

## Thick slice and thin slice teaching evaluations

Gail Tom · Stephanie Tom Tong · Charles Hesse

Received: 30 April 2009 / Accepted: 2 July 2009 / Published online: 4 August 2009  
© Springer Science+Business Media B.V. 2009

**Abstract** Student-based teaching evaluations are an integral component to institutions of higher education. Previous work on student-based teaching evaluations suggest that evaluations of instructors based upon “thin slice” 30-s video clips of them in the classroom correlate strongly with their end of the term “thick slice” student evaluations. This study’s results support previous findings, but also find that thick slice student evaluations are better than thin slice evaluations. The findings are interpreted to suggest that thick slice evaluations capture course content variables not reflected in thin slice evaluations.

**Keywords** Faculty evaluation · Teaching effectiveness · Teacher assessment · Thin slice

An annual ritual marking the end of the semester that occurs in college classes is the administration of student evaluations of teachers, an activity exacerbated in part by an era of increasing demands of accountability. Given its near universal usage and the accumulation of thousands of articles resulting from more than 40 years of research, one would expect a converging consensus rather than divisive controversy on the topic of student evaluations of teaching (SET) (Kulik 2001; Cashin 1995). However, equally strong research exists to support as to counter the validity, reliability and usefulness of SET. Typical of research that support SET is the conclusion Marsh and Roche (1997) offer following their extensive and thorough literature review: SETs are “(a)

---

G. Tom (✉) · C. Hesse  
College of Business Administration, California State University, Sacramento, 6000 J Street,  
Sacramento, CA 95819, USA  
e-mail: tomgk@csus.edu

S. T. Tong  
Department of Communication, Michigan State University, East Lansing, MI 48824, USA

multidimensional; (b) reliable and stable; (c) primarily a function of the instructor who teaches a course rather than the course that is taught; (d) relatively valid against a variety of indicators of effective teaching; (e) relatively unaffected by a variety of variables hypothesized as potential biases (e.g., grading leniency, class size, workload, prior subject interest); and (f) useful in improving teaching effectiveness when coupled with appropriate consultation.” Typical of research that questions the value of SET is the oft-cited research by Naftulin et al. (1973). They found that a very expressive, animated, and authoritative lecturer introduced as “Dr. Fox” received high student evaluations regardless of the fact that he delivered meaningless, sometimes false, content. The fact that an instructor could “seduce” a class into believing that he had expertise in a subject simply through animated nonverbal behavior seems to demonstrate the triumph of classroom stage presence over content.

Even more remarkable is Ambady and Rosenthal’s study (1993) which demonstrates that after less than a minute and after as little as 6 s of observation, not only were students’ ratings of the instructors’ nonverbal behaviors in agreement with one another, they also predicted the instructors’ long term SET. Ambady and Rosenthal (1993) coined the term “thin slices” to describe judgments about a person’s traits based on brief observations of the person. In their study, students unacquainted with the instructors were shown 13, 30 s silent videotapes. The videotapes were created by randomly selecting 10 s from the first 10 min of class, 10 s from the middle of the class and 10 s from the last 10 min of class for each of 13 instructors who taught a wide variety of undergraduate courses ranging from the natural sciences, humanities, social sciences.

Recent findings validate thin slice findings. Babad et al. (2004) report that positive judgments of brief instances of nonverbal lecturing behavior predicted positive post-course student ratings of teaching components pertaining to the instructor. Clayson and Sheffet (2006) had students rate their instructor on personality variables (e.g., agreeable—disagreeable, conscientious—not conscientious, emotionally stable—emotionally unstable, introverted—extroverted, unimaginative-uncreative—imaginative—creativity) at the beginning of the academic term, after less than 5 min of exposure. They found that the students’ ratings of their perceptions of the instructor’s personality at the beginning of the term were associated with the end of course SET scores. In addition to the initial personality assessment taken at the beginning of the term, the students’ perception of their instructor’s personality was also assessed later on in the term when students were more familiar with the instructor and classroom environment, Clayson and Sheffet (2006) found that changes in the measures of the student perception of their instructor’s personality taken later in the course were more strongly associated with long term SET than student ratings of instructor personality taken earlier in the course. Thus, it appears that nonverbal, personality traits impact on SET increases over the length of the term.

Thin slice findings have been demonstrated with visual only presentations as well as with presentations that incorporated both sound and vision. We wanted to determine whether the visual component alone contributes differently to thin slice evaluations and correlates differently with long term evaluations compared to presentations that include visual and verbal components. To address this, we created two conditions: a 30 s video of instructor presentations with the sound turned off and the same 30 s

video of instructor presentations with sound. Thus, two thin slice SET scores were obtained: thin slice no sound SET and thin slice sound SET scores. To represent the instructor's long term SET, we decided to use the average of all of the instructor's end of term SETs provided by their university department. We use the term "thick slice" to describe multiple term average SET judgments as a complement to the term thin slice (Ambady and Rosenthal 1993). The use of three different measurements of SET allows us to make comparisons not specifically addressed in previous research.

## 1 Methodology

Instructors who teach televised classes were contacted and invited to participate in the study. The university library houses a collection of video tapes of all the instructors' televised presentations during the active term. These are made available to students enrolled in the courses who may have missed a televised lecture or who want to review a presentation. In addition, the library also maintains video tapes of the instructors' courses that were broadcasted the immediate previous term. The ten instructors who agreed to participate in the study granted the researchers access to the video tapes and to their long term average student evaluation scores provided by their departments.

For each instructor, a video tape was randomly selected and a random 30 s segment from the selected video tape of the instructor addressing the class was copied. It was necessary that the student participants in this study had no previous knowledge or interaction with the instructors in the study. The selection of lower classmen enrolled in introductory Communication Studies courses met this requirement. A total of 20 classes participated in this study. Half the classes viewed 30 s clips of the ten instructors with the sound turned off. Half the classes viewed the same 30 s clips of the ten instructors with the sound turned on. The no sound condition replicates stimuli used in prior studies (e.g., Ambady and Rosenthal 1993) whereas the use of both nonverbal and verbal stimuli reflects the norm in classrooms and was a more realistic representation of the classroom environment.

### 1.1 Procedure

The researcher introduced the study on "first impressions" at the beginning of class and passed out a survey. The survey reiterated that the purpose of the research was to study first impressions and listed the ten instructors. Students who had a course from any of the instructors or who were familiar with the instructors were asked not to participate. In all the classes, three students were familiar with one or more of the listed instructors. The researcher explained that 30 s clips of ten instructors were going to be shown one at a time. Following the showing of each 30 s clip, students were asked to rate the instructor on the 15 scales provided (accepting, active, attentive, competent, confident, dominant, empathic (understanding), enthusiastic, honest, likable, anxious, optimistic, professional, supportive, warm 1 = very descriptive, 2 = descriptive, 3 = neutral, 4 = not descriptive, 5 = not at all descriptive). These were the same scales used in the previous study by Ambady and Rosenthal (1993). The researcher showed a clip of an instructor, then provided students with about 30 s

**Table 1** Mean (Scale: 1 = high to 5 = low) ratings of instructors on scale items for thin slice conditions

Scale item	Thin slice condition	
	No sound	Sound
Accepting	2.63	2.35
Anxious	2.86	2.78
Active	2.56	2.50
Attentive	2.35	2.32
Competent	2.22	2.15
Dominant	2.53	2.52
Empathic	2.83	2.47
Enthusiastic	2.82	2.68
Honest	2.53	2.27
Dominant	3.47	3.41
Optimistic	2.79	2.60
Professional	2.32	2.21
Supportive	2.46	2.71
Warmth	2.69	2.85

to answer the survey. The researcher was able to assess when students had completed their ratings by monitoring their behaviors: pens/pencils stopped moving, survey page was turned, students looked up from the survey, etc. Data collection was completed within 10–15 min. Each student and the instructor were provided with a five dollar gift card to a campus eatery as a gratuity for their participation.

## 2 Results

Table 1 shows the mean ratings of the instructors on the 15 scale items for the thin slice sound and no sound conditions. Table 2 shows the mean ratings of the instructors for the thin slice sound, thin slice no sound and thick slice SET.

A correlation analysis was conducted to determine the relationship between thick slice SET and thin slice SET. The results showed that the thick slice SET is related to both the average thin slice no sound SET ( $r = .35$   $p = .000$ ) and the average thin slice sound SET ( $r = .10$   $p = .03$ ).

To understand the specific traits influencing thick slice SET, a correlation analysis of the thick slice evaluation score to each of the traits measured in the questionnaire was conducted. The results show that in the no sound condition, 14 (all except the trait anxious) of the 15 traits are significantly correlated to the thick slice SET. In the verbal condition 5 traits (dominant, honest, likable, optimistic, warmth) are significantly correlated to the thick slice SET (See Table 3).

An ANOVA conducted to compare the average ratings of the instructors' thick slice SET (mean = 1.49) (Scale 1 = higher to 5 = lower) and thin slice SET in the sound

**Table 2** Mean (Scale: 1 = high to 5 = low) ratings of instructors for thin slice and thick slice conditions

Instructor	Condition		
	Thin slice		Thick slice
	No sound	Sound	Long term average
A	2.41	2.39	1.24
B	2.68	2.24	1.31
C	3.32	2.61	1.35
D	2.52	2.29	1.80
E	3.19	3.12	1.80
F	2.66	2.43	1.45
G	2.49	2.49	1.49
H	2.06	1.89	1.37
I	2.53	2.53	Not available
J	2.27	2.08	Not available

**Table 3** Correlation of traits with thick slice SET

Traits	Condition	
	No sound	Sound
Accepting	.242**	.038
Active	.248**	.026
Anxious	.027	.067
Attentive	.221**	.053
Competent	.238**	.007
Confident	.192**	.001
Dominant	.140**	.099*
Empathic	.172**	.043
Enthusiastic	.228**	.064
Honest	.154**	.108**
Likable	.227**	.138**
Optimistic	.196**	.078*
Professional	.091*	.026
Supportive	.194**	.018
Warmth	.246**	.120**

\*\* .01 (2 tailed)

\* .05 (2 tailed)

condition (mean = 2.51) and thin slice SET in the no sound condition (mean = 2.75) was significant ( $F = 35.266$ ,  $df = 2, 18$ ,  $p = .000$ ). A Tukey post hoc test shows that the instructors' thick slice SET is higher than the average SET in both the no sound thin slice condition ( $p = .00$ ) and sound thin slice condition ( $p = .00$ ); the average thin slice SET sound and average thin slice SET no sound were comparable ( $p = .304$ ).

### 3 Discussion

Similar to previous studies (e.g., [Ambady and Rosenthal 1993](#); [Babad et al. 2004](#); [Clayson and Sheffet 2006](#)), this study supported thin slice findings. We should not

be surprised that thin slice ratings are correlated with end of term thick slice ratings because the nonverbal behaviors/personality traits that are assessed in thin slice are relatively immutable and unchanging and are reflected throughout the course (Clayson and Haley 1990; Clayson and Sheffet 2006; Erdle et al. 1985; Marsh and Hocevar 1991; Murray et al. 1990; Sherman and Blackburn 1975).

The thin slice findings in this study were not as strong as reported in previous studies (e.g., Ambady and Rosenthal 1993). This may be due to the difference between the instructors participating in this study and those in previous studies. The instructors in this study were self selected, long-distance-learning instructors who teach televised courses. They have uniformly high thick slice SET and do not reflect the range of department scores found in other studies.

Also different from previous research is the use in this study of thick slice SET averaged over multiple terms compared to previous studies use of single term thick slice SET. Multiple term thick slice SET may be less variable than single term thick slice SET.

The finding that thick slice SET is higher than both thin slice no sound SET and thin slice sound SET suggest that thick slice SET is measuring variables not captured by the traits measured in this study such as those relating to course content. Given their lack of expertise in the course content, the students thin slice evaluations were most likely based upon judgments of delivery variables. On the other hand, students' thick slice evaluations made towards the end of the term most likely include judgments of both delivery variables and course content variables: significance, quantity, quality, relevance of their learning, the organization, format, pace of the classroom presentations, etc. Thus, the thick slice SET likely reflects both delivery variables/personality variables and course content variables whereas the thin slice SET captures only the delivery/personality variables. To investigate the effect of expertise of course content on thin slice SET, future thin slice research could select participants who are well-versed in the subject matter being presented, but who have no prior knowledge of the instructor. Would the thin slice SET be more similar to the thick slice SET for an expert population compared to a novice/student population?

In the present study, the two average thin slice SETs were similar, indicating that the addition of verbalizations/sound to behaviors did not offer sufficient information to affect evaluations. Although verbalizations/sound affected the number of traits affecting thin slice evaluations (14 of the 15 traits to arrive at their evaluations in the no sound condition, but only 5 traits in the sound condition) the average thin slice SETs were comparable. Verbalizations would not inform raters who are not knowledgeable with the content.

The instructor's nonverbal behaviors are likely the observable factors that students use to infer the latent qualities that are related to course content. In the majority of cases, thin slice reveals an authentic link. For example, an instructor who is enthusiastic about his subject matter does naturally exhibit facial expressions and hand gestures that easily communicate his enthusiasm. The question of the validity of instructor evaluations arise when this link is weak (Yunker and Yunker 2003), erroneous (Chonko 2004) or made inauthentic by manipulation. In the case of Dr. Fox (Naftulin et al. 1973) the manipulation of the nonverbal behaviors is purposely not linked to or indicative of the content. However, we would suspect that a Dr. Fox could only get away with

vacuous content for a short while. A class of students would become suspect after more than a few class sessions.

From day one, students arrive at thin slice SET based upon delivery/personality variables which then intertwine, interact, add to and affect the reception and evaluation of subsequent course content material. It has been suggested that thin slice-type evaluations create a halo effect (Clayson and Sheffet 2006; Clayson and Haley 1990). Personality/Delivery variables are likely an inherent and inextricably integrated component of course content. That personality/delivery variables are part and parcel to SET does not invalidate SET. Studies report that students select instructors with personality traits that enhance the learning environment: understanding, energetic, empathetic (e.g., Faranda and Clark 2004) and centers for teaching improvements can offer useful techniques to improve instructor delivery and resultant SET. However, given the stability of personality variables, they likely set an upper limit to SETs. This means that some instructors' SETs do not reflect their students learning. Case in point, Simpson and Siguaw (2000) report of a finance instructor who raised student scores on a national exam from the 13th to 97th percentile, whose instructor SETs consistently resided in lowest third of all instructor SETs. On the other hand there are Dr. Fox-like instructors whose SETs are more generous than their students' learning. Such cases may tend to be the extremes of a continuum. It is much more likely that most instructor SETs place between these extremes. This study suggests the recognition of the limits imposed by instructor personality variables to thick slice SET based upon student perceptions. Given the likelihood that current SET instruments are likely to remain and are not likely to be replaced any time soon by an evaluation system that does not include student perceptions, serious thought to the inclusion of additional inputs to teaching evaluations such as standardized instruments that assess student learning, consideration of the course materials: syllabus, exams, lecture materials, etc. would improve the validity of assessment of teaching effectiveness.

**Acknowledgments** This research was sponsored by an academic grant from the College of Business Administration, California State University, Sacramento.

## References

- 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*, 431–441.
- Babad, E., Avni-Babad, D., & Rosenthal, R. (2004). Prediction of students' evaluations from brief instances of professors' nonverbal behavior in defined instructional situations. *Social Psychology of Education*, *7*(1), 3–33.
- Cashin, W. E. (1995). Student ratings of teaching: The research revisited. *Instructional Development and Effectiveness Assessment Paper No. 32*. Manhattan: Center for Faculty Evaluation and Development, Kansas State University.
- Chonko, L. B. (2004). If it walks like a duck...: Concerns about quackery in marketing education. *Journal of Marketing Education*, *26*(1), 4–17.
- Clayson, D. E., & Haley, D. (1990). Student evaluations in marketing: What is actually being measured? *Journal of Marketing Education*, *12*(3), 9–17.
- Clayson, D. E., & Sheffet, M. J. (2006). Personality and the student evaluation of teaching. *Journal of Marketing Education*, *28*(2), 149–160.

- Erdle, S., Murray, H. G., & Rushton, J. P. (1985). Personality, classroom behavior and student ratings of college teaching effectiveness: A path analysis. *Journal of Educational Psychology*, 77(4), 394–407.
- Faranda, W. T., & Clark, I., III. (2004). Student observations of outstanding teaching: Implications for marketing education. *Journal of Marketing Education*, 26(3), 271–281.
- Kulik, J. A. (2001). Student ratings: Validity, utility and controversy. *New Directions for Institutional Research*, 109, 9–25.
- Lantos, G. P. (1997). Motivating students: The attitude of the professor. *Marketing Education Review* 2, 27–38.
- Marsh, H. W., & Hocevar, D. (1991). Students' evaluations of teaching effectiveness: The stability of mean rating of the same teachers over a 13-year period. *Teaching & Training Education*, 7(1), 303–314.
- Marsh, H. W., & Roche, L. A. (1997). Making students' evaluations of teaching effectiveness effective. *American Psychologist*, 52(11), 1187–1197.
- Murray, H. G., Ruston, J. P., & Paunonen, S. V. (1990). Teacher personality traits and student instructional ratings in six types of university courses. *Journal of Educational Psychology*, 82(2), 250–261.
- Naftulin, D. H., Ware, J. E., & Donnelly, F. A. (1973). The Doctor Fox lecture: A paradigm of educational seduction. *Journal of Medical Education*, 48, 630–635.
- Petty, R. E., Cacioppo, J. T., & Schumann, D. (1983). Central and peripheral routes to advertising effectiveness: The moderating role of involvement. *The Journal of Consumer Research*, 10, 135–146.
- Sherman, B. R., & Blackburn, R. T. (1975). Personal characteristics and teaching of college Faculty. *Journal of Educational Psychology*, 67(1), 124–131.
- Simpson, P. M., & Siguaw, J. A. (2000). Student evaluations of teaching: An exploratory study of the faculty response. *Journal of Marketing Education*, 22(3), 199–213.
- Slomon, S. A. (2002). Two systems of reasoning. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *The Psychology of intuitive judgment: Heuristics and biases*. Cambridge: Cambridge University Press.
- Williams, W. M., & Ceci, S. J. (1997). How'm I doing? Problems with the use of student ratings of instructors and courses. *Change*, 29(5), 12–23.
- Yunker, P. J., & Yunker, J. A. (2003). Are student evaluations of teaching valid? Evidence from an analytical business core course. *Journal of Education for Business*, 78(6), 313–317.

## Author Biographies

**Gail Tom** is a Professor of Marketing at California State University, Sacramento. Her research interests are in consumer behavior decision making.

**Stephanie Tom Tong** is a Ph.D. candidate at the Department of Communication at Michigan State University. Her research interests include how communication technology impacts interpersonal relationships and educational processes.

**Charles Hesse** is a graduate of California State University, Sacramento. He is interested in research in promotions and effectiveness.

Copyright of Social Psychology of Education is the property of Springer Science & Business Media B.V. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.