DOES SOCIAL DESIRABILITY AFFECT FACULTY AND STUDENTS' PERCEPTIONS OF EXCELLENCE IN TEACHING?

A DESEJABILIDADE SOCIAL AFETA AS PERCEPÇÕES DE PROFESSORES E ESTUDANTES SOBRE EXCELÊNCIA NO ENSINO?

Marcelo Henrique Oliveira Henklain
ORCID 0000-0001-9884-8592
Federal University of Roraima, UFRR
Department of Computer Science
Boa Vista, Roraima, Brazil
marcelo.henklain@ufrr.br

Juliana Rosa Lira
ORCID 0000-0001-5563-3022
Federal University of Roraima, UFRR
Department of Psychology
Boa Vista, Roraima, Brazil
juliana.lira@ifrr.edu.br

João dos Santos Carmo
ORCID 0000-0003-3913-7023
Federal University of São Carlos, UFSCar
Department of Psychology
São Carlos, São Paulo, Brazil
joaocarmo.dpsi@gmail.com

Jared Wayne Keeley
ORCID 0000-0002-1437-5810
Virginia Commonwealth University, VCU
Department of Psychology
Richmond, Virgínia, USA
jwkeeley@vcu.edu

Abstract. Social desirability may adversely impact studies of teaching excellence and its evaluation in the classroom. However, to date, no studies have examined to what extent social desirability might impact student or faculty ratings of excellent teaching. This study investigated whether a tendency to provide socially desirable answers was related to relevance ratings of TBC items among 98 faculty and 545 university students from various Brazilian states. The results of this study indicated no association between student responses and social desirability, but a weak relationship for faculty. It is encouraging that social desirability did not seem to impact students' ratings, which would support the valid interpretation of student evaluations of TBC items' relevance; however, there is a clear need for studies that deepen and expand these findings and investigate other variables involved in the influence of social desirability, while also investigating social desirability when using TBC to evaluate teachers' performance.

Keywords: Psychometrics, Educational quality, Teacher evaluation.

Resumo. A desejabilidade social pode impactar negativamente estudos sobre a excelência no ensino e sua avaliação na sala de aula. No entanto, até o momento, nenhum estudo examinou em que medida a desejabilidade social pode afetar as avaliações de excelente ensino feitas por estudantes ou professores. Este estudo investigou se a tendência a fornecer respostas socialmente desejáveis estava relacionada às avaliações de relevância dos itens de TBC entre 98 professores e 545 estudantes universitários de vários estados brasileiros. Os resultados deste estudo indicaram ausência de associação entre as respostas dos alunos e a desejabilidade social, mas uma relação fraca para os professores. É encorajador que a desejabilidade social não pareça impactar as avaliações dos alunos, o que apoiaria a interpretação válida das avaliações dos alunos sobre a relevância dos itens de TBC; no entanto, há uma clara necessidade de estudos que aprofundem e ampliem esses achados e investiguem outras variáveis envolvidas na influência da desejabilidade social, enquanto também investigam a desejabilidade social ao usar TBC para avaliar o desempenho dos professores.

Palavras-chave: Psicometria, Qualidade educacional, Avaliação docente.

1. INTRODUCTION

Excellence in teaching has been the subject of extensive investigation across various countries, institutions, and disciplines, and by different evaluators, including students and faculty members. These studies utilize an educational instrument known as the Teacher

Behavior Checklist (TBC; Buskist et al., 2002; Stigall & Blicoe, 2015; Hermosa-Bosano & Keeley, 2021). Comprising 28 items, the TBC amalgamates qualities and behaviors exhibited by excellent teachers. It can be used to identify key characteristics of excellence in teaching according to a specific population and to evaluate teachers focusing on their professional development (Keeley et al., 2006; Buskist & Keeley, 2018). Since the term technology means the systematic application of scientific knowledge to solve practical problems, instruments like the TBC must be recognized as a form of technology (Andrade & Valentini, 2018; Zanini et al., 2022). They are based on scientific knowledge to solve the problem of measuring psychological phenomena, which are inherently complex due to their mutable, multidetermined, and ephemeral nature.

Therefore, the TBC serves a dual purpose within the realm of educational technology, functioning both as an assessment tool for researchers and educational administrators, and as a guiding framework for teachers who seek to enhance their capabilities of promoting a meaningful and enjoyable learning experience (Henklain et al., 2023). With further empirical support, the integration of TBC into educational systems could enable the automated extraction of information regarding teacher performance, facilitating broader improvement of teachers' performance with reduced manual effort. Hence, we advocate for research focused on establishing the psychometric evidence supporting the efficacy of TBC.

Most TBC studies have focused on evaluating to what extent it is a valid and reliable tool. In the main group of studies, researchers have been trying to determine the convergence degree between students and teachers regarding their perceptions of the TBC's 10 most important qualities (e.g., Buskist & Keeley, 2018; Ismail & Groccia, 2017). This is necessary, for example, to identify what qualities are priorities to promote learning and if there are universal principles for teaching – which could dramatically impact faculty's development and assessment worldwide. Another group of studies – which we focus on here – is concerned with the extent to which teachers and undergraduate students think that the TBC items are relevant to excellence in teaching (e.g., Lammers et al., 2010; Liu et al., 2015). This is a key issue if we want to empirically show that the TBC items represent what those directly involved in the teaching and learning process consider necessary to leverage educational outcomes.

When we obtain data indicating that the items of an instrument measuring a specific phenomenon are recognized by experts in that phenomenon (or individuals who experience it directly) as relevant and appropriate, psychometricians refer to it as evidence of content validity (Cunha, Almeida-Neto, & Stackfleth, 2016). Several studies have documented the content validity of the TBC through the convergence of item ratings across students and faculty, various cultures (Henklain et al., 2018; Keeley et al., 2012; Lammers et al., 2010; Liu, Keeley, & Buskist, 2015), and academic disciplines (Liu, Keeley, & Buskist, 2016). Despite the progress made in TBC studies showing its content validity, there are still problems to address.

One of the main issues requiring investigation is that the studies cited previously are mainly based on verbal reports which could be biased by variables such as social desirability (Henklain et al., 2019). Therefore, a relevant factor that can be analyzed when evaluating the participants' results is that they may present answers according to what they believe the researcher or other important people desire (Gouveia et al., 2009; Ribas et al., 2004). That's our goal with this research.

2. THEORETICAL FRAMEWORK

2.1 Social desirability

Socially desirable responses could be due to different reasons for students and faculty (Henklain et al., 2019; Balan et al., 2022). When students rate the importance of various qualities of excellent teaching, they could rate them highly just because such qualities seem

important for good teaching or because other people say they are. Students may also think that they do not have the technical knowledge of how a teacher should be evaluated, and so they assent to the list of characteristics. Students might view the TBC favorably simply to please the teacher who is using it, as they value maintaining a positive relationship with their teacher.

For faculty, the reasons could be different. They may feel uncomfortable saying that TBC qualities are not important, even if they personally do not value them, because of the perception that their colleagues do value these qualities. For example, a particular teacher may not prioritize establishing strong relationships with their students, but recognize that his/her colleagues may consider it important. Additionally, just like students, faculty may believe that because the TBC was developed by researchers, it must represent an accurate definition of teaching since the researchers must know about what they are investigating.

In summary, social desirability may arise for several reasons and could impact participants' ratings on the TBC in multiple ways like when one is asked to evaluate to what extent TBC items are important to evaluate excellence in teaching (Paulhus, 1984; Ribas et al., 2004; Gouveia et al., 2009; Gnambs & Kaspar, 2017). Because of this, without specific data, it could be argued that the strong content validity evidence for TBC item relevance is, at least in part, due to participants' efforts to conform to perceived social expectations. Our study rationale is that if we want to develop robust educational technologies to assess teachers' performance and foster their professional capabilities, our first step is to identify what we mean by excellent performance. The TBC already does that, but we need to rule out social desirability as an explanation for its content validity evidence before we assume that its items do reflect students' and faculty's perceptions of teaching excellence.

Social desirability has been studied for a long time (e.g., Paulhus, 1984), and researchers have developed instruments to measure it. One of the most used instruments in the USA and Brazil is the Marlowe-Crowne Scale, which was adapted for use in Brazil by Gouveia et al. (2009). Gouveia et al. (2009) found in two studies that a 20-item version of the Marlowe-Crowne Social Desirability Scale (MCSDS) is more appropriate for the Brazilian population.

2.2 Excellence in teaching

There is no consensus in the scientific literature regarding the characteristics of an excellent teacher (Guerra et al., 2021; Barreto et al., 2023). Therefore, it is crucial to continue investigating this issue. In this study, we adopt two starting points for this investigation. The first concerns the social function of the teacher.

We adopt the perspective, aligned with analytic-behavioral theory (Kienen et al., 2021), that the teacher's social function is to teach, encompassing all the work involved in creating conditions that promote valuable learning for students. This work begins with lesson planning, including the setting of learning objectives that, when achieved, can help students better address their social realities. It then involves constructing teaching and assessment conditions based on these objectives and proceeds to the implementation and refinement of teaching based on assessment results.

Teaching thus involves facilitating learning in a process where educational stimuli do not assume aversive value (Henklain et al., 2020). However, what behaviors constitute a teacher's repertoire and can aid in teaching excellence? We believe that the Teacher Behavior Checklist (TBC) may be useful in this investigation, potentially assisting researchers in this examination and educators in assessing teaching excellence in practical classroom settings (Buskist & Keeley, 2018; Keeley et al., 2024). Therefore, we will present the characteristics and existing studies on this instrument in the next section, highlighting a gap in the scientific literature on excellent teachers related to examining the relationship between social desirability and student evaluations of teaching.

3. RELATED STUDIES

Since its inception in 2002 (Buskist et al., 2002), the TBC has typically been investigated through two types of research: (a) evaluation of teachers by their students (e.g., Keeley et al., 2006; Henklain et al., 2020) or in self-assessments (e.g., Guerra et al., 2021), including the examination of the TBC's validity evidence based on its internal structure; and (b) the use of the TBC to characterize perceptions of diverse groups, such as teachers and students from different countries, regarding to what extent its items are relevant to excellence in teaching and what they consider to be the 10 most important qualities for an excellent teacher, thereby conducting comparisons of convergences between these groups (e.g., Buskist & Keeley, 2018; Hermosa-Bosano & Keeley, 2021; Ismail & Groccia, 2022; Barreto et al., 2023).

Regarding the first line of research, the work of Keeley et al. (2006) stands out as the first contribution. The researchers investigated the psychometric properties of the TBC when used by American university students to evaluate their teachers. The findings led to the proposal of a two-factor model for the TBC, composed of "Caring and Supportive" and "Professional Competence and Communication Skills". In Brazil, Henklain et al. (2020) conducted a cross-cultural adaptation of the TBC and successfully identified a factor structure like that proposed based on American samples. Both studies found favorable evidence of the TBC's validity and reliability.

Concerning the second line of research, we must mention the extensive literature review conducted by Buskist and Keeley (2018), encompassing 18 studies with 1765 teachers and 6290 students. Among the teachers, 100% agreed that knowledgeable about subject matter, enthusiasm for teaching, and promotion of critical thinking are essential qualities for an excellent teacher. Students, on the other hand, reached complete agreement only on the quality of content mastery, although 86% also recognized enthusiasm for teaching. In a more recent study, Ismail and Groccia (2022) found that American and internationally educated teachers agreed on nine out of 10 qualities, with content mastery and enthusiasm for teaching being the top two qualities for both groups.

In addition to typical studies, there are TBC studies that use it to develop and evaluate models that explain how teacher characteristics relate to student satisfaction (e.g., Geier, 2021) and teacher characteristics and student effort (e.g., Geier, 2022). There are also adaptations of the TBC for new contexts, such as online teaching (Keeley & Wilson-Doenges, 2024). Notwithstanding, according to our literature review, only two studies tried to deal with the possible biases in verbal reports about the relevance of TBC items.

Henklain et al. (2019) selected six items of the TBC and used the Implicit Relational Assessment Procedure (IRAP) to measure the time in milliseconds for each one of the 47 participants to relate the six TBC qualities and their antonyms to the stimuli "good" and "bad teacher". The IRAP is an implicit measure less affected by biases like social desirability because the participant must answer correctly and rapidly within a time constraint. In this task, the quicker the participant relates two stimuli, the stronger the evidence that they learned this relation prior to the experiment. Researchers found that participants associated "good teacher" with TBC qualities faster than with their opposites, providing evidence that this association existed before the experiment.

Nonetheless, despite the contributions of Henklain et al. (2019), this study has three main caveats that warrant consideration. First, researchers allotted a relatively long response time (2 seconds) for participants, which may have diminished the IRAP's effectiveness in revealing pre-experimental learning histories. Additionally, the IRAP preparation included pre-block rules on how participants should relate stimuli, which Finn et al. (2016) demonstrated can significantly influence IRAP scores, potentially confounding the reflection of participants' pre-experimental histories. The second caveat is that the study included only undergraduate students, and there were no TBC data regarding faculty and social desirability. Finally, the

researchers did not explicitly measure social desirability within their sample. Consequently, they cannot provide evidence on whether, or to what extent, social desirability affects TBC ratings, nor whether this impact is similar between students and faculty. Since we use the TBC to evaluate teachers, investigate excellence in teaching, and even make pedagogical decisions, it is reasonable to say that we must understand the characteristics of students and faculty ratings and which variables may influence them.

It is important to note that Balan et al. (2022) replicated Henklain et al. (2019) and confirmed the initial study's findings, thereby strengthening the evidence supporting the content validity of the six TBC items tested. However, while this replication was significant, it did not address the limitation of not including a direct measure of social desirability.

We believe there is substantial reason for social desirability to influence the evaluation of TBC items' relevance and, especially, teaching evaluations, given the sensitivity of the topic. One may feel uncomfortable with the idea of giving low scores to items of an instrument since the researchers have worked hard to develop it. This discomfort tends to be higher when people are asked to evaluate teachers. They may fear the potential consequences for themselves and the teachers regarding the evaluation results. Socially desirable responding can be technically conceptualized as a form of response bias (Gouveia et al., 2009; Gnambs & Kaspar, 2017). Specifically, it is a source of variability in student ratings of their teachers that is not associated with the target construct (i.e., the actual quality of instruction or, in our study, the relevance of the TBC item). Instead, it is variability due to a characteristic of the rater.

A variety of other kinds of response bias have been shown to affect student ratings of their teachers. For example, the grade that a student expects to receive in a class is significantly related to how highly they rate their teacher (Gump, 2007). Although the extent and nature of the grade effect are vigorously debated (Centra, 2003; Smith et al., 2011), the presence of response bias is well documented. Similarly, student ratings are often subject to a halo effect, whereby students' global opinion of how much they like their teacher or otherwise positive attitudes towards them inflates their ratings of that teacher (Feeley, 2002; Keeley et al., 2013). In the same way, we hypothesize that the degree to which a person is likely to respond in a socially desirable way will be related to how highly they rate the qualities of their teachers. However, to date, no studies have examined the impact of social desirability directly on student evaluations of their teachers with the TBC.

Considering this scenario and the lack of studies that investigate the relationship between social desirability and responses to the TBC, we consider that it would represent an important goal to measure the social desirability of faculty and students after they evaluated TBC items. The present study investigated how often teachers exhibit the TBC qualities and corresponding behaviors according to Brazilian faculty and undergraduates, and to what extent this evaluation is associated with the tendency to hold the bias of social desirability.

4. METHODS

4.1. Participants

Participants were 545 undergraduate students and 98 faculty from various disciplines and Brazilian states. In the student sample, there were 182 men, 362 women, and one person who identified differently. The average age of the students was 23.8 (*SD* = 5.8). Our undergraduate students were mainly from the Northeast region of Brazil. The students that participated were from 131 different disciplines, mainly classified as Applied Social Sciences. Within the faculty sample, there were 43 men and 55 women. The faculty average age was 38.4 years (*SD* = 10.1). Most of our teachers were also from the Brazilian Northeast region and were teaching disciplines related to the Applied Social Sciences. Table 1 shows the characterization of our participants.

Table 1. Participant's characteristics.

Variables	Undergraduate	Teachers
Gender		
Male	182	43
Female	362	55
Other	1	0
Age		
Mean (SD)	23.8 (5.8)	38.4 (10.1)
Region		
Northeast	35.4%	37.8%
South	24.2%	21.4%
Southeast	22.9%	12.2%
North	13.6%	17.3%
Midwest	3.3%	10.2%
Federal District	0.6%	1.0%
Discipline*		
Applied Social Sciences	16.8%	23.3%
Linguistics, Arts and Letters	9.9%	7.0%
Exact Sciences and Earth Sciences	7.6%	11.6%
Health Sciences	7.6%	14%
Engineering Sciences	6.9%	11.6%
Agricultural Sciences	6.9%	4.7%
Human Sciences	5.3%	7.0%
Biological Sciences	3.8%	4.7%
Multidisciplinary	0.8%	0.0%
Other	34.4%	16.3%
Total Sample (*n*)	**545**	**98**

Source: Author (year)

Note. We categorized each discipline according to criteria of a Brazilian government agency called Coordination of Improvement of Higher-Level Personnel (CAPES).

The procedures in this study were approved by the Brazilian Platform for Ethical Committees (Brazilian Platform, CAAE 54448416.6.0000.5302). Informed consent was obtained from all individual participants included in the study.

4.2. Instruments

There were two instruments used in this study: the Teacher Behavior Checklist (TBC) and the Marlowe-Crowne Social Desirability Scale – Abbreviated version (MCSDS-20). The TBC is an instrument used in teacher evaluation that has 28 items rated on a Likert frequency scale ("1 = Never exhibit" to "5 = Always exhibit"). It was translated and adapted into Brazilian Portuguese by Henklain et al. (2020), and the complete instrument can be accessed through their work. Figure 1 illustrates the 28 qualities of the TBC.

The MCSDS-20, adapted by Gouveia et al. (2009), is used to measure a person's tendency to respond in a socially desirable way. The items were sourced from personality instruments and reflect behaviors indicative of a need for approval from others. For each item, the participant must answer true or false.

The two instruments were adapted to a Google Forms format to facilitate online data collection. This Google Form also collected demographic information to describe the sample in terms of age, gender, Brazilian region of residence, and discipline.

Teacher Behavior Checklist (TBC) 28-qualities			
01 Accessible	**08** Enthusiastic about teaching	**15** Prepared	**22** Rapport
02 Approachable	**09** Establishes term goals	**16** Presents current information	**23** Realistic expectations
03 Authoritative	**10** Flexible	**17** Professional	**24** Respectful
04 Confident	**11** Good listener	**18** Promotes class discussion	**25** Sensitive and Persistent
05 Creative	**12** Positive attitude	**19** Promotes critical thinking	**26** Strives to be a better teacher
06 Effective communicator	**13** Humble	**20** Provides feedback	**27** Technologically competent
07 Cares for students	**14** Knowledgeable	**21** Manages class time	**28** Understanding

Figure 1. The 28 Qualities of the TBC.

4.3. Data collection procedure

The invitation to participate was sent via email with a link to the Google Form. Additionally, we used communication apps and social media groups of higher education students and teachers in Brazil to disseminate the research link.

All participants provided informed consent for the research. The study instructed participants to describe how often an excellent teacher exhibits each TBC quality and its corresponding behaviors. Participants were also asked to complete the MCSDS-20 which presented social attitudes that may or may not represent their daily behavior, with no right or wrong answers. All participants answered the instruments in the same order.

4.4. Data analysis procedure

The Shapiro-Wilk normality test ($\alpha = 0.05$) indicated that TBC scores did not follow a normal distribution; given the nature of the ratings, they are almost always skewed. For the MCSDS-20, the scores in the student sample did not follow a normal distribution, whereas the scores in the teacher sample did ($W = 0.98$, $p = 0.11$). However, since the sample was defined by convenience, the TBC scale is ordinal and the MCSDS-20 scale is nominal, we decided to use non-parametric statistics to analyze the data.

The software used was R Studio based on R version 3.4.1 (R Development Core Team, 2017). First, we calculated an average for the 28 TBC items for each participant to serve as a total score (Keeley et al., 2006). For the MCSDS-20, we summed the participants' responses, assigning one point for each response indicative of social desirability. We then calculated the average, median, standard deviation, and interquartile range for both instruments. Additionally, we computed Cronbach's Alpha for the TBC and the Kuder-Richardson-20 coefficient for the MCSDS-20. Mann-Whitney tests were conducted to examine median differences in the TBC

between Group 1 (with MCSDS-20 scores below the median) and Group 2 (with MCSDS-20 scores at or above the median).

5. RESULTS

Table 2 summarizes the main data obtained in this study concerning student and faculty samples. We found high averages in the TBC for both students and faculty (i.e., well above the midpoint of 3). Faculty exhibited a significantly higher score than students (U = 17098, p < 0.001, r = -0.22, n = 643). We found moderate to high scores on the MCSDS-20 for students and faculty. Again, faculty obtained a significantly higher score (U = 16557, p < 0.001, r = -0.22, n = 643) which means a higher tendency to respond in a socially desirable way. The internal consistency of both measures was acceptable, but higher for the TBC than for the MCSDS-20.

Table 2. Summary of main results.

Statistics	TBC		MCSDS-20	
	Students	Faculty	Students	Faculty
Average (Standard-deviation)	3.9 (1.1)	4.4 (0.8)	10.2 (3.2)	12.6 (3.6)
Median (Interquartile range)	4.1 (0.9)	4.5 (0.6)	10 (5.0)	13 (5.0)
75th Percentile	4.54	4.75	13	15
25th Percentile	3.64	4.11	8	10
Cronbach's Alpha	0.96	0.92	---	---
Kuder-Richardson (KR_{20})	---	---	0.64	0.74

Source: Author (year)

To answer our main research question, we compared high (above the median of the group) and low (bellow median) social desirability ratings to the TBC. We found no difference in students' TBC scores for low social-desirability (*TBC Median* = 4.107; *IQR* = 1.000) and high (*Median* = 4.179; *IQR* = 0.786) ratings in the MCSDS-20 (U = 33989, p = 0.236, r = -0.051, n = 545). There was also no association between TBC and MCSDS-20 scores within the student's samples (ρ = 0.047, p > 0.05). However, teachers with high scores in social desirability did have statistically higher TBC scores (*TBC Median* = 4.607; *IQR* = 0.554) than teachers with lower social desirability (*TBC Median* = 4.393; *IQR* = 0.554; U = 883, p = 0.025, r = -0.22, n = 98). There was a small but statistically significant correlation between these variables for teachers (ρ = 0.26, p < 0.05).

6. DISCUSSION

This study aimed to investigate whether the answers provided by faculty and students to the TBC reflect a pattern of response related to social desirability. For students, there was no association between their responses on the TBC and the MCSDS-20. This finding is encouraging as it does not appear that social desirability is substantially related to student ratings of TBC items' relevance. We hope that this finding might also hold for student use of the TBC to evaluate individual teachers, which we should investigate in future studies. We were unable to find any examples of empirical studies that have investigated the impact of social desirability on teacher evaluations with the TBC, although some authors have suggested the importance of studies like ours (e.g., Spooren et al., 2012). Our data also agree with the findings of Henklain et al. (2019), indicating that the relation between TBC items and the

concept of a good teacher is not likely an effect of the participants' intentions to please the researchers or to conform to what participants may think that researchers are expecting.

However, faculty data revealed a relationship between the pattern of responses on the TBC and MCSDS-20, although it was weak. This result suggests that the faculty of this sample may have answered to the TBC seeking to confirm a social expectation about what would be an excellent teacher, even if they do not necessarily believe in it. In this sense, they could have provided socially desirable answers (Gouveia et al., 2009).

This possibility raises the question of whether other findings regarding faculty's opinions about the TBC also suffered this type of bias, which encourages new studies to answer this question, maybe using implicit measures like Henklain et al. (2019) in combination with the MCSDS-20. On the other hand, it is necessary to consider that the correlation was weak and may not generalize to new samples. It would be important to determine if this effect replicates.

Finally, the sample in this study was selected based on convenience, and data were collected using Google Forms. These two aspects may favor biases. Our sampling method made it possible that individuals who care about teaching were more likely to enroll in the study, thereby misrepresenting the relationship that may exist between social desirability and TBC ratings in the population. Bigger and more diverse samples would be important improvements to future studies. Using Google Forms could also be a source of bias, as completing the study online may predispose a person to respond differently than if they were in person in front of the researcher. However, a meta-analysis has shown that completing psychological measures online versus in-person is not substantially related to social desirability (Gnambs & Kaspar, 2017). Thus, we believe our method may have produced similar effects to an in-person study. Nonetheless, this study represents a useful first step in evaluating the possible effect of social desirability in the evaluation of good teaching.

7. CONCLUSION

We found no evidence that social desirability may have affected students' evaluations of what constitutes the profile of an excellent teacher. However, our data showed that this variable probably impacted faculty responses to the TBC. Given that faculty perceptions of teaching excellence are crucial for understanding the effectiveness of TBC as an educational technology for assessing teacher performance and enhancing learning experiences, future research should examine whether our findings hold true across a broader sample. Additionally, we recommend further studies using both scales and implicit measures to evaluate students' and faculty's perceptions of the TBC. It is also essential to conduct experimental investigations to identify the independent variables influencing feedback on TBC items. This approach is vital for a thorough understanding of TBC's content validity.

CONFLICT OF INTEREST

The authors declare they have no conflict of interest.

FUNDING

This research was part of the scientific program of the National Institute of Science and Technology on Behavior, Cognition, and Teaching (INCT-ECCE), supported by the São Paulo Research Foundation (FAPESP, grants #2008/57705-8 and #2014/50909-8) and the National Council for Scientific and Technological Development (CNPq, Grants #573972/2008-7, and #465686/2014-1).

REFERENCES

Andrade, J. M., & Valentini, F. (2018). Diretrizes para a Construção de Testes Psicológicos: a Resolução CFP n° 009/2018 em Destaque. *Psicologia: Ciência e Profissão*, *38*(spe), 28-39. https://doi.org/10.1590/1982-3703000208890

Balan, R. N., Haydu, V. B., Almeida, J. H., Henklain, M. H. O., & Zacarin, M. R. Z. (2022). Teacher Behavior Checklist e Implicit Relational Assessment Procedure en la evaluación docente. *Revista de Psicología*, 40(1), 553-577. https://doi.org/10.18800/psico.202201.018

Barreto, Y., Urquiza, M., Henklain, M., Almada, N., Leal, S., Silva, H., Lima, G., Silva, S., Almeida, C., & Alves, G. (2023). Quais são as qualidades docentes mais importantes no ensino de informática para mulheres? In *Anais do II Workshop de Aplicações Práticas de Learning Analytics em Instituições de Ensino no Brasil (pp. 108-117)*. Porto Alegre: SBC. https://doi.org/10.5753/wapla.2023.236176

Buskist, W., & Keeley, J. W. (2018). Searching for universal principles of excellence in college and university teaching. *NewColo Directions for Teaching and Learning, 156*, 95-105. https://doi.org/10.1002/tl.20321

Buskist, W., Sikorski, J., Buckley, T., & Saville, B. K. (2002). Elements of master teaching. Em S. F. Davis & W. Buskist (Orgs.), *The teaching of psychology: Essays in honor of Wilbert J. McKeachie and Charles L. Brewer* (pp. 30-39). New York: Psychology Press.

Centra, J. A. (2003). Will teachers receive higher student evaluations by giving higher grades and less course work? *Research in Higher Education*, *44*, 495-518. https://doi.org/10.1023/A:1025492407752

Cunha, C. M., de Almeida-Neto, O. P., & Stackfleth, R. (2015). Principais métodos de avaliação psicométrica da validade de instrumentos de medida [Main psychometric evaluation methods of the validity of measuring instruments]. *Revista de Atenção de Saúde, 14*(47), 75-83. https://doi.org/10.13037/rbcs.vol14n47.3391

Feeley, T. (2002). Evidence of halo effects in student evaluations of communication instruction. *Communication Education*, *51*, 225-236.

Finn, M., Barnes-Holmes, D., Hussey, I., & Graddy, J. (2016). Exploring the behavioral dynamics of the Implicit Relational Assessment Procedure: The impact of three types of introductory rules. *Psychological Record*, *66*(2), 309-321. http://doi.org/10.1007%2Fs40732-016-0173-4

Geier, M. T. (2022). The Teacher Behavior Checklist: The Mediation Role of Teacher Behaviors in the Relationship Between the Students' Importance of Teacher Behaviors and Students' Effort. *Teaching of Psychology*, *49*(1), 14-20. https://doi.org/10.1177/0098628320979896

Gerônimo-Júnior, V. C., Henklain, M. H. O., Carmo, J. S., & Kelley, J. W. (2023). Utility of the Teacher Behavior Checklist beyond Psychology Majors: Replication with Brazilian Physical Education Students. *Acta Colombiana de Psicología*, *26*(1), 214-225. https://doi.org/10.14718/ACP.2023.26.1.14

Gnambs, T., & Kaspar, K. (2017). Socially desirable responding in web-based questionnaires: A meta-analytic review of the candor hypothesis. *Assessment*, *24*, 746-762. https://doi.org/10.1177/1073191115624547

Gouveia, V. V., Guerra, V. M., Sousa, D. M. F., Santos, W. S., & Costa, J. M. (2009). Escala de Desejabilidade Social de Marlowe-Crowne: evidências de sua validade fatorial e consistência interna [Marlowe-Crowne Social Desirability Scale: evidences of its factorial validity and reliability]. *Avaliação Psicológica*, *8*(1), 87-98. Retrieved from http://bit.ly/39mRvqK

Guerra, V. M., Castello, N. F. V., Cacciari, M. B., & Cintra, C. L. (2021). Inventário de autoavaliação do professor: Proposta de um instrumento baseado no Teacher Behaviors Checklist. *Psico*, *52*(2), e30095. https://doi.org/10.15448/1980-8623.2021.2.30095

Gump, S. E. (2007). Student evaluations of teaching effectiveness and the leniency hypothesis: A literature review. *Educational Research Quarterly*, *30*, 55-68.

Henklain, M. H. O, Carmo, J. S., Haydu, V. B., & Muniz, M. (2018). Brazilian faculty and student perspectives on excellent teaching. *New Directions for Teaching and Learning, 156*, 31-39. https://doi.org/10.1002/tl.20314

Henklain, M. H. O, Haydu, V. B., Carmo, J. S., Muniz, M. & Perez, W. F. (2019). Expanding the evidence of content validity for the Teacher Behavior Checklist using the IRAP. *The Psychological Record*, *69*, 205-214. https://doi.org/10.1007/s40732-019-00334-9

Henklain, M. H. O., Carmo, J. S., Haydu, V. B., Muniz, M., Buskist, W., & Keeley, J. W. (2020). Teacher Behavior Checklist: Evidências Psicométricas na Avaliação Docente por Estudantes Universitários Brasileiros [Teacher Behavior Checklist: Psychometric evidence in teacher evaluation by brazilian university students]. *Paidéia*, *30*(e3025), 1-11. https://doi.org/10.1590/1982-4327e3025

Hermosa-Bosano, C., & Keeley, J. W. (2021). Faculty and students' perceptions of excellent teaching: A study in Ecuador using the Teacher Behavior Checklist. *Trends in Psychology*, *29*(4), 752-765. https://doi.org/10.1007/s43076-021-00087-x

Ismail, E. A., & Groccia, J. E. (2017). Foreign and U.S.-educated faculty members' views on what constitutes excellent teaching: Effects of gender and discipline. *To Improve the Academy, 36*(1), 20–38. https://doi.org/10.1002/tia2.20056

Ismail, E. A., & Groccia, J. E. (2022). Foreign- and US-educated faculty members' views on what constitutes excellent teaching: Effects of prior graduate training, rank, and teaching experience. *New Directions for Teaching and Learning*, *169*(special issue), 47-64. https://doi.org/10.1002/tl.20482

Keeley, J. W., & Wilson-Doenges, G. (2024). An adaptation of the teacher behavior checklist for online and hybrid classes: Lessons about online teaching evaluations. *Scholarship of Teaching and Learning in Psychology*. Advance online publication. https://dx.doi.org/10.1037/stl0000400

Keeley, J., Christopher, A. N., & Buskist, W. (2012). Emerging evidence for excellent teaching across borders. In: J. E. Groccia, M. Al-Sudairy, & W. Buskist (Ed.), *Handbook of college and university teaching: Global perspectives* (pp. 374-390). Thousand Oaks, CA: Sage.

Keeley, J., English, T., Irons, J., & Henslee, A. M. (2013). Investigating halo and ceiling effects in student evaluations of instruction. *Educational and Psychological Measurement*, *73*, 440-457. https://doi.org/10.1177/0013164412475300

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. https://doi.org/10.1207/s15328023top3302_1

Kienen, N., Panosso, M. G., Nery, A. G. S., Waku, I., & Carmo, J. S. (2021). Contextualização sobre a Programação de Condições para Desenvolvimento de Comportamentos (PCDC): Uma experiência brasileira. *Perspectivas em Análise do Comportamento*, *12*(2), 360-390. https://doi.org/10.18761/PAC.2021.jul110

Liu, S., Keeley, J., & Buskist, W. (2015). Chinese college students' perceptions of characteristics of excellent teachers. *Teaching of Psychology, 42*(1), 83-86. https://doi.org/10.1177/0098628315620888

Liu, S., Keeley, J., & Buskist, W. (2016). Chinese college students' perceptions of excellent teachers across three disciplines: Psychology, chemical engineering and education. *Teaching of Psychology, 47*(1), p. 70-74. https://doi.org/10.1177/0098628315620888

Paulhus, D. (1984). Two-component model of socially desirable responding. *Journal of Personality and Social Psychology*, *46*, 598-609. https://doi.org/10.1037/0022-3514.46.3.598

Ribas Jr., R. C., Moura, M. L. S., & Hutz, C. S. (2004). Adaptação brasileira da Escala de Desejabilidade Social de Marlowe-Crowne [Brazilian adaptation of the Marlowe-Crowne Social Desirability Scale]. *Avaliação Psicológica*, *3*(2), 83-92. Retrieved from http://bit.ly/3bnylmx

Smith, D., Cook, P., & Buskist, W. (2011). An experimental analysis of the relation between assigned grades and instructor evaluations. *Teaching of Psychology*, *38*, 225-228. https://doi.org/10.1177/0098628311421317

Spooren, P., Mortelmans, D., & Thijssen, P. (2012). 'Content' versus 'style': Acquiescence in student evaluation of teaching? *British Educational Research Journal*, *38*, 3-21. https://doi.org/10.1080/01411926.2010.523453

Stigall, L., & Blincoi, S. (2015). Student and instructor use of the teacher behavior checklist. *Teaching of Psychology*, *42*(4), 299-306. https://doi.org/10.1177/0098628315603061

Zanini, D. S., Oliveira, K. S., Oliveira, K. L., & Henklain, M. H. O. (2022). Desafios da Avaliação Psicológica no Brasil: Nova realidade, velhas questões. *Avaliação Psicológic*a, *21*(4), 407-417. https://dx.doi.org/10.15689/ap.2022.2104.24162.04