SOCIETY FOR THE TEACHING OF PSYCHOLOGY

Teaching of Psychology
40(4) 300-303

DOI: 10.1177/0098628313501043
top.sagepub.com
SAGE

The Reliability of Student Ratings of Master Teacher Behaviors

R. Eric Landrum[1] and Jeffrey R. Stowell[2]

Abstract

Faculty members ($N = 39$) from two universities had their lectures recorded. From these videos, we selected 5-min vignettes centered on eight dimensions of the Teacher Behavior Checklist. Undergraduate students from two universities ($N = 753$) viewed three 5-min video clips of college teachers across various disciplines. Ratings on the eight dimensions (knowledgeable, respectful, approachable, realistic, enthusiastic, encouraging, flexible, and creative) were moderately to strongly intercorrelated, and students had high agreement among themselves in rating each instructor's teaching behaviors. The methodology to establish reliability of student ratings of video clips may be useful in future validity studies to assist in providing evidence-based examples of effective teaching.

Keywords

teaching, evaluative dimensions, vignettes, Teacher Behavior Checklist

Watching video recordings of one's own teaching is a potentially powerful method for gaining insights into how to improve teaching (Berg & Smith, 1996). Peer observations of teaching also provide an opportunity to learn how to improve teaching by watching others and can be valuable in providing formative feedback regarding teaching performance (Ackerman, Gross, & Vigneron, 2009; Courneya, Pratt, & Collins, 2008; Keig, 2000; Peel, 2005). However, Ackermanet et al. (2009) pointed out disadvantages with peer observation, such as including unintentional and intentional biases of the observer (e.g., manipulating the suitability of a promotion and tenure candidate) or atypical performance by the teacher being observed (e.g., the act of observation changes the natural teaching environment).

Student evaluations of teaching (SETs) can also be informative in discovering ways to improve teaching (Wilson & Ryan, 2012). Students express their opinions about the teaching behaviors that they like (Cunnane, 2010; Wilson & Taylor, 2001) and dislike (Kelsey, Kearney, Plax, Allen, & Ritter, 2004; Miley & Gonsalves, 2003; Perlman & McCann, 1998). Unfortunately, these ratings are rarely tied to specific instances of behavior that occur in the classroom; rather, students provide ratings at the end of the semester that reflect overall performance. However, in a study by Ambady and Rosenthal (1993), there was considerable agreement between judges' ratings of short video segments of teaching (three 2-s muted video clips taken from a 50-min lecture) and students' end-of-semester SETs. Based on the nonverbal behaviors in the video, judges' ratings of likable, active, supportive, optimistic, enthusiastic, honest, and confident were most strongly correlated with SETs ($rs > .60$).

The classroom behaviors rated in Ambady and Rosenthal study (1993) overlap with a list of positive teaching behaviors that occur in the classroom, identified in the Teacher Behavior Checklist (TBC; Keeley, Furr, & Buskist, 2010; Keeley, Smith, & Buskist, 2006). The TBC measures 28 behaviors of teaching that comprise two broad factors: a caring/supportive factor and a professional competency/communication factor (Keeley et al., 2006). Eight of the 28 behaviors have been agreed upon by faculty and students as being displayed by master teachers: approachable/personable, creative and interesting, encouraging and caring, enthusiastic, flexible/open-minded, knowledgeable about the subject, realistic expectations of students/fair testing and grading, and respectful (Buskist, Sikorski, Buckley, & Saville, 2002; Schaeffer, Epting, Zinn, & Buskist, 2003). The TBC has been used as part of an exercise for teachers to engage in deliberate self-reflection for teaching improvement (e.g., McGovern & Miller, 2008).

Combining the elements of videotaping, student feedback, and the TBC, our long-term goal is to create an online video library of master teacher behaviors that would be available to others for teacher improvement purposes. However, we needed to first establish that students have some level of agreement in identifying these master teacher behaviors. Thus, the purpose of our study was to determine the reliability of student ratings of teaching behaviors in videotaped segments of university instructors from various disciplines. Our data analyses focused on the following two questions: (1) Are the eight TBC behaviors

[1] Department of Psychology, Boise State University, Boise, ID, USA
[2] Eastern Illinois University, Charleston, IL, USA

Corresponding Author:
R. Eric Landrum, Department of Psychology, Boise State University, MS 1715, 1910 University Drive, Boise, ID 83725, USA.
Email: elandru@boisestate.edu

Table 1. Correlations of Teacher Behavior Checklist (TBC) Behavior Ratings by Students.

	Interesting	Encouraging	Enthusiastic	Flexible	Knowledgeable	Realistic	Respectful
Approachable	.660*	.714*	.622*	.700*	.252*	.593*	.679*
Interesting		.666*	.685*	.641*	.308*	.543*	.530*
Encouraging			.604*	.705*	.297*	.638*	.657*
Enthusiastic				.564*	.360*	.496*	.516*
Flexible					.248*	.641*	.630*
Knowledgeable						.321*	.310*
Realistic							.576*

*$p < .001$.

highly correlated as part of a unidimensional aspect of quality teaching (i.e., master teacher)? (2) What level of agreement do students have on TBC behaviors when viewing the same video segment of teaching?

Method

Participants

Faculty volunteers from various departments at Boise State University (BSU; $n = 22$) and Eastern Illinois University (EIU; $n = 14$) provided consent to have from one to three of their classroom lectures videotaped by a research assistant. Students enrolled in introductory psychology courses at BSU ($n = 659$) and EIU ($n = 94$) participated in the study for course credit to partially satisfy a course-based research exposure requirement. Participants volunteered through a web-based experiment management program.

Materials

Research assistants used high-definition cameras mounted on tripods to record faculty members' class sessions onto mini-digital video tapes, which were then digitized. We selected 19 segments of video approximately 5-min long that a research assistant identified as containing one or more examples of the eight TBC criteria selected for study: approachable/personable, creative/interesting, encourages/cares for students, enthusiastic about teaching/topic, flexible/open-minded, knowledgeable about subject matter, realistic expectations/fair grading, and respectful. After watching the video, student participants rated the instructor in the video on each of the eight TBC behaviors using a TurningPoint student response keypad (clicker) to indicate the degree to which the trait was present in the video on a scale ranging from 1 = *not at all* to 4 = *extremely*.

Procedure

After providing informed consent, the researchers gave participants (tested in groups) a clicker and a list of each of the eight TBC behaviors accompanied by a brief description of the teaching behavior that they could refer to when providing their ratings. In each session, students watched three different faculty videos (randomized across sessions). For each of the three faculty videos, students rated the instructor in the video on the eight teaching dimensions after watching 5 min of the video. Students at EIU viewed only instructors from BSU, whereas students at BSU rated faculty from both institutions. At the conclusion of the study, participants were debriefed.

For statistical analyses, we treated participants' ratings of a single video as a separate observation, resulting in up to three observations for each participant. This allowed us to include as many observations as possible in certain statistical analyses that otherwise would exclude more than a third of the participants because they were missing one or more of the eight ratings across three video segments. Thus, if they were missing one or more TBC ratings from one video, we still used their ratings on the other two video clips. The number of students who viewed each video ranged from 46 to 212 due to institutional differences in sample size and daily variation in the number of participants.

Results

We center our presentation of the results on questions of interest mentioned in the introduction.

1. Are the eight TBC behaviors highly correlated as part of a unidimensional aspect of quality teaching? Yes; Pearson correlations between the mean instructor ratings on each of the eight teaching behaviors appear in Table 1. Notably, the correlations are all strong (range .50 to .71, ns = 2,005–2,192, ps < .001), except for the correlations of "Knowledgeable" with the other behaviors (range .25–.36, ns = 2,066–2,153, ps < .001).
2. What level of agreement do students have on TBC behaviors when viewing the same video segment of teaching? For each instructor, we conducted a separate analysis using data from participants who provided all eight TBC ratings for the instructor. The resulting intraclass correlation coefficients (ICC; see McGraw & Wong, 1996; Shrout & Fleiss, 1979) reflected the absolute agreement of raters across the mean of the eight TBC behavior ratings. Across instructors, the average ICC was .83, with a range from 0.68 to 0.91, suggesting high agreement. Table 2 shows the mean ratings for each of the eight TBC behaviors.

Table 2. Mean Teacher Behavior Checklist (TBC) Behavior Ratings.

TBC Behavior	*N*	*M*	*SD*
Enthusiastic	2,173	3.36	0.76
Knowledgeable	2,212	3.26	0.88
Approachable/personable	2,225	3.12	0.99
Realistic expectations	2,072	2.97	0.84
Encouraging/caring	2,179	2.91	0.97
Respectful	2,209	2.86	0.98
Creative/interesting	2,134	2.79	0.95
Flexible/open-minded	2,194	2.51	1.01

Note. Items were rated on a scale from 1 = *not at all* to 4 = *extremely*.

Discussion

The moderately high correlations among the eight different teaching behaviors suggest that if a teacher is portraying one of the master teacher behaviors, he or she is also likely displaying others. The one exception to this conclusion is the behavior rating of "Knowledgeable." It is possible that a teacher could display a high amount of knowledge, but not be effective in conveying that knowledge to students, or conversely, display a lower amount of knowledge, but still maintain high levels of the other teaching behaviors. In our sample, the first explanation is more likely because ratings of "knowledgeable" were the second highest of all teaching behaviors, probably reflecting student's beliefs that university teachers are knowledgeable.

Students appeared cautious about making strong judgments about teachers' behaviors after only 5 min, but all behaviors were rated at or above the scale midpoint potentially due to the possibility that skilled teachers may have been more likely than unskilled teachers to volunteer to have their teaching publicly reviewed. We did not collect actual SETs from the instructors in our study to verify this, but prior researchers suggested that these TBC ratings teaching behaviors would also be correlated with other measures of teacher effectiveness (Buskist et al., 2002). Alternately, a methodology that can help identify master teachers (without relying on SETs) may be valuable to those educational researchers exploring scholarship of teaching and learning issues.

SETs are widely used to evaluate teaching, and video-recording lectures is nothing new. What is the unique contribution of our study? We were able to successfully combine these student ratings of behavior with video clips in hopes that this methodology can be used to establish reliability, setting the stage for subsequent validity studies where segments of video may be shared with faculty who desire to learn from viewing master teacher behaviors in the classroom.

Our research has limitations that can be explored in future studies. During the 5 min of video viewing, multiple TBC behaviors were often present, making it difficult to tease out the influence of specific behaviors on the overall ratings because they were not orthogonal. One possible solution would be to present clips of isolated behaviors (e.g., a few seconds of "enthusiasm") to see if specific scale ratings were affected more so than others. We hope that the methodology and the results presented here will provide the foundation for future research on the reliability and validity of student-rated videos of effective teachers.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article: This research was supported in part by a Technology Incentive Grant provided by the Office of Academic Technologies at Boise State University.

References

Ackerman, D., Gross, B. L., & Vigneron, F. (2009). Peer observation reports and student evaluations of teaching: Who are the experts? *Alberta Journal of Educational Research*, *55*, 18–39.

Ambady, N., & Rosenthal, R. (1993). Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *Journal of Personality and Social Psychology*, *64*, 431.

Berg, M. H., & Smith, J. P. (1996). Using videotapes to improve teaching. *Music Educators Journal*, *82*, 31–37. doi:10.2307/3398914

Buskist, W., Sikorski, J., Buckley, T., & Saville, B. (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. 27–39). Mahwah, NJ: Lawrence Erlbaum.

Courneya, C.-A., Pratt, D. D., & Collins, J. (2008). Through what perspective do we judge the teaching of peers? *Teaching and Teacher Education*, *24*, 69–79. doi:10.1016/j.tate.2007.01.009

Cunnane, S. (2010, July 15). Creating the perfect instructor. *Inside Higher Ed*. Retrieved from http://www.insidehighered.com/layout/set/print/news/2010/07/15/perfect

Keeley, J., Furr, R. M., & Buskist, W. (2010). Differentiating psychology students' perceptions of teachers using the teacher behavior checklist. *Teaching of Psychology*, *37*, 16–20. doi:10.1080/00986280903426282

Keeley, J., Smith, D., & Buskist, W. (2006). The teacher behaviors checklist: Factor analysis of its utility for evaluating teaching. *Teaching of Psychology*, *33*, 84–91.

Keig, L. (2000). Formative peer review of teaching: Attitudes of faculty at liberal arts colleges toward colleague assessment. *Journal of Personnel Evaluation in Education*, *14*, 67–87.

Kelsey, D. M., Kearney, P., Plax, T. G., Allen, T. H., & Ritter, K. J. (2004). College students' attributions of teacher misbehaviors. *Communication Education*, *53*, 40–55. doi:10.10/0363452032000135760

McGovern, T. V., & Miller, S. L. (2008). Integrating teacher behaviors with characters strengths and virtues for faculty development. *Teaching of Psychology*, *35*, 278–285. doi:10.1080/00986280802374609

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlations coefficients. *Psychological Methods*, *1*, 30–46.

Miley, W. M., & Gonsalves, S. (2003). What you don't know can hurt you: Students' perceptions of professors' annoying teaching habits. *College Student Journal*, *37*, 447–455.

Peel, D. (2005). Peer observation as a transformatory tool? *Teaching in Higher Education*, *10*, 489–504. doi:10.1080/13562510500239125

Perlman, B., & McCann, L. I. (1998). Students' pet peeves about teaching. *Teaching of Psychology*, *25*, 201–202.

Schaeffer, G., Epting, K., Zinn, T., & Buskist, W. (2003). Student and faculty perceptions of effective teaching: A successful replication. *Teaching of Psychology*, *30*, 133–136.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*, 420–428.

Wilson, J. H., & Ryan, R. G. (2012). Formative teaching evaluations: Is student input useful? In M. E. Kite (Ed.), *Effective evaluation of teaching: A guide for faculty and administrators: Society for the teaching of psychology*. Retrieved from http://teachpsych.org/ebooks/evals2012/index.php

Wilson, J. H., & Taylor, K. W. (2001). Professor immediacy as behaviors associated with liking students. *Teaching of Psychology*, *28*, 136–138.