NO TEACHER BEHAVIOR CHECKLIST (TBC)					
Grau de relevância das qualidades do TBC segundo professores e universitários					
Itens	Brasil* Média (DP)		Itens	Brasil* Média (DP)	
	Professor	Universitário		Professor	Universitário
01. Acessível	5,84	6,06	15. Preparado(a)	6,47	6,32
02. Atencioso(a)	6,49	6,08	16. Usa informações atuais	6,32	6,29
03. Autoritativo(a)	5,41	4,74	17. Postura profissional	5,35	4,59
04. Confiante	6,63	6,38	18. Promove discussões	6,01	5,52
05. Criativo(a)	6,38	6,15	19. Promove postura crítica	6,44	6,07
06. Comunicador(a) eficaz	6,60	6,33	20. Fornece <i>feedback</i>	6,40	6,32
07. Encoraja os alunos	6,30	5,87	21. Pontual	6,13	5,97
08. Entusiasmado(a)	6,56	6,18	22. Estabelece vínculo	5,67	5,18
09. Estabelece objetivos	6,46	5,98	23. Avalia de modo justo	6,12	6,25
10. Flexível	6,04	6,00	24. Respeitoso(a)	6,76	6,59
11. Bom(a) ouvinte	6,40	6,13	25. Sensível e persistente	6,12	5,99
12. Atitude alegre	5,21	4,93	26. Buscar ser melhor	6,37	6,27
13. Humilde	6,29	6,06	27. Hábil no uso de TIC's	5,65	5,33
14. Domina o tema	6,65	6,50	28. Compreensivo(a)	5,92	5,96

Nota. \* = Dados de 278 universitários e 161 professores de universidades públicas e privadas brasileiras; **Céulas na cor cinza (■)**: indicam os maiores valores de média de relevância atribuídos por professores ou estudantes; **Céulas na cor verde (■)**: indicam as qualidades com maior relação entre controle pelo professor e impacto sobre o aprendizado.

Resultados obtidos pelo(a) professor(a) na avaliação pelo discente			
Itens	Média	Itens	Média
01. Acessível	4,58	15. Preparado(a)	5,00
02. Atencioso(a)	5,00	16. Usa informações atuais	4,42
03. Autoritativo(a)	4,25	17. Postura profissional	4,92
04. Confiante	4,83	18. Promove discussões	4,75
05. Criativo(a)	4,75	19. Promove postura crítica	4,75
06. Comunicador(a) eficaz	4,92	20. Fornece <i>feedback</i>	4,75
07. Encoraja os alunos	5,00	21. Pontual	4,92
08. Entusiasmado(a)	4,92	22. Estabelece vínculo	4,83
09. Estabelece objetivos	4,75	23. Avalia de modo justo	4,92
10. Flexível	4,67	24. Respeitoso(a)	4,92
11. Bom(a) ouvinte	4,75	25. Sensível e persistente	4,83
12. Atitude alegre	4,92	26. Buscar ser melhor	4,50
13. Humilde	5,00	27. Hábil no uso de TIC's	3,58
14. Domina o tema	4,92	28. Compreensivo(a)	4,83

Nota. **Céulas na cor azul (■)**: indicam os destaques do professor e que ele deve manter; **Céulas pintadas de amarelo (■)**: indicam qualidades que o professor deve refletir se precisa aperfeiçoar.

Resultados obtidos pelo(a) professor(a) na autoavaliação			
Itens	Média	Itens	Média
01. Acessível	3,50	15. Preparado(a)	4,00
02. Atencioso(a)	4,00	16. Usa informações atuais	3,00
03. Autoritativo(a)	3,00	17. Postura profissional	5,00
04. Confiante	5,00	18. Promove discussões	4,00
05. Criativo(a)	4,00	19. Promove postura crítica	4,00
06. Comunicador(a) eficaz	3,50	20. Fornece <i>feedback</i>	4,00
07. Encoraja os alunos	5,00	21. Pontual	4,00
08. Entusiasmado(a)	4,00	22. Estabelece vínculo	5,00
09. Estabelece objetivos	4,00	23. Avalia de modo justo	5,00
10. Flexível	5,00	24. Respeitoso(a)	5,00
11. Bom(a) ouvinte	4,00	25. Sensível e persistente	5,00
12. Atitude alegre	3,70	26. Buscar ser melhor	2,00
13. Humilde	3,80	27. Hábil no uso de TIC's	3,00
14. Domina o tema	5,00	28. Compreensivo(a)	5,00

Nota. **Céulas na cor azul (■)**: indicam os destaques do professor e que ele deve manter; **Céulas pintadas de amarelo (■)**: indicam qualidades que o professor deve refletir se precisa aperfeiçoar.

a. Você considera que precise aperfeiçoar alguma qualidade dentre as 28 apresentadas? Se sim, qual(is)? Caso contrário, justifique a sua resposta.

b. O que você considera que possa ser aperfeiçoado na avaliação de professores com o TBC e/ou neste documento de feedback?

**Figura 1.**  
Modelo de Protocolo da Feedback e Plano de Ação.

## Procedimento de coleta de dados

Após dois meses do início do semestre, o pesquisador combinou dia e horário para fazer a coleta nas turmas de cada professor participante. Foram apresentados os objetivos do estudo aos alunos, cuja participação requeria assinatura do TCLE. Em seguida, o TBC foi aplicado. Foi assegurado aos alunos que o professor só teria acesso ao resultado geral da turma. Durante a aplicação o professor não ficou na sala de aula. Ele foi para o seu gabinete realizar a autoavaliação com o TBC. Na segunda aplicação, o procedimento foi o mesmo. O período entre as duas avaliações foi, no mínimo, de 29 dias (primeiro dia da 2ª aplicação menos último dia da 1ª) e, no máximo, de 71 (último dia da 2ª aplicação menos primeiro dia da 1ª). No teste-reteste com os estudantes, o TBC apresentou evidência favorável de fidedignidade:  $r(\text{unicaudal}) = 0,78$ ;  $p < 0,001$ ;  $R^2 = 0,60$ ;  $N = 74$ .

*Autoavaliação e Feedback:* Após a primeira aplicação, foi realizada uma sessão de feedback com os docentes dos grupos G2 e G3, que consistia na apresentação das qualidades consideradas mais importantes no TBC segundo professores e universitários brasileiros, seguida dos resultados da avaliação discente e da autoavaliação. Após examinar esses dados, o professor era questionado sobre sua concordância com aqueles dados e se considerava que poderia aperfeiçoar a sua prática. Em caso positivo, o docente era convidado a escrever as qualidades que poderia alterar, sem que nenhuma ação fosse determinada nesta pesquisa para essa modificação de prática.

*Plano de ação:* No G3, os professores participantes realizavam a mesma sequência descrita no tópico anterior, no entanto, após a primeira aplicação do TBC, os professores deste grupo eram convidados a elaborar um plano de ação. Nesse momento, logo após a apresentação do feedback supracitado, o professor era solicitado a escolher uma qualidade do TBC que considerasse mais simples de ser aperfeiçoada e, concomitantemente, de maior impacto no aprendizado. Em seguida, ele deveria escrever como pretendia agir para melhorar a percepção dos estudantes sobre a qualidade escolhida. O professor ficou livre para escolher a qualidade que julgasse mais simples e de maior impacto.

*Devolutiva:* Após finalizadas as duas aplicações do TBC, foi agendado com os seis professores, de forma individual, uma sessão de devolutiva. Nessa ocasião foi explicada a divisão dos docentes em grupos, o que foi feito em cada um deles e todos os resultados. Foi indagado aos professores os motivos que poderiam ter influenciado nos

resultados da avaliação e sua opinião sobre o estudo, uma vez que poderiam apresentar análises sobre seu próprio comportamento em relação aos dados obtidos.

### **Procedimento de análise de dados**

Para as duas aplicações, foram calculadas as medianas, os intervalos interquartil, as médias e os desvios-padrão de cada um dos três professores em relação aos 28 itens do TBC, bem como as diferenças entre a média da segunda avaliação e da primeira. Valores positivos de diferença eram indicativos de aumento da pontuação obtida no TBC (i.e., melhora no desempenho), valores iguais a zero indicavam manutenção e valores negativos indicavam piora. Foi realizado o teste dos postos sinalizados de Wilcoxon apenas com os dados dos 74 participantes que participaram das suas avaliações. Esses dados foram analisados em conjunto e separados por grupos (G1, G2 e G3) – nesse último caso foi adotada a correção de Bonferroni, conforme sugestão de Field (2009), sendo 1,6% o critério de significância. Essa estatística não paramétrica foi utilizada porque no teste de Shapiro-Wilk os dados das duas avaliações foram estatisticamente significativos ( $p < 0,05$ ).

### **Resultados e Discussão**

Os resultados das duas avaliações foram sintetizados na Figura 2. Foram calculadas as médias por item e a geral em cada avaliação, bem como o desvio-padrão para verificação da dispersão dos dados.

I.	G1 – Controle				G2 – Feedback				G3 – Feedback + Plano			
	P1		P4		P2		P5		P3		P6	
	MAP (DP)		MAP (DP)		MAP (DP)		MAP (DP)		MAP (DP)		MAP (DP)	
	1ª	2ª										
1	4,43 (0,79)	4,57 (0,79)	4,17 (0,86)	4,63 (0,62)	4,20 (1,00)	4,33 (1,20)	4,62 (0,59)	4,41 (1,06)	4,62 (0,90)	4,30 (0,82)	4,54 (0,71)	4,41 (1,05)
2	4,57 (0,53)	4,86 (0,38)	4,33 (0,91)	4,81 (0,40)	4,10 (1,20)	3,81 (1,33)	4,81 (0,51)	4,88 (0,33)	5,00 (0,00)	5,00 (0,00)	4,54 (1,10)	4,64 (0,90)
3	5,00 (0,00)	4,71 (0,49)	3,50 (1,25)	4,00 (0,82)	4,40 (1,12)	4,76 (0,54)	<u>3,71</u> (0,96)	<u>4,41</u> (0,71)	<u>4,31</u> (1,14)	<u>4,50</u> (0,97)	4,08 (1,16)	4,36 (1,05)
4	4,86 (0,38)	5,00 (0,00)	4,44 (1,10)	4,56 (0,51)	4,70 (0,46)	4,81 (0,60)	4,62 (0,59)	4,53 (0,72)	4,85 (0,39)	4,90 (0,32)	4,15 (1,19)	4,41 (1,05)
5	4,43 (0,98)	4,71 (0,49)	2,89 (1,02)	3,38 (0,81)	<u>3,45</u> (1,25)	<u>3,81</u> (0,98)	4,24 (0,77)	4,18 (0,88)	4,77 (0,62)	5,00 (0,00)	3,81 (1,23)	4,14 (1,25)
6	5,00 (0,00)	5,00 (0,00)	4,50 (0,71)	4,75 (0,45)	4,10 (1,31)	4,38 (0,86)	4,67 (0,66)	4,76 (0,44)	4,92 (0,29)	5,00 (0,00)	4,35 (1,06)	4,50 (1,01)
7	4,00 (1,00)	4,43 (0,98)	4,44 (1,04)	4,63 (0,62)	<u>3,55</u> (1,36)	<u>4,05</u> (1,16)	4,33 (0,8)	4,71 (0,77)	5,00 (0,00)	5,00 (0,00)	4,38 (0,98)	4,27 (1,16)
8	4,57 (0,53)	4,57 (0,79)	4,00 (0,97)	4,19 (0,83)	4,00 (0,86)	4,29 (0,96)	4,00 (1,00)	4,53 (0,62)	4,92 (0,29)	4,90 (0,32)	4,35 (0,89)	4,41 (1,05)
9	5,00 (0,00)	4,71 (0,49)	4,39 (0,70)	4,56 (0,63)	4,25 (0,90)	4,48 (0,75)	4,43 (0,81)	4,59 (0,51)	4,77 (0,45)	4,90 (0,32)	4,46 (0,99)	4,27 (0,98)
10	3,29 (0,95)	4,14 (1,07)	4,06 (1,00)	4,13 (1,02)	3,45 (1,29)	3,38 (1,16)	4,48 (0,81)	4,65 (0,49)	4,69 (0,89)	4,80 (0,42)	4,19 (1,06)	4,45 (1,01)
11	4,57 (0,53)	4,71 (0,49)	4,28 (1,02)	4,25 (0,68)	3,85 (0,96)	4,19 (1,08)	4,81 (0,4)	4,71 (0,47)	4,69 (0,62)	5,00 (0,00)	4,5 (0,76)	4,27 (0,98)
12	3,43 (0,98)	4,29 (0,76)	3,44 (1,20)	4,06 (0,85)	3,50 (1,12)	3,52 (1,17)	<u>3,95</u> (1,07)	<u>4,59</u> (0,71)	4,77 (0,29)	5,00 (0,00)	4,19 (0,85)	4,18 (1,18)
13	4,57 (0,53)	4,57 (0,53)	4,39 (0,98)	4,81 (0,4)	3,65 (1,38)	3,86 (1,15)	4,67 (0,66)	4,71 (0,59)	5,00 (0,00)	5,00 (0,00)	4,58 (0,76)	4,68 (0,78)
14	4,71 (0,49)	5,00 (0,00)	4,06 (1,11)	4,63 (0,62)	4,65 (0,91)	4,81 (0,51)	4,62 (0,59)	4,71 (0,59)	4,92 (0,29)	4,80 (0,63)	4,46 (1,03)	4,45 (1,06)
15	5,00 (0,00)	5,00 (0,00)	4,11 (0,9)	4,44 (0,73)	4,7 (0,56)	4,57 (0,98)	4,19 (0,68)	4,53 (0,62)	5,00 (0,00)	5,00 (0,00)	4,08 (1,13)	4,09 (0,92)
16	4,71 (0,76)	4,86 (0,38)	3,56 (1,25)	3,63 (0,96)	4,05 (0,94)	4,43 (0,81)	4,48 (0,68)	4,71 (0,47)	<u>4,46</u> (0,90)	<u>4,60</u> (0,52)	<u>3,73</u> (1,25)	<u>3,95</u> (1,29)
17	5,00 (0,00)	4,86 (0,38)	4,83 (0,51)	4,94 (0,25)	4,90 (0,44)	4,86 (0,36)	4,81 (0,40)	4,94 (0,24)	4,92 (0,29)	5,00 (0,00)	4,81 (0,40)	4,55 (0,96)
18	4,57 (0,79)	4,71 (0,49)	2,94 (1,06)	3,75 (0,86)	4,15 (1,08)	4,19 (0,87)	4,48 (1,03)	4,65 (0,61)	4,77 (0,62)	5,00 (0,00)	3,96 (1,15)	4,27 (1,08)
19	4,86 (0,38)	4,86 (0,38)	3,50 (1,04)	3,56 (1,09)	3,80 (1,20)	4,48 (0,93)	4,48 (0,68)	4,53 (0,62)	4,77 (0,62)	4,80 (0,42)	3,85 (1,08)	4,05 (1,13)
20	4,43 (0,53)	5,00 (0,00)	3,56 (1,25)	3,94 (1,12)	4,05 (1,14)	4,29 (1,10)	4,19 (0,87)	4,59 (0,62)	4,77 (0,45)	4,80 (0,42)	<u>4,00</u> (1,06)	<u>4,27</u> (1,28)
21	4,86 (0,38)	4,86 (0,38)	4,11 (0,76)	4,19 (0,66)	4,40 (0,75)	4,86 (0,36)	4,05 (0,74)	3,76 (0,90)	4,92 (0,29)	4,80 (0,63)	<u>3,42</u> (1,14)	<u>3,68</u> (1,09)
22	3,71 (1,11)	4,00 (0,58)	3,56 (0,92)	4,25 (0,68)	<u>3,30</u> (1,16)	<u>3,38</u> (1,24)	4,05 (1,02)	4,65 (0,61)	4,85 (0,58)	5,00 (0,00)	4,08 (1,13)	4,23 (1,27)
23	4,43 (0,79)	4,57 (0,53)	4,17 (1,04)	4,56 (0,63)	3,9 (1,12)	3,81 (0,98)	4,14 (0,65)	4,65 (0,49)	4,92 (0,29)	5,00 (0,00)	4,19 (1,06)	4,36 (1,14)
24	3,57 (1,51)	4,71 (0,49)	4,22 (1,17)	4,75 (0,45)	4,15 (1,08)	4,05 (1,32)	4,90 (0,30)	5,00 (0,00)	4,92 (0,29)	5,00 (0,00)	4,50 (0,99)	4,73 (0,70)
25	3,86 (1,68)	5,00 (0,00)	4,11 (0,83)	4,25 (0,93)	3,50 (1,03)	3,24 (1,3)	4,48 (0,68)	4,65 (0,61)	4,85 (0,39)	4,80 (0,63)	4,27 (1,15)	4,36 (1,00)
26	4,57 (0,79)	4,57 (0,53)	3,33 (1,03)	3,94 (1,24)	4,00 (0,92)	4,48 (0,75)	4,10 (0,94)	4,59 (0,62)	4,54 (1)	4,60 (0,97)	4,27 (1,15)	4,32 (1,21)
27	4,57 (0,53)	4,71 (0,49)	2,78 (1,11)	3,25 (1,13)	4,45 (0,98)	4,67 (0,66)	4,67 (0,66)	4,88 (0,49)	<u>3,69</u> (1,44)	<u>4,30</u> (1,06)	4,65 (0,75)	4,77 (0,69)
28	4,00 (0,82)	4,71 (0,49)	4,33 (0,84)	4,44 (0,89)	3,45 (1,33)	3,95 (0,97)	4,52 (0,6)	4,76 (0,44)	4,77 (0,39)	5,00 (0)	4,62 (0,8)	4,41 (1,14)
M.	<b>4,45</b> (0,63)	<b>4,70</b> (0,44)	<b>4,06</b> (1,03)	<b>4,20</b> (0,93)	<b>4,76</b> (0,49)	<b>4,85</b> (0,30)	<b>3,93</b> (0,98)	<b>4,26</b> (0,75)	<b>4,41</b> (0,72)	<b>4,62</b> (0,58)	<b>4,25</b> (1,00)	<b>4,34</b> (1,05)
Dif.	<b>0,26</b>		<b>0,18</b>		<b>0,09</b>		<b>0,33</b>		<b>0,21</b>		<b>0,09</b>	

Figura 2.

Comparação entre os resultados das duas aplicações do TBC.

*Nota.* DP = Desvio Padrão; I. = Itens do TBC; M. = Média; Dif. = Diferença entre as médias gerais das duas avaliações; Sublinhado simples = Qualidade que o professor concorda que deveria melhorar; Sublinhado duplo = Qualidade selecionada pelo G3 no Plano de Ação.

Com base nas médias gerais exibidas na Figura 2, nota-se que as avaliações realizadas pelos estudantes foram positivas desde o início, com médias acima de 3,9 (sendo que o valor máximo era 5 e o menor era 1). Esse resultado é compatível com os achados de Henklain (2017), cujos dados mostram uma tendência de os universitários realizarem avaliações mais positivas do que negativas dos seus professores. Segundo Henklain, seus dados revelaram que mesmo quando a tarefa do estudante foi avaliar o pior professor que já teve na graduação a partir do TBC, não se observaram os menores resultados possíveis na escala (no caso, 1). Além disso, é possível inferir que os professores que aceitaram participar da pesquisa possuem um histórico de aprovação social por parte dos alunos. Pesquisas futuras podem incluir como critério de inclusão de participantes, que existam professores com baixo índice de aprovação dos discentes. Desta forma, será possível verificar a replicabilidade dos dados obtidos nesta pesquisa, assim como as de Henklain (2017).

Verificou-se, ainda, que para todos os participantes, as médias aumentaram entre a primeira e a segunda avaliação. Quando se calcula a mediana, verifica-se que cinco participantes melhoraram e apenas P6 apresentou uma pequena redução, de 4,57 na primeira avaliação para 4,53 na segunda (queda de 0,4). Com relação ao desvio-padrão, exceto no caso de P6, para todos os demais professores essa estatística reduziu entre a primeira e a segunda aplicação, indicando que as pontuações ficaram mais concentradas em torno da média.

Outra forma de avaliar se houve melhora entre as avaliações é utilizar os dados dos 74 universitários que, seguramente, responderam às duas avaliações. A partir deles foi possível calcular o teste dos postos sinalizados de Wilcoxon no qual foi observada uma diferença estatisticamente significativa entre a primeira (*Média* = 4,26; *DP* = 0,61; *Mediana* = 4,45; *IQR* = 0,81) e a segunda aplicação (*Média* = 4,43; *DP* = 0,59; *Mediana* = 4,54; *IQR* = 0,55), tendo sido a média da segunda ligeiramente superior ( $z = 579,500$ ;  $p < 0,001$ ). O tamanho do efeito dessa diferença pode ser considerado, segundo Field (2009), grande ( $r_{pbiss} = -0,520$ ).

Esse aumento na pontuação obtida com o TBC pode estar associado ao efeito que a avaliação produz no sentido de estimular o professor a obter melhores resultados. Apesar desses resultados de melhora, cumpre responder à questão sobre se houve diferença na avaliação recebida pelo professor em função do grupo a que pertencia. Para o exame dessa questão é importante analisar os dados da Figura 2 tanto individualmente,

quanto por grupo. Conforme já destacado, é preciso cautela nas comparações entre os grupos deste estudo.

No caso do G1, que não recebeu qualquer *feedback*, a expectativa era que os professores não apresentassem mudanças de comportamento, ou que essas mudanças fossem pequenas se comparadas a G2 e G3. Contudo, P1 obteve uma diferença de 0,26 entre as duas aplicações, e P4 de 0,18. Essas diferenças indicam que houve aumento de pontuação entre as duas avaliações. O total de mudança, 0,26 mais 0,18, no G1 foi de 0,44. No teste de Wilcoxon foi observada diferença estatisticamente significativa, ao nível de significância de 1,6%, entre a primeira (*Média* = 4,09; *DP* = 0,57; *Mediana* = 4,30; *IQR* = 0,80) e a segunda avaliação (*Média* = 4,40; *DP* = 0,44; *Mediana* = 4,50; *IQR* = 0,63). A média na segunda foi superior ( $z = 21,00$ ;  $p = 0,003$ ). O tamanho do efeito foi grande ( $r_{pbiss} = -0,754$ ).

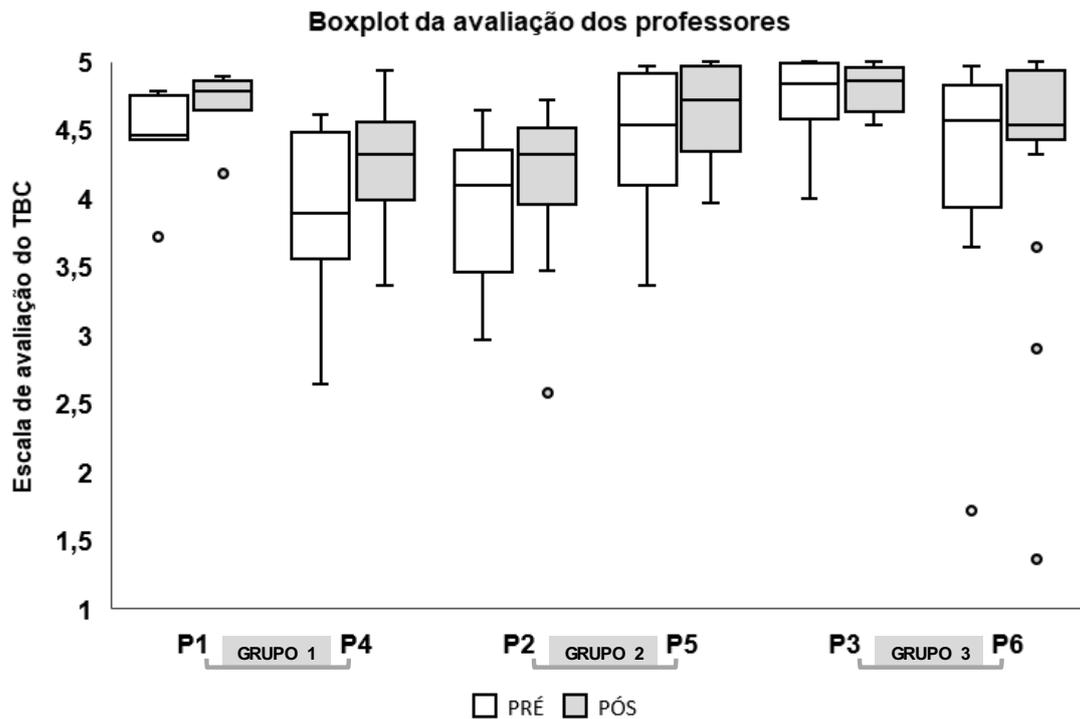
No caso de G2, que recebeu *feedback*, a expectativa era que ocorresse alguma melhora e que fosse superior ao G1. A melhora ocorreu. P2 exibiu uma diferença sutil de 0,09 e P5 de 0,33. Não obstante, o total de mudança alcançada foi de 0,42, ou seja, menor que G1, contrariando as expectativas teóricas. No teste de Wilcoxon foram obtidos resultados não significativos entre a primeira avaliação (*Média* = 4,19; *DP* = 0,56; *Mediana* = 4,25; *IQR* = 0,59) e a segunda (*Média* = 4,38; *DP* = 0,52; *Mediana* = 4,50; *IQR* = 0,52). A média da segunda foi superior ( $z = 104,50$ ;  $p = 0,022$ ). O tamanho do efeito foi médio ( $r_{pbiss} = -0,447$ ).

Finalmente, do G3 era esperado o maior grau de melhora. P3 apresentou uma diferença de 0,21, enquanto P6 mostrou uma diferença pequena de 0,09. No total, o G3 ficou com o menor grau de diferença entre avaliações, 0,30, contrariando as hipóteses desta pesquisa. No teste de Wilcoxon não foi observada diferença estatisticamente significativa entre a primeira (*Média* = 4,46; *DP* = 0,68; *Mediana* = 4,71; *IQR* = 0,54) e a segunda avaliação (*Média* = 4,82; *DP* = 0,43), tendo sido a média da segunda superior ( $z = 100,00$ ;  $p = 0,079$ ). O tamanho do efeito foi entre pequeno e médio ( $r_{pbiss} = -0,333$ ).

Os dados agregados contrariaram as expectativas iniciais. Não obstante, é preciso examinar o que ocorreu com a avaliação relativa às qualidades que os professores do G2 e do G3 concordaram que deveriam melhorar, bem como com as qualidades que o G3 assumiu o compromisso de que deveria melhorar. P2, do G2, concordou que deveria melhorar na Qualidade 5 e foi o que ocorreu: pontuou 3,45 na primeira avaliação e 3,81 na segunda, melhora de 0,36. Esse participante também disse que pretendia melhorar na

Qualidade 7 e alcançou essa melhora, saiu de 3,55 para 4,05, crescimento de 0,5. P2 indicou, ainda, a Qualidade 22, cuja melhora foi sutil, de 3,30 para 3,38, crescimento de 0,08. P5, do G2, relatou que precisaria melhorar nas seguintes qualidades, 3 e 12. Na Qualidade 3 melhorou, saindo de 3,71 para 4,41, diferença de 0,7. Na Qualidade 12 também cresceu, de 3,95 para 4,59, uma diferença de 0,64. Desse grupo, o crescimento de P5 foi maior que o de P2, mas ambos exibiram, no mínimo, manutenção de uma avaliação geral desde o início positiva, e um crescimento maior na pontuação das qualidades que relataram ser necessário aprimorar.

Verifica-se, no G3, que existem dados de qualidades que os professores concordaram que poderiam melhorar e dados da qualidade selecionada para aperfeiçoar no plano de ação. Das qualidades que concordava que poderia melhorar, P3 indicou a 3 e a 27. Na primeira, cresceu de 4,31 para 4,50, diferença de 0,19. Na segunda também melhorou, saindo de 3,69 para 4,30, diferença de 0,61. P6 indicou as qualidades 16 e 21. Na Qualidade 16 apresentou melhora de 3,73 para 3,95, diferença de 0,22, e na Qualidade 21 também houve melhora, de 3,42 para 3,68, diferença de 0,26. Finalmente, com relação à qualidade cujo compromisso foi assumido, verificou-se que P3 selecionou a Qualidade 16 e P6 a 20. P3 apresentou melhora de 4,46 para 4,60, diferença de 0,14, e P6 apresentou melhora de 4,00 para 4,27, diferença 0,27. Nota-se, portanto, que a tendência foi de crescimento em todas as qualidades, sem que a diferença fosse maior no caso da qualidade selecionada no Plano de Ação. Quando comparado aos dados de G2, não foram observadas diferenças substantivas. A Figura 3 resume as informações descritas por meio de boxplots que mostram os valores de cada professor na primeira e na segunda avaliação.



**Figura 3.**

Boxplot da avaliação dos professores.

A Figura 3 deixa clara a melhora de todos os professores, sendo que, exceto P6, os demais apresentaram aumento de mediana. Fica evidente também que P1 já era uma professora muito bem avaliada desde o início. Destaca-se o fato de que, à exceção de uma avaliação, as duas medidas de P1 ficaram entre 4,50 e 5,00, tendo ocorrido uma redução na variabilidade das respostas na segunda avaliação, o que está indicado pelo achatamento da caixa do bloxplot.

Para P4 verificou-se um crescimento grande de desempenho, pois mais de 50% das avaliações na segunda medida ficaram acima da mediana da primeira avaliação. P2 foi uma professora que pediu para o pesquisador se poderia conversar com a turma de modo a conseguir mais exemplos de como poderia melhorar seu desempenho, e seus dados se aproximaram do que ocorreu com P4, pois quase 50% das avaliações na segunda medida superaram a mediana da primeira. Nos casos de P3, P5 e P6, destaca-se o fato que, desde a primeira medida, esses docentes já apresentaram medianas em torno de 4,5, ou seja, quase o limite da escala (5,00). No caso de P5, mais de 25% das avaliações na segunda medida superaram a primeira. Nos casos de P3 e P6 o crescimento foi menor. Mas, para os dois professores, pelo menos, 25% das avaliações feitas ao final do semestre superaram aquelas feitas inicialmente. No caso de P6 verificou-se uma sutil queda na segunda mediana, bem como maior número de outliers.

Para entender esses resultados, destaca-se, inicialmente, a natureza dos itens selecionados pelos professores. Alguns deles referem-se a comportamentos que são mais simples de modificar. Os níveis de complexidade desses itens foram analisados por Henklain e Carmo (no prelo). Por exemplo, pode ser mais simples resolver problemas de pontualidade do que de atitude alegre e positiva. Logo, a depender do comportamento que o professor escolhe, pode ser mais fácil obter sucesso. P3, por exemplo, escolheu melhorar no Item 16 (“Usa informações Atuais”), que tende a ser mais simples por depender apenas do próprio professor. Neste caso, ele propôs como plano de ação utilizar materiais impressos em sala com informações atualizadas sobre a sua disciplina. Por outro lado, na devolutiva, P3 indicou dificuldade em manter a sua estratégia, uma vez que possui grande quantidade de disciplinas a ministrar. Ou seja, essa variável associada ao custo de resposta do professor para mudar e manter a mudança é de fundamental relevância. Além disso, P3 ressaltou que existem condições complexas que não estão ao alcance do professor, principalmente no que concerne às diferentes atribuições do professor do magistério superior (e.g., Leis 8.112/1990 e 12.772/2012) que podem concorrer entre si. De fato, há grande demanda e sobrecarga de trabalho para o professor, o que dificulta a manutenção dos comportamentos considerados de qualidade (Veira-Santos & Henklain, 2017).

Outro aspecto a ser considerado é o engajamento do participante para mudar. Por exemplo, a participante P2 explicitou no feedback da primeira aplicação que precisaria aprimorar-se em dois itens que se referem às relações entre professor e aluno (Itens 2 e 22, ver Figura 2) e acrescentou que gostaria de perguntar aos alunos como poderia melhorar enquanto professora em relação a esses itens. Na devolutiva, ao ver o resultado de seu desempenho, confirmou ter conversado com seus alunos sobre sua prática. Esse dado indica que ao analisar a eficácia de um sistema avaliativo, é fundamental considerar características dos participantes. P2 contou que em sua conversa com os alunos, eles não souberam apontar comportamentos específicos a serem melhorados, apenas indicações generalistas e que, em sua opinião, já fazia. Isso sugere que, dada a complexidade das relações sociais em sala de aula (Del Prette, Del Prette, Garcia, Silva, & Puntel, 1998), pode ser difícil obter um *feedback* claro dos discentes sem a ajuda de um instrumento.

Destaca-se também que, ao analisar um sistema avaliativo, é preciso examinar eventuais divergências nos valores de importância atribuídos aos itens de um instrumento, como o TBC, por alunos e professores. O motivo é que o professor pode priorizar itens

percebidos como menos importantes pelos estudantes e vice-versa. Isso se torna uma variável interveniente uma vez que os novos comportamentos apresentados pelo professor podem ser menos valorizados pelos alunos e, assim, receber menos atenção na reavaliação. Também pode ocorrer que, por acharem pouco relevantes, os estudantes não reforcem e até punam os novos comportamentos do professor, reduzindo a sua frequência (Vieira-Santos & Henklain, 2017).

Importa lembrar que divergências entre avaliadores, capazes de enviar resultados de uma avaliação, podem ocorrer quando os itens de um instrumento não são suficientemente claros em termos dos comportamentos do professor a que se referem (sobre clareza na descrição de comportamentos ver Kubo & Botomé, 2001). Os exemplos de ações do professor disponíveis no TBC ajudam a reduzir o problema da clareza, mas não dispensam um processo cuidadoso de decomposição e derivação de componentes de comportamentos. Henklain, Kienen, Haydu, Carmo e Muniz (no prelo), identificaram no TBC, 85 classes de comportamentos que foram organizadas pelos pesquisadores em cinco classe gerais compostas por 107 classes de ações, 27 classes de eventos antecedentes e 26 classes de eventos subsequentes. Essa profusão de informações a que o TBC se refere explica como alunos e professores podem divergir sobre quais comportamentos refletem cada qualidade, o que impacta em suas avaliações e na percepção que possuem de melhora.

Também merece atenção o fato de que professores que já obtenham escores muito altos no TBC na primeira avaliação, não deveriam continuar no estudo. Considere o caso de P3. Esse participante foi o que obteve melhor avaliação pelos alunos em ambas as aplicações do TBC. Considerando que sua pontuação já está no limite máximo desde a primeira aplicação, a porcentagem de melhora só tenderia a indicar manutenção, enviando a comparação entre os grupos, uma vez que cada um foi composto por apenas dois membros. Pesquisas futuras devem considerar um número maior de participantes por grupos, assim como incluir professores com baixa pontuação em todos eles.

Quanto as análises do desempenho entre participantes, destacam-se características que podem ter influenciado no responder dos alunos e, conseqüentemente, na análise do desempenho dos professores após feedback. Quando se observa os dados de P1 e se compara com P6, os participantes, respectivamente, com maior e menor porcentagem de melhora entre as avaliações, algumas características parecem dar indicadores de variáveis intervenientes. P1 está ministrando uma disciplina de Práticas, enquanto P6 está em uma

disciplina de conteúdos conceituais (de acordo com as ementas das disciplinas). Desta forma, o próprio currículo do curso pode representar uma variável que interfere na avaliação do professor, uma vez que existirão disciplinas que podem ser pior ou melhor avaliadas pelos alunos e, assim, pareadas com o professor que as conduz.

Essa interpretação é reafirmada por P3 após a devolutiva, quando afirma que existem disciplinas que são interdependentes (uma sendo pré-requisito para outra) e, com isso, as qualidades e dificuldades de uma são automaticamente estendidas para a outra. É o possível caso da disciplina ministrada por P6, “Estrutura de Dados II”, ou seja, é possível que os alunos tenham avaliado o professor baseado não apenas no período solicitado pelo pesquisador, mas, também, pelo histórico construído em relação à disciplina de “Estrutura de Dados I”. Tais variáveis não foram controladas neste estudo. Pesquisas futuras devem realizar uma sondagem de conhecimento prévio dos alunos sobre a disciplina e sobre o professor, objetivando identificar possíveis variáveis que podem influenciar na resposta do TBC.

De um ponto de vista psicométrico, importa destacar que o TBC na versão adaptada para o Brasil ainda não possui evidências de invariância da medida entre grupos com diferentes características (ver Damásio, 2013). Logo, é possível que o uso que é feito do TBC por estudantes de Ciências Biológicas não seja o mesmo daqueles que estudam Análise e Desenvolvimento de Sistemas, como também pode ser que existam diferenças em função de outras variáveis como gênero e presença ou não de deficiência. Portanto, esse é um aspecto que precisa ser investigado futuramente.

Com base nas variáveis apresentadas, torna-se evidente a complexidade da avaliação de desempenho do professor, principalmente nos itens que dependem de outras condições para melhoria, que não apenas das ações do professor (Clinton et. al, 2015). Essa realidade apenas fortalece o entendimento de que pesquisas futuras devem continuar investigando meios de otimizar o *feedback* e de criar recursos como o plano de ação, que possam ser úteis para estimular o aperfeiçoamento do professor. Tais recursos devem considerar os níveis de complexidade de variáveis (e.g., relação com a disciplina, condições de trabalho do professor) que afetam o desempenho do professor, assim como o tempo necessário para alcançar a melhora, uma vez que comportamentos mais complexos e que dependam de variáveis que o professor não controla, podem demorar mais para mudar ou para que sejam aprendidos (Bolsoni-Silva, 2002).

Por fim, os dados obtidos possibilitam afirmar que o TBC pode ser um recurso para avaliações formativas, confirmando os apontamentos de Buskist et al. (2002) e Buskist e Keeley (2018), embora ajustes sejam necessários para que a sua contribuição possa ser detectada em um estudo científico. A afirmação sobre a utilidade do TBC considera como possibilidade que, ao ter contato com descrições de comportamentos que representam qualidades de um professor eficaz, mesmo que não haja um plano de ação associado, o professor pode se engajar em comportamentos que criem condições para o aperfeiçoamento profissional, conforme visto com o G1. Essa melhora pode ser maior e requerer menos esforço se pesquisadores conseguirem identificar uma forma prática e eficaz de fornecer feedback, que descreva com maior clareza o que o professor pode fazer. Também é fundamental que sejam propostos mecanismos de reforçamento de comportamentos profissionais de excelência no ensino. A Lei 8.112/90, por exemplo, no Art. 237, prevê a possibilidade de incentivos funcionais contingentes ao desempenho do servidor público, tais como prêmios, medalhas, diplomas de honra ao mérito, condecoração, entre outros.

Além disso, uma ação realizada pelo governo Australiano, denominado *Australian Professional Standards for Teaching*, avaliada em 2015 (Clinton et. al, 2015), apresenta, de modo análogo ao TBC, embora com maior profundidade, diretrizes para a atuação do professor por meio de ações concretas que podem ser colocadas em prática e os recursos necessários para o pleno desenvolvimento dessas ações. Esse documento revela como a educação australiana melhorou a partir do cumprimento de tais diretrizes. Assim, mostrar o que o professor pode fazer e os ganhos que ele poderá obter, pode ser uma estratégia de aperfeiçoamento de um sistema de avaliação de desempenho.

### **Considerações finais**

Esta pesquisa teve como objetivo propor uma estratégia para fornecer feedback para professores – como parte de uma avaliação de desempenho formativa - e investigar evidências de mudanças no comportamento de professores universitários a partir dessa intervenção. Para a coleta de dados e estruturação do feedback, foi utilizado o TBC. Os resultados obtidos foram inconclusivos em relação ao impacto do TBC para o desenvolvimento profissional dos professores. Contudo, dadas as limitações metodológicas deste estudo e os dados de melhora dos professores, bem como resultados

positivos de outras pesquisas com o TBC, é possível supor que esse instrumento possa ser eficaz para auxiliar no aperfeiçoamento profissional do professor.

Destaca-se que era esperado que o grupo de professores que, além de *feedback*, também desenvolveu um plano de ação, obtivesse o melhor desempenho na comparação com os outros grupos. Entretanto, os dados obtidos não indicaram isso. As hipóteses para justificar esse baixo desempenho estão relacionadas as características dos itens do TBC, as condições de trabalho do professor, aos critérios de seleção e distribuição dos participantes entre os grupos e fatores externos a sala de aula que não foram controlados. Desta forma, pesquisas futuras devem incorporar procedimentos que resolvam ou atenuem essas limitações.

Outra questão que merece atenção está relacionada a complexidade existente na prática do professor do magistério superior. Verificou-se que apenas apresentar o que precisa ser melhorado, assim como questionar como poderia melhorar essa característica parece insuficiente para promover mudanças. É importante que pesquisas futuras apresentem estratégias para o professor e que acompanhem a implementação dessa estratégia, pois pode acontecer de o avaliando não ter repertório para saber como melhorar.

Por fim, acredita-se que esta pesquisa traz elementos importantes para o planejamento e a elaboração de avaliações formativas relativas ao ensino de nível superior. Sem reduzir as limitações existentes ou justificá-las, é preciso lembrar que esta pesquisa teve natureza exploratória. Desse ponto de vista, os resultados obtidos foram positivos. Além da construção de um protótipo de avaliação formativa, houve também o teste da sua eficácia. Isso ajudou a identificar as suas limitações e, assim, orientar estudos futuros que tenham maior probabilidade de sucesso no aperfeiçoamento do desempenho profissional de professores do magistério superior.

### **Referências**

- Bolsoni-Silva, A. T. (2002). Habilidades Sociais: Breve análise da teoria e da prática à luz da análise do comportamento. *Interações em Psicologia*, 6(2), 233-242. doi: 10.5380/psi.v6i2.3311
- Brasil. (1990). Lei n. 8.112 de 11 de dezembro de 1990. Dispõe sobre o regime jurídico dos servidores públicos civis da União, das autarquias e das fundações públicas federais. Brasília. Recuperado de <https://bit.ly/3cbTAs1>

- Brasil. (2004). Lei n. 10.861 de 14 de abril de 2004. Institui o Sistema Nacional de Avaliação da Educação Superior – SINAES e dá outras providências. Brasília. Recuperado de <https://goo.gl/Guc1L3>
- Brasil. (2012). Lei n 12.772, de 28 de dezembro de 2012. Dispõe sobre a estruturação do Plano de Carreiras e Cargos de Magistério Federal. Brasília. Recuperado de <https://goo.gl/Whiktv>
- Buskist, W., & Keeley, J. W. (2018). Searching for universal principles of excellence in college and university teaching. *New Directions for Teaching and Learning*, 156, 95-105. doi: 10.1002/tl.20321
- Buskist, W., Sikorski, J., Buckley, T., & Saville, B. K. (2002). Elements of master teaching. In S. F. Davis & W. Buskist (Eds.), *The teaching of psychology: Essays in honor of Wilbert J. McKeachie and Charles L. Brewer* (pp. 30-39). New York: Psychology Press.
- Cassettari, N. (2014). Avaliação de professores: Uma questão de escolhas. *Estudos em Avaliação Educacional*, 25(57), 166-197. doi: 10.18222/ae255720142829
- Clinton, J., Dinham, S., Savage, G., Aston, R., Dabrowski, A., Gullickson, A., Calnin, G., & Arbour, G. (2015). Evaluation of the Implementation of the Australian Professional Standards for Teachers: Final Report. Centre for Program Evaluation, The University of Melbourne.
- Damásio, B. F. (2013). Contribuições da Análise Fatorial Confirmatória Multigrupo (AFCMG) na avaliação de invariância de instrumentos psicométricos. *Psico-USF*, 18(2), 211-220. doi: 10.1590/S1413-82712013000200005
- Del Prette, Z. A. P., Del Prette, A., Garcia, F. A., Silva, A. T. B., & Puntel, L. P. (1998). Habilidades sociais do professor em sala de aula: um estudo de caso. *Psicologia: Reflexão e Crítica*, 11(3), 591-603. doi: 10.1590/S0102-79721998000300016
- Field, A. (2009). *Descobrendo a estatística usando o SPSS*. Porto Alegre: Artmed.
- Hanklain, M. H. O., & Carmo, J. S. (no prelo). Teacher Behavior Checklist: um recurso para favorecer o aperfeiçoamento contínuo de professores. In: A. B. de C. Menezes (org.), *Ensinar e Aprender: Desafios da Educação do Séc. XXI*. Fortaleza: Imagine Tecnologia Comportamental.

- Henklain, M. H. (2017). *Aplicações do Teacher Behavior Checklist à formação e avaliação de professores de nível superior: Contribuições analítico-comportamentais e psicométricas* (Tese de doutorado, Programa de Pós-Graduação em Psicologia, Universidade Federal de São Carlos, São Carlos).
- Henklain, M. H. O., Carmo, J. dos S., & Haydu, V. B. (2018). Contribuições analítico-comportamentais para descrever o repertório de professores universitários eficazes. *Revista Brasileira de Orientação Profissional*, 19(2), 197-207. doi: 10.26707/1984-7270/2019v19n2p197
- Henklain, M. H. O., Carmo, J. S., Haydu, V. B., Muniz, M., Buskist, W., & Keeley, J. W. (no prelo). Teacher Behavior Checklist: Psychometric evidence in teacher assessment by Brazilian college students. Paideia.
- Henklain, M. H. O., Kienen, N., Haydu, V. B., Carmo, J. dos S. (no prelo). Characterization of college teacher behavioral classes based on the Teacher Behavior Checklist. *Temas em Psicologia*, 28(4).
- Keeley, J., Smith, D., & Buskist, W. (2006). The Teacher Behaviors Checklist: Factor Analysis of its utility for evaluating teaching. *Teaching of Psychology*, 33(2), 84-91. doi: 10.1207/s15328023top3302\_1
- Kubo, O. & Botomé, S. P. (2001). Ensino-Aprendizagem: uma interação entre dois processos comportamentais. *Interação*, Curitiba, 5, 123-132.
- Mangiapanello, K. A., & Hemmes, N. S. (2015). An analysis of feedback from a behavior analytic perspective. *The Behavior Analyst*, 38(1), 51-75. doi: 10.1007/s40614-014-0026-x
- Primi, R., Muniz, M., & Nunes, C. H. S. S. (2009). Definições contemporâneas de validade de testes psicológicos. In C. S. Hutz (Org.), *Avanços e polêmicas em avaliação psicológica* (pp. 243-265). São Paulo: Casa do Psicólogo.
- Ribeiro, E. A. (2012). As atuais políticas públicas de avaliação para a educação superior e os impactos na configuração do trabalho docente. *Avaliação: Revista da Avaliação da Educação Superior*, 17(2), 299-316. doi: 10.1590/S1414-40772012000200002
- Silva, M. A. DA, Machado, W. de L., Pilotto, L. M., Backes, B., Zanon, R. B., Machado, P. V., Zoltowski, A. P. C., Vieira, R. V. de A., Endres, R. G., Francalanci, M., Oliveira, V. A. de J., Krug, J. S.A, & Bandeira, D. R. (2017). Construção e estudo de evidências de validade da Escala de Avaliação Docente. *Revista Brasileira de Educação*, 22(70), 690-707. doi: 10.1590/s1413-24782017227035

Uttl, B., White, C. A., & Gonzalez, D. W. (2017). Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation*, 54, 22-42. doi: 10.1016/j.stueduc.2016.08.007

Vieira-Santos, J. & Henklain, M. H. O. (2017). Contingências sociais que dificultam o engajamento do professor universitário em relações de qualidade com seus alunos. *Perspectivas em análise do comportamento*, 8(2), 200-214. doi: 10.18761/PAC.2016.032

Vieira-Santos, J. & Henklain, M. H. O. (no prelo). Instrumentos com evidências psicométricas para avaliar o desempenho de professores universitários: revisão de literatura. *Revista Psicologia para América Latina*.

Submetido: 06/04/2020

Aprovado: 30/04/2020