JEM JOURNAL of EDUCATIONAL MEASUREMENT
NCME national council on measurement in education
Journal of Educational Measurement
Fall 2016, Vol. 53, No. 3, pp. 313–331

No Second Chance to Make a First Impression: The "Thin-Slice" Effect on Instructor Ratings and Learning Outcomes in Higher Education

Preeti G. Samudra, Inah Min, Kai S. Cortina, and Kevin F. Miller
University of Michigan

Prior research has found strong and persistent effects of instructor first impressions on student evaluations. Because these studies look at real classroom lessons, this finding fits two different interpretations: (1) first impressions may color student experience of instruction regardless of lesson quality, or (2) first impressions may provide valid evidence for instructional quality. By using scripted lessons, we experimentally investigated how first impression and instruction quality related to learning and evaluation of instruction among college students. Results from two studies indicate that quality of instruction is the strongest determinant of student factual and conceptual learning, but that both instructional quality and first impressions affect evaluations of the instructor. First impressions matter, but our findings suggest that lesson quality matters more.

It is common practice in college courses to ask students to evaluate their instructors at the end of each course. These evaluations are often made available to other students to use in selecting courses, and for promotion committees to use in evaluating faculty. Due to their consequential nature, these ratings should ideally reflect careful analysis across an entire term and hence be a reliable and valid measure of the quality of instruction. Student ratings of instructors do correlate with student achievement (Cohen, 1987), but many other factors also affect these ratings, such as initial student interest, workload, and difficulty of the course (see Benton & Cashin, 2012, for a review).

Ambady and Rosenthal (1993) presented data that challenge the validity of course evaluations. They found that course evaluations could be very accurately predicted from personality judgments made by different and untrained students who watched a 30-second silent video clip from the first day of class. This was the first in a long series of studies showing that "thin slices" of behavior are sufficient for people to make a range of judgments, ranging from which candidate is likely to win an election (Rule et al., 2010) to whether a surgeon is likely to be sued for malpractice (Ambady et al., 2002).

Striking as these results are, however, they leave open two quite different interpretations. The first is that first impressions color our later experience such that a teacher who makes a bad impression on the first day of class has irrevocably tarnished his or her reputation. This could be an example of a confirmatory bias—the tendency for initial impressions to affect later judgments even after exposure to contradictory

---

Authors Preeti G. Samudra and Inah Min contributed equally to this publication.

evidence. The expectations we build from our initial impressions influence our interpretation of later events, leading us to favor, remember, or selectively gather information consistent with our initial beliefs (Rabin & Schrag, 1999; Rosenzweig, 2007).

But there may be another reason that first impressions predict actual end-of-term evaluations. They may, in fact, be reliable indicators of the quality of a course. The impression that a teacher makes on the first day of class may be consistent with the impression she makes throughout the term. In fact, prior research has shown that nonverbal, relatively automatic behaviors that are linked to first impression formation are quite stable across different contexts (Weisbuch, Slepian, Clarke, Ambady, & Veenstra-VanderWeele, 2010). Thus, the impressions students have of instructors may be relatively consistent between the first minute of a lesson and the remainder of the class or term. First impressions may be predictive because they tend to be valid indicators of the quality of instruction students will receive.

These two explanations of the thin slice effect on student evaluations of instruction have different real-world implications. If first impressions color how students experience the instruction they later receive, then instructors should put effort into shaping those first impressions. Additionally, this would cast doubt on the validity of teacher evaluations as indicators of quality throughout the course rather than simply during the first few minutes.

The alternative view has very different implications. If first impressions have their effect because they tend to be consistent with the course as a whole, then there is no shortcut to being perceived as an effective instructor. Instructors should focus on ensuring that they are competent and passionate about what they teach. This would in turn support the validity of teacher evaluations as being reflective of genuine quality of instruction.

Existing studies of the thin slice effect show how strong and pervasive this phenomenon is, but due to their correlational nature, they cannot determine whether first impressions shape student evaluations directly, or whether they are valid predictors of the quality of the course as a whole. In the current studies, we extend the research on the thin-slice effect by using an experimental paradigm to systematically vary the quality of first impressions and instruction. We additionally investigate the role of first impressions and quality of instruction on student learning. Existing research finds a positive link between perceived instructional quality and student learning (e.g., Helmke, Schneider, & Weinert, 1986; Keith & Cool, 1992). The paradigm used here allows us to look at how first impressions and overall instructional quality relate to three kinds of outcomes in a lecturing context: (1) student learning, (2) student evaluation of instructor personality dimensions, and (3) student evaluations of instructor effectiveness. Study 1 provides an initial experimental investigation into whether first impressions and instructional quality impact learning and teacher evaluations. Study 2 replicates and extends Study 1 by changing the instructor, topic, and the nature of the questions used to assess learning.

Study 1

The basic paradigm used in both studies involved random assignment of subjects to four conditions resulting from crossing (1) videotaped first impressions designed

to be positive or negative with (2) lectures designed to be more or less effective. The demands of a scripted, videotaped lesson limited us to examining a single lecture without interaction between the instructor and students. At the university level, learning occurs through a variety of means, and is often less unidirectional than lectures. However, lectures remain a very common form of instruction in higher education (McKeachie & Svinicki, 2006).

To vary the quality of first impression, we manipulated variables found to be important in the Ambady and Rosenthal (1993) study, such as confidence and enthusiasm. To vary the quality of instruction, we manipulated factors that relate to effective lecturing, such as quality of explanations, elaborations, and organization/connections between ideas, using examples, and including recaps within the lecture (Atkins & Brown, 1988; McKeachie & Svinicki, 2006).

If first impressions have an effect on evaluation of instruction, we would expect a strong effect of the introduction independent of the quality of later instruction. Alternatively, if the power of first impressions lies in their generally accurate predictions of what follows, then the quality of instruction should be the main predictor of both learning and evaluations, independent of the quality of the first impression.

Methods

Participants. Participants were 192 undergraduate students (87 males, 105 females) enrolled in an introductory psychology course at a large Midwestern university in the United States. The participants were typically first or second year students ($M_{\text{age}} = 18.77$ years, $SD_{\text{age}} = .90$). They received course credit for taking part in the study. The sample was predominantly Caucasian.

Materials.

First impression videos. A Caucasian, middle-aged male actor portrayed the instructor for all videos used in Study 1. In the first impression video, the actor introduced himself and described his interest in the subject matter. A similar verbal script was used for the good and bad first impression videos. For the good first impression, the actor projected confidence, enthusiasm, and an interest in teaching the subject matter. This was accomplished by using a strong and positive tone of voice as well as enthusiastic and relevant gestures and facial expressions. To make a bad first impression, the actor displayed lack of interest in the subject matter and in teaching. This was demonstrated by a relatively monotonous and negative tone of voice, a disinterested facial expression, and frequent fidgeting. The good first impression lasted 43 seconds, while the bad first impression comprised the first 49 seconds of the video.

Instructional videos. The topic of instruction was topography and reading topographic maps. The final portion of the instructional video was a practice quiz where the majority of concepts were reviewed. In the instruction videos, the actor stood behind a podium in a large lecture hall while delivering a scripted PowerPoint lecture. Both good and bad instruction videos used the same slides and covered the same material, but the good instruction video was well organized and included complete explanations and elaborations. Additionally, it included three mid-lecture recaps to

break up the lecture and provide review. The good lesson was 19 minutes, 15 seconds long. In the bad instruction video, the instructor appeared less organized by needing time to remember what to say for some slides, and using scripted filler words such as "um." Additionally, within each slide, information was covered in less detail and was sometimes presented in a less coherent order. The bad instruction video was 15 minutes, 30 seconds long. Example scripts are included in Appendix 1.

Design and Procedure. The study employed a 2 (first impression: good/bad) × 2 (instruction: good/bad) experimental design (First impression × Instruction). Participants were randomly assigned to one of the four conditions. The experiment took approximately 45 minutes, and participants were tested independently, viewing the lecture on their own computer. Participants did not interact during the experiment and could not see each other's computer screens.

Participants watched the assigned video and then completed an online quiz that included a measure of student learning and a teacher evaluation questionnaire.

Measures.

Demographic variables. To assess comparability across conditions, participants answered questions about their age, gender, race/ethnicity, year in school, and English as a first language status.

Student learning. Learning was assessed through a 15-question, multiple-choice quiz that focused on factual recall. Questions assessed important concepts and definitions covered in the instructional video, and additionally required participants to read new topographic maps. The sum of correct responses to quiz questions (0–15) was used as an indicator of student learning.

Teacher evaluation. The final portion of the online questionnaire asked participants to rate the instructor on a scale of 1–10 on the following 14 dimensions (following Ambady & Rosenthal, 1993): accepting, active, anxious, attentive, competent, confident, dominant, empathetic, honest, likeable, optimistic, professional, supportive, and warm. In order to obtain ratings of student perceptions of instructional quality, participants rated instructional quality on a scale of 1–5 for the extent to which the teacher was: (1) excellent, (2) clear and understandable, and (3) well prepared, as well as how interesting the lesson was.

Results

Analysis. Descriptive statistics for both student learning outcomes and teacher evaluations can be found in Table 1. No significant differences were found between the four experimental groups regarding age, year in school, gender, race, and English as a first language status; therefore these variables were not included in further analyses. Dependent measures were further analyzed using a 2 (first impression: good or bad) × 2 (instructional quality: good or bad) multivariate analysis of variance (MANOVA), with both factors between subjects.

Student Learning. As shown in Table 2, there was a significant main effect of instructional quality on quiz scores, $F(1, 188) = 4.47$, $p = .036$, $\eta_p^2 = .02$, with

Table 1
Study 1 Descriptive Statistics: Mean Scores for Each Condition

		Condition							
		GG		BG		GB		BB	
Dependent Variable		Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Student Learning	Total Quiz Score	11.00	2.56	11.11	2.53	10.20	2.85	10.28	2.67
Perceived Teacher	Accepting	7.54	2.35	7.15	2.13	6.57	1.74	5.86	2.46
Personality	Active	5.52	2.58	5.30	2.31	4.71	1.97	3.74	2.16
Traits	Anxious	2.74	1.88	3.81	2.59	6.04	2.59	4.64	2.66
	Attentive	6.59	2.45	6.00	2.55	5.49	2.19	4.62	2.44
	Competent	7.26	2.08	7.62	2.08	5.43	2.52	4.94	2.71
	Confident	7.17	2.53	6.91	2.47	4.63	2.51	4.10	2.38
	Dominant	4.70	2.48	4.70	2.36	4.16	2.13	2.82	1.66
	Empathetic	5.54	2.71	5.45	2.52	5.02	2.05	4.86	2.31
	Honest	7.80	2.38	7.23	2.10	7.22	1.91	6.12	2.72
	Likeable	7.02	2.34	6.45	2.79	5.80	2.44	5.38	2.83
	Optimistic	7.09	2.32	6.34	2.76	5.78	2.48	5.18	2.38
	Professional	7.65	2.13	7.38	2.45	5.39	2.35	4.88	2.80
	Supportive	7.35	2.67	6.77	2.62	5.73	2.35	4.86	2.64
	Warm	6.74	2.82	6.26	3.12	6.06	2.44	4.82	2.66
Overall Teacher	Excellent Teacher	3.57	1.00	3.47	.97	2.57	.96	2.52	1.07
Ratings	Clear and Understandable	4.37	.71	4.26	.82	3.04	1.14	3.06	1.22
	Well Prepared	4.26	.88	4.30	.83	2.53	1.21	2.76	1.24
	Interesting	2.63	1.22	2.62	1.09	2.43	1.10	2.28	1.09

Notes. GG = good first impression/good instruction; BG = bad first impression/good instruction; GB = good first impression/bad instruction; BB = bad first impression/bad instruction.

good instruction ($M = 11.05$, $SD = 2.53$) producing higher quiz scores than bad instruction ($M = 10.24$, $SD = 2.75$). There was no significant main effect of first impression on student learning, and no interaction between instructional quality and first impression (Fs < 1), suggesting that the instructor first impression did not affect student learning.

Teacher Evaluations.

Instructional quality influences. There was a significant main effect of instructional quality on participant ratings for 13 of the 14 specific instructor trait dimensions (see Table 2); participants who received good instruction rated the instructor more favorably across multiple personality traits than did participants who received poor instruction. Specifically, compared to the bad instruction condition, participants rated the instructor in the good instruction condition as significantly more *accepting*, $F(1,188) = 12.81$, $p < .001$, $\eta_p^2 = .06$, *active*, $F(1, 188) = 13.15$, $p < .001$, $\eta_p^2 = .07$, *attentive*, $F(1,188) = 12.67$, $p < .001$, $\eta_p^2 = .06$, *competent*, $F(1,188) = 38.94$, $p < .001$, $\eta_p^2 = .17$, *confident*, $F(1,188) = 56.11$, $p < .001$, $\eta_p^2 = .23$, *dominant*,

Table 2
Study 1 Main Effects of First Impression and Instructional Quality

		Independent Variables							
		First Impression				Instructional Quality			
Dependent Variable		*F*	*df*	Partial η^2	Significant	*F*	*df*	Partial η^2	Significant
Student Learning	Total Quiz Score	.06	1/188	.00	.81	4.47	1/188	.02	.04
Perceived Teacher Personality Traits	Accepting	3.07	1/188	.02	.18	12.81	1/188	.06	<.01
	Active	3.37	1/188	.02	.07	13.15	1/188	.07	<.01
	Anxious	.22	1/188	.00	.64	33.96	1/188	.15	<.01
	Attentive	4.38	1/188	.02	.04	12.67	1/188	.06	<.01
	Competent	.03	1/188	.00	.86	38.94	1/188	.17	<.01
	Confident	1.23	1/188	.01	.27	56.11	1/188	.23	<.01
	Dominant	4.56	1/188	.02	.03	14.87	1/188	.07	<.01
	Empathetic	.14	1/188	.00	.71	2.56	1/188	.01	.11
	Honest	6.35	1/188	.03	.01	6.50	1/188	.03	.01
	Likeable	1.73	1/188	.01	.19	9.24	1/188	.05	<.01
	Optimistic	3.49	1/188	.02	.06	11.83	1/188	.06	<.01
	Professional	1.20	1/188	.01	.27	45.23	1/188	.19	<.01
	Supportive	3.85	1/188	.02	.05	22.46	1/188	.11	<.01
	Warm	4.66	1/188	.02	.03	6.99	1/188	.04	.01
Overall Teacher Ratings	Excellent Teacher	.26	1/188	.00	.61	44.85	1/188	.19	<.01
	Clear and Understandable	.11	1/188	.00	.74	76.21	1/188	.29	<.01
	Well Prepared	.75	1/188	.00	.39	113.52	1/188	.38	<.01
	Interesting	.25	1/188	.00	.62	2.77	1/188	.02	.10

Table 3
Study 1 Significant Interaction Effects: First Impression × Instruction Interaction

	First Impression × Instructional Quality			
Dependent Variables	*F*	*df*	Partial η^2	Significant
Anxious	12.13	1/188	.06	<.01
Dominant	4.65	1/188	.02	.03

$F(1,188) = 14.87$, $p < .001$, $\eta_p^2 = .07$, *honest*, $F(1,188) = 6.50$, $p = .012$, $\eta_p^2 = .03$, *likeable*, $F(1,188) = 9.24$, $p = .003$, $\eta_p^2 = .05$, *optimistic*, $F(1,188) = 11.83$, $p = .001$, $\eta_p^2 = .06$, *professional*, $F(1,188) = 45.23$, $p < .001$, $\eta_p^2 = .19$, *supportive*, $F(1, 188) = 22.46$, $p < .001$, $\eta_p^2 = .11$, and *warm*, $F(1, 188) = 6.99$, $p = .009$, $\eta_p^2 = .04$, and significantly less *anxious*, $F(1,188) = 33.96$, $p < .001$, $\eta_p^2 = .15$. There was no main effect of instructional quality on ratings of *empathetic* ($F < 3$).

Instructional quality also resulted in significant differences on the three overall quality of instructor ratings, with higher ratings in good than bad instruction conditions for being an *excellent teacher*, $F(1,188) = 44.85$, $p < .001$, $\eta_p^2 = .19$, *clear and understandable*, $F(1,188) = 76.21$, $p < .001$, $\eta_p^2 = .29$, and *well prepared*, $F(1, 188) = 113.52$, $p < .001$, $\eta_p^2 = .38$; for ratings of how *interesting* the material was, this difference approached significance ($p = .10$).

First impression influences. The good first impression condition produced significantly higher ratings than the bad first impression condition on 4 of the 14 specific instructor traits: *attentive*, $F(1,188) = 4.38$, $p = .038$, $\eta_p^2 = .02$, *dominant*, $F(1, 188) = 4.56$, $p = .034$, $\eta_p^2 = .02$, *honest*, $F(1,188) = 6.35$, $p = .013$, $\eta_p^2 = .03$, and *warm*, $F(1,188) = 4.66$, $p = .032$, $\eta_p^2 = .02$. The difference between instructor ratings in the good and bad first impression conditions approached significance for three additional traits: *active, optimistic*, and *supportive* ($.05 < p < .09$). First impression condition did not produce any differences in the three overall quality of instructor ratings or the rating of how *interesting* the lecture was.

First impression by instructional quality interactions and principal component analysis. As shown in Table 3, there were two significant interactions between instructional quality and first impression condition on the instructor ratings as *anxious*, $F(1,188) = 12.13$, $p < .001$, $\eta_p^2 = .06$ and *dominant*, $F(1,188) = 4.65$, $p = .032$, $\eta_p^2 = .02$.

A principal component analysis on all 14 instructor traits revealed an optimal two-factor solution where all traits excluding *anxious* loaded on a single factor that explained 57.19% of the raw variance (rescaled factor loadings ranged from .729 to .869). The reverse-scored trait of anxious loaded on a second factor that explained an additional 10.72% of the raw variance (rescaled factor loading of .804). This suggested that participant ratings on traits of a positive valence (e.g., *confident*, *supportive*) were quite similar to one another, whereas the one negatively worded trait

(*anxious*) was rated differently. Since the positive valence traits all loaded on the same factor, we considered any first impression by instructional quality interactions for those traits to be trivial due to their small effects sizes and inconsistency between traits. The interaction for the second factor, *anxious*, revealed that the rating differed more between the good and bad instruction conditions following a good first impression than a bad first impression.

Discussion

The purpose of Study 1 was to compare two potential hypotheses about what accounts for the relationship between first impressions and instruction, and to see whether these effects influence learning as well as student evaluations. Two explanations consistent with existing research were compared. The first is that first impressions have a persistent effect that determines how students experience later instruction. The second is that the validity of first impressions stems from their consistency with the actual quality of the lesson that follows.

Study 1 looked at these factors in the context of a videotaped lesson that enabled us to vary the relation between first impression and lesson quality. We did find some effects of first impression on course evaluations that are independent of the quality of instruction. These effects last at least across a single lesson.

However, the effects of first impressions on teacher evaluations were much smaller than the effects of the actual instruction received. First impressions affected ratings of only 4 of 14 instructor traits and did not influence the three overall instructor effectiveness ratings. Instructional quality, on the other hand, strongly influenced 13 of the 14 instructor traits, as well as all three overall instructor effectiveness ratings.

Our results are more consistent with the view that first impressions predict course evaluations because they can be valid predictors of later instruction, although there were some independent effects of first impression. Overall, our results suggest that teacher evaluations are more affected by instructional quality than by first impressions.

Even though there were some effects of first impressions on teacher evaluations, they did not impact learning in any way. Instructional quality, however, affected how much students remembered from the lesson. Good instruction, characterized by good organization, complete explanations, elaborations of difficult concepts, and mid-lecture recaps resulted in the highest level of student learning, as seen by student quiz scores. This may be because a well-organized, fully explained lesson helps students sustain their attention, leading to better understanding, and subsequently, more learning. When students learn more, they might feel positively about not only the lesson itself, but also about the person providing the lesson.

While Study 1 indicated that instructional quality is the strongest determinant of both learning and student evaluations of teachers, the generalizability of our findings was limited in numerous ways. Our lesson was portrayed by a single actor, and we used only one topic of instruction (topography). The subject matter used was factually oriented, and it may be that different factors would affect student engagement with more conceptual subject matter.

Study 2 extended this paradigm to look at a very different subject matter that enabled us to look at both factual recall and higher level conceptual learning. The topic was relevant to psychology, making it more applicable to their learning context. We also varied the instructor and adopted a pretest–posttest design to account for prior knowledge.

Study 2

The lecture used in Study 2 focused on international comparisons in education, a topic quite different from that used in Study 1. We used a young, female instructor for Study 2 in order to substantially vary instructor characteristics from Study 1.

The change in topic permitted us to include quiz questions that assessed different levels of learning. The learning assessment in Study 1 largely incorporated lower level, factual knowledge–based questions, according to Bloom's Taxonomy (Anderson et al., 2001; Bloom, 1956), while the quiz in Study 2 included conceptual questions involving application and analysis that require higher order thinking skills. Due to the possibility that instructional quality may differentially affect learning at different levels of critical thinking, the posttest quiz in Study 2 incorporated both factual and conceptual questions.

Overall, we sought to test the robustness of the findings of Study 1 when the instructor and topic of instruction were different, and to see whether first impression and instructional quality would benefit learning at both the factual and conceptual levels.

Methods

Participants. Participants were 238 undergraduate students (102 males, 136 females) from the same subject pool used for Study 1. The participants were typically first or second year students ($M_{age} = 18.74$ years, $SD_{age} = .99$). Participants received course credit for taking part in the study. The sample was predominantly Caucasian.

Materials.

First impression videos. For study 2, a younger, East Asian female actor portrayed the instructor for the videos. As in Study 1, the first impression video consisted of the actor introducing herself and her interest in the subject matter. A similar script was used for the good and bad first impression videos, with quality of first impression manipulated through facial expression and tone of voice.

For the good first impression, the actor projected confidence, enthusiasm for the subject matter, and an interest in teaching the subject matter. This was accomplished by using a strong and positive tone of voice as well as enthusiastic and relevant gestures and facial expressions. To make a bad first impression, the actor displayed lack of interest in both the subject matter and in teaching. This was demonstrated by a relatively monotonous and negative tone of voice, a disinterested facial expression, frequent fidgeting, and looking at a cellphone. The good first impression video was 1 minute, 21 seconds long, and the bad first impression video was 1 minute, 50 seconds long.

Instructional videos. The topic of instruction for Study 2 was international comparisons in education. Unlike Study 1, no practice quiz was given at the end of the lecture, and there were no mid-lecture recaps in the good instruction condition. Identical slides were used in the good and bad instruction PowerPoint presentations.

The good instruction video was well organized and included complete explanations and elaborations. The good instruction video was 18 minutes, 58 seconds long. The bad instruction video maintained the same order of slides, but the instructor appeared less organized by needing time to remember what she needed to say for some slides, appearing unaware of when topic transitions were occurring, and using scripted pauses and filler words such as "um." Additionally, within each slide, information was covered in less detail and was sometimes presented in a less coherent order than in the good-instruction version. The bad instruction video was 22 minutes, 31 seconds long. All information later tested on the quiz was fully covered in the good and the bad instruction video.

Design and Procedure. Study 2 employed a 2 (first impression: good/bad) $\times$ 2 (instruction: good/bad) experimental design (First impression $\times$ Instruction). Participants were randomly assigned to one of the four conditions.

The procedure was identical to Study 1 with one main difference. In Study 2, before watching the video, participants completed a short pretest questionnaire that included demographic variables and pretest questions about the material covered in the video.

Measures.

Pretest. To assess comparability of conditions, participants answered questions about their age, gender, race/ethnicity, year in school, and English as a first language status. Participants also answered five multiple-choice questions and one open-ended question about the topic of the video to ensure group comparability in prior knowledge.

Student learning. Learning was assessed through a quiz with 18 multiple-choice questions and 1 open-ended question. Questions assessed important concepts and definitions covered in the instructional video (see Appendix 2 for examples). Twelve of the questions were designed to be lower level definitional questions that were exclusively based on recalling the information in the video. Six additional questions were more conceptual in nature, and required applying the information in the video or extending it to a new context. The sum of correct responses to quiz questions in total (0–18) as well as for the basic (0–12) and conceptual (0–6) questions separately was used as an indicator of student learning.

Teacher evaluation. The teacher evaluation questionnaire used in Study 2 was identical to that in Study 1.

Results

Analysis. Descriptive statistics for both student learning outcomes and teacher evaluations can be found in Table 4. No significant differences were found between the four experimental groups regarding age, year in school, gender, race, and English

Table 4
Study 2 Descriptive Statistics: Mean Scores for Each Condition

		Condition							
		GG		BG		GB		BB	
Dependent Variable		Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Student Learning	Total Quiz Score	11.47	2.70	11.45	2.40	10.17	2.21	9.73	2.63
	Conceptual Portion	3.67	1.26	3.53	1.26	3.15	1.11	2.98	1.28
	Basic Factual Portion	7.80	1.95	7.92	1.63	7.02	1.72	6.75	2.08
Perceived Teacher Personality Traits	Accepting	7.22	1.81	6.87	2.14	5.76	2.42	5.31	2.60
	Active	5.22	2.55	4.75	2.40	3.63	2.21	3.68	2.52
	Anxious	3.88	2.42	5.88	2.72	5.36	2.95	5.83	3.12
	Attentive	5.85	2.16	6.02	2.48	3.69	2.06	3.39	2.27
	Competent	6.77	2.17	6.50	2.36	3.92	2.63	3.20	2.25
	Confident	5.80	2.07	4.67	2.42	3.10	1.99	2.97	2.18
	Dominant	3.97	2.05	3.15	1.95	2.47	1.81	2.14	1.67
	Empathetic	5.38	2.34	5.33	1.95	3.51	2.04	3.29	2.34
	Honest	7.35	2.15	7.70	1.78	6.22	2.36	5.20	2.72
	Likeable	6.28	2.26	5.32	2.52	3.97	2.48	3.29	2.49
	Optimistic	6.73	1.93	5.85	2.39	4.08	2.49	3.69	2.58
	Professional	7.75	2.03	5.72	2.48	3.36	2.43	2.80	2.25
	Supportive	6.42	2.08	5.60	2.48	3.80	2.39	3.41	2.44
	Warm	6.30	2.30	5.30	2.61	3.78	2.49	3.53	2.67
Overall Teacher Ratings	Excellent Teacher	3.27	.92	2.60	.96	1.67	.92	1.66	.98
	Clear and Understandable	3.70	1.09	3.50	1.02	1.93	1.03	1.90	1.03
	Well Prepared	3.97	.88	3.57	.91	1.54	.80	1.61	.98
	Interesting	3.08	1.11	2.88	1.15	2.42	1.22	2.51	1.32

Notes. GG = good first impression/good instruction; BG = bad first impression/good instruction; GB = good first impression/bad instruction; BB = bad first impression/bad instruction.

as a first language status; therefore these variables were not included in further analyses. Dependent measures were further analyzed using a 2 (first impression: good or bad) × 2 (instructional quality: good or bad) MANOVA, with both factors between subjects and pretest scores as a covariate.

Student Learning. Instructional quality condition had a significant main effect on overall posttest quiz scores, $F(1, 233) = 22.17$, $p < .001$, $\eta_p^2 = .09$, where quiz scores were higher in the good instruction condition ($M = 11.46$, $SD = 2.55$) than the bad instruction condition ($M = 9.95$, $SD = 2.43$). This held true for both the subset of conceptual questions, $F(1, 233) = 11.16$, $p = .001$, $\eta_p^2 = .05$ and basic factual questions, $F(1, 233) = 16.92$, $p < .001$, $\eta_p^2 = .07$ (see Table 5). Students who received a well-organized, fully explained lecture were more likely to score higher

Table 5
Study 2 Main Effects of First Impression and Instructional Quality

		Independent Variables							
		First Impression				Instructional Quality			
Dependent Variable		*F*	*df*	Partial η^2	Significant	*F*	*df*	Partial η^2	Significant
Student Learning	Total Quiz Score	.22	1/233	.00	.64	22.17	1/233	.09	<.01
	Conceptual Portion	.72	1/233	.00	.39	11.16	1/233	.05	<.01
	Basic Factual Portion	.00	1/233	.00	.95	16.92	1/233	.07	<.01
Perceived Teacher Personality Traits	Accepting	1.74	1/233	.01	.19	26.36	1/233	.10	<.01
	Active	.48	1/233	.00	.49	17.88	1/233	.07	<.01
	Anxious	13.04	1/233	.05	<.01	3.85	1/233	.02	.05
	Attentive	.02	1/233	.00	.90	67.16	1/233	.22	<.01
	Competent	2.41	1/233	.01	.12	100.82	1/233	.30	<.01
	Confident	5.00	1/233	.02	.03	60.73	1/233	.21	<.01
	Dominant	5.95	1/233	.03	.02	26.51	1/233	.10	<.01
	Empathetic	.18	1/233	.00	.67	48.26	1/233	.17	<.01
	Honest	1.30	1/233	.01	.26	37.55	1/233	.14	<.01
	Likeable	6.60	1/233	.03	.01	46.98	1/233	.17	<.01
	Optimistic	3.80	1/233	.02	.05	61.75	1/233	.21	<.01
	Professional	18.15	1/233	.07	<.01	149.33	1/233	.39	<.01
	Supportive	3.42	1/233	.01	.07	62.33	1/233	.21	<.01
	Warm	3.22	1/233	.01	.07	43.10	1/233	.15	<.01
Overall Teacher Ratings	Excellent Teacher	7.04	1/233	.03	.01	107.69	1/233	.32	<.01
	Clear and Understandable	.68	1/233	.00	.41	154.55	1/233	.40	<.01
	Well Prepared	1.87	1/233	.01	.17	355.40	1/233	.60	<.01
	Interesting	.13	1/233	.00	.72	10.98	1/233	.05	<.01

on the posttest quiz compared to students who received a poorly organized and less-detailed lecture.

There was no main effect of first impression on student learning, suggesting that the first impression did not have an effect on how much students ultimately learned from the lecture. We also found no interaction between instructional quality condition and first impression. The covariate of pretest score was significantly related to overall posttest quiz score, $F(1, 233) = 5.31$, $p = .022$ as well as the subscore for basic factual questions, $F(1, 233) = 6.24$, $p = .013$. Pretest scores were not significantly related to the subset of conceptual quiz questions.

Teacher Evaluations.

Instructional quality influences. As shown in Table 5, instructional quality produced a significant main effect on participant teacher evaluation ratings for 13 of the 14 specific instructor trait dimensions, suggesting that participants who received good instruction rated the instructor more favorably across multiple personality traits compared to participants who received poor instruction. Participants rated the instructor in the good instruction condition as significantly more *accepting*, $F(1, 233) = 26.36$, $p < .001$, $\eta_p^2 = .10$, *active*, $F(1, 233) = 17.88$, $p < .001$, $\eta_p^2 = .07$, *attentive*, $F(1, 233) = 67.16$, $p < .001$, $\eta_p^2 = .22$, *competent*, $F(1, 233) = 100.82$, $p < .001$, $\eta_p^2 = .30$, *confident*, $F(1, 233) = 60.73$, $p < .001$, $\eta_p^2 = .21$, *dominant*, $F(1, 233) = 26.51$, $p < .001$, $\eta_p^2 = .10$, *empathetic*, $F(1, 233) = 48.26$, $p < .001$, $\eta_p^2 = .17$, *honest*, $F(1, 233) = 37.55$, $p < .001$, $\eta_p^2 = .14$, *likable*, $F(1, 233) = 46.98$, $p < .001$, $\eta_p^2 = .17$, *optimistic*, $F(1, 233) = 61.75$, $p < .001$, $\eta_p^2 = .21$, *professional*, $F(1, 233) = 149.33$, $p < .001$, $\eta_p^2 = .39$, *supportive*, $F(1, 233) = 62.33$, $p < .001$, $\eta_p^2 = .21$, and *warm*, $F(1, 233) = 43.10$, $p < .001$, $\eta_p^2 = .15$ than the instructor in the bad instruction condition. The higher ratings of *anxious* in the bad instruction condition compared to the good instruction condition approached significance, $F(1, 233) = 3.85$, $p = .05$.

The instructor in the good instruction condition received significantly higher ratings for being an *excellent teacher*, $F(1, 233) = 107.69$, $p < .001$, $\eta_p^2 = .32$, being *clear and understandable*, $F(1, 233) = 154.55$, $p < .001$, $\eta_p^2 = .40$, and being *well prepared*, $F(1, 233) = 355.40$, $p < .001$, $\eta_p^2 = .60$. There was also a significant difference in ratings of how *interesting* the material in the lecture was, $F(1, 233) = 10.98$, $p = .001$, $\eta_p^2 = .05$, with the good instruction condition rated as more interesting than the bad instruction condition.

First impression influences. Our analyses showed that first impressions did not have a significant effect on student learning, as measured by the posttest quiz score controlling for prior knowledge. Nonetheless, the good first impression condition produced significantly higher ratings than the bad first impression condition on 4 of the 14 specific instructor traits: *confident*, $F(1, 233) = 5.00$, $p = .026$, $\eta_p^2 = .02$, *dominant*, $F(1, 233) = 5.95$, $p = .015$, $\eta_p^2 = .03$, *likable*, $F(1, 233) = 6.60$, $p = .011$, $\eta_p^2 = .03$, and *professional*, $F(1, 233) = 18.15$, $p < .001$, $\eta_p^2 = .07$. The good first impression condition produced significantly lower ratings than the bad first impression condition for the trait of *anxious*, $F(1, 233) = 13.04$, $p < .001$, $\eta_p^2 = .05$. The

difference between first impression conditions on the traits of *warm, optimistic*, and *supportive* approached significance ($.05 < p < .08$). This shows that, even though not all instructional traits were affected, the instructor was rated more favorably when she gave a confident, enthusiastic introduction than when she gave a monotonous, disinterested introduction.

The good first impression condition produced higher scores than the bad first impression condition for the dimension of being an excellent teacher, $F(1, 233) = 7.04$, $p = .009$, $\eta_p^2 = .03$. First impression condition did not produce any differences in the rating of how interesting or clear and understandable the lecture was, or how well prepared the instructor was.

First impression by instructional quality interactions and principal component analysis. There were five significant interactions between instructional quality and first impression condition. Three interactions involved the specific instructor traits of *anxious*, $F(1, 233) = 4.88$, $p = .028$, $\eta_p^2 = .02$, *honest*, $F(1, 233) = 5.27$, $p = .023$, $\eta_p^2 = .02$, and *professional*, $F(1, 233) = 5.94$, $p = .016$, $\eta_p^2 = .03$. A significant interaction was also found for two of the three overall instructor effectiveness ratings: *excellent teacher*, $F(1, 233) = 7.24$, $p = .008$, $\eta_p^2 = .03$, and well *prepared*, $F(1, 233) = 3.94$, $p = .048$, $\eta_p^2 = .02$ (see Table 6).

We conducted a principal component analysis on all 14 instructor traits and the three overall instructor effectiveness ratings. This analysis revealed an optimal two-factor solution where all traits and overall ratings excluding anxious loaded on a single factor that explained 61.55% of the variance (factor loadings ranged from .705 to .900). The reverse-scored trait of anxious loaded on a second factor that explained an additional 9.22% of the variance (factor loading of .944). As in Study 1, we considered any first impression by instructional quality interactions for the traits and overall ratings that loaded on the first factor (where the traits had a positive valence) to be trivial due to their small effects sizes and inconsistency between traits.

The interaction for the final trait that loaded on the second factor, *anxious*, revealed similar results to Study 1. Differences in anxious ratings between good and bad instruction conditions were greater after the good first impression than after the bad first impression.

Table 6
Study 2 Significant Interaction Effects: First Impression × Instruction Interaction

First Impression × Instructional Quality				
Dependent Variables	*F*	*df*	Partial η^2	Significant
Anxious	4.88	1/233	.02	.03
Honest	5.27	1/233	.02	.02
Professional	5.94	1/233	.03	.02
Excellent Teacher	7.24	1/233	.03	.01
Well Prepared	3.94	1/233	.02	.05

Discussion

Study 2 tested whether the findings of Study 1 would replicate when we changed the instructor used and the topic, and looked at conceptual as well as factual learning. In general, our findings from Study 2 did replicate what we found in Study 1. Even when many factors related to the instructor and lesson were changed, instructional quality had a strong effect on both learning and teacher evaluations, whereas first impressions did not affect learning and had smaller effects on the evaluations students gave the instructor.

This suggests that students are able to focus on the quality of instruction rather than just the initial impression an instructor makes. This finding supports the overall validity of teacher evaluations, although some first impression effects do persist across at least one class session.

It is also important to note that in both Study 1 and Study 2, for the trait of *anxious*, the difference in ratings between the good and bad instruction conditions were greater when the teacher had made a good first impression. While this effect was not large, it was found with different instructors and topics, and suggests that students may infer that an instructor who makes a bad first impression followed by a disorganized lecture does so out of anxiety.

Finally, we found that instructional quality influenced both factual and conceptual learning whereas first impressions influenced neither. This suggests that good instruction facilitates both lower level factual processes related to remembering and understanding, and higher level conceptual thinking in the form of applying or extending information to new contexts.

General Discussion

By using an experimental paradigm, we were able to distinguish between two possible interpretations of why evaluations of a short, silent clip of an instructor's first class are quite similar to the end-of-term course evaluations of that instructor. We found that good first impressions do, in fact, increase teacher evaluation ratings for different instructors, in line with findings by Ambady and Rosenthal (1993). However, these effects are small when compared to the impact of instructional quality on teacher evaluations and factual and conceptual learning.

This suggests that making a strong first impression is no shortcut to obtaining a positive evaluation of instruction by students. At least among college students, in settings where we could experimentally control the relation between first impression and later instruction, what students learned and their evaluation of the quality of instruction were predominantly determined by the quality of instruction and not the qualities their instructor showed during the first minute of class. That introduction did influence their evaluation of the instructor, but since we presented a single short lesson, it may be that these effects would diminish or disappear across the course of the semester if not reinforced by similar experiences.

These results are encouraging from a pedagogical point of view, because they suggest that a teacher may overcome a bad first impression by providing good instruction. Conversely, it suggests that making a good first impression on students is only the beginning of the work of being an effective instructor. The results also support

the validity of students' end-of-term evaluations as measures of instruction quality and not mere reflections of first impressions.

Our study was limited to a single lecture-based session that students participated in as an experiment. This allowed us to experimentally vary relations between first impressions and the instruction that followed, but also limits its generalizability to real instructional settings. First impressions may have stronger effects in discussion settings, where a poor initial impression may cause students to opt out of engagement and participation. Situations where students receive meaningful grades may lead to a different dynamic between first impressions and instructional quality. It may also be that other factors come into play across longer time intervals.

We lack a good method of quantifying differences between good and bad first impressions or good and bad instructional quality, so we cannot connect the differences we observed with the range of variation in these factors in real-world settings. Finally, although we looked at two very different topics, students generally found both to be rather uninteresting. It may be that topics perceived as more interesting would show a different pattern of results.

Despite these limitations, an experimental approach to looking at relations between first impression and instructional quality provides the only method for disentangling factors that are inextricable in real classrooms.

Conclusions

Is there a second chance to make a first impression? Our results suggest that, in fact, there is. Consistent good instruction throughout the term should be sufficient to overcome any negative impressions formed by a poor first class.

By experimentally manipulating initial impression and instructional quality, we were able to demonstrate that instructional quality has by far the bigger impact on student learning and evaluation of instruction. This supports the validity of student evaluations of instruction and suggests that students are able to look beyond the first impression an instructor makes and evaluate the instruction that follows. In natural settings, however, the same factors that lead to a poor first impression may persist throughout a class, reinforcing the conclusions drawn in an initial class. We still have much to learn about the processes that instructors can use to enlist student engagement and interest, but we hope that these results will be a source of encouragement to every instructor who has taught a bad first class, as well as to every student who has endured one.

Acknowledgments

This study was supported by a grant from the National Center for Educational Research (NCER) at the Institute of Educational Science (R305A100178, PI: Kevin Miller; Making Room for Student Thinking: Using Automated Feedback, Video-Based Professional Development, and Evidence-Based Practice Recommendations to Improve Mathematical Discussion). The authors would like to thank Fiona Nowlin for her contributions to the conception of the project and the development of lesson materials in Study 1. They would also like to acknowledge the efforts of the research assistants who assisted in collecting the data for these studies. They also greatly

appreciate the efforts of Ron Miller and Shirley Wang, who portrayed the instructors in our study materials.

APPENDIX 1

Good and Bad Instruction Topography Script for Three Lecture Slides

Lecture Slide Number	Good Instruction Script	Bad Instruction Script
3	Elevation is the height of a topographic feature or landform relative to sea level. Many mountains start from the ocean bed, so it is important to remember that elevation represents the height above sea level only, not the height from the ocean floor all the way to the mountain peak.	So ....elevation.....this is important. It is the height relative to sea level. You can see it on this figure, and it is important to remember that it is sea level and not the ocean bed.
11	If you look closely at the map here, you will notice that these contours come in two forms. Some are bold, thicker contours, like the two indicated by the arrows. These are called *index elevation contours.* They act as markers for elevation changes, and usually have a label, or an index, for the elevation of that contour. In this case, the circled index elevation contour has an elevation of 7,012 feet above sea level.	If you look closely, you will notice that these contours come in a couple of different forms. Some are bold and thicker contours, and are elevation contours because they have elevation markers. See here, this contour has an elevation of 7,012 feet above sea level. So, the units of elevation can vary, they can be in feet or meters, or other units depending on where the map is made.
12	Now that we know the bold, darker lines represent index elevation contours, let us address the other fainter lines. All the other contours that are not bolded are simply called *elevation contours*.	All the other contours that are not bolded are called elevation contours. Wait . . . what did I call the other contours? Um . . . I think I called them elevation contours as well . . . .let us see . . . oh yes . . . like it said in the picture on the last slide, those ones that were thicker are actually index elevation contours, and these are elevation contours.

APPENDIX 2

Sample Factual and Conceptual Question From Study 2

Sample Factual Question:

Which factors are associated with higher academic achievement within a single country?

A. Enjoyment of subject
B. Greater classroom socioeconomic diversity
C. Higher academic self-concept
D. A and C only
E. All of the above

Reason this is factual: This information was specifically mentioned in the lecture and needed to be recalled to answer this question correctly.

Sample Conceptual Question:

Researchers from Qatar argue that they perform poorly on the Trends in International Mathematics and Science Study (TIMSS) tests *only* because their mathematics curriculum is so different from that of Western countries. If the Qatar researchers are correct, which of the following should be True?

A. Students from Qatar should improve their TIMSS performance on the next wave of data collection
B. Students from Qatar should decline in TIMSS performance on the next wave of data collection
C. Students from Qatar should perform similarly on TIMSS and the Programme for International Student Assessment (PISA)
D. Students from Qatar should perform better on PISA than TIMSS
E. None of the above

Reason this is conceptual: The lecture discusses how TIMSS tends to be more closely tied to the curriculum than PISA. Students need to identify the required knowledge and assess what is likely to occur in this hypothetical scenario based on that knowledge.

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.

Ambady, N., LaPlante, D., Nguyen, T., Rosenthal, R., Chaumeton, N., & Levinson, W. (2002). Surgeons' tone of voice: A clue to malpractice history. *Surgery*, *132*(1), 5–9.

Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., Raths, J., & Wittrock, M. C. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York, NY: Longman.

Atkins, M., & Brown, G. (1988). *Effective teaching in higher education*. London, UK: Routledge. Retrieved April 10, 2016, from http://www.ebrary.com.

Benton, S., & Cashin, W. E. (2012). *Student ratings of teaching: A summary of research and literature* (IDEA Paper no. 50). Manhattan, KS: The IDEA Center. (http://www.theideacenter.org/research-and-papers/idea-papers). Accessed April 17, 2016.

Bloom, B. S. (1956). *Taxonomy of educational objectives, Handbook I: Cognitive domain*. New York, NY: David McKay.

Cohen, P. A. (1987, April). *A critical analysis and reanalysis of the multisection validity meta-analysis*. Paper presented at the annual meeting of the American Educational Research Association, Washington, DC.

Helmke, A., Schneider, W., & Weinert, F. E. (1986). Quality of instruction and classroom learning outcomes: The German contribution to the IEA classroom environment study. *Teaching and Teacher Education*, *2*(1), 1–18.

Keith, T. Z., & Cool, V. A. (1992). Testing models of school learning: Effects of quality of instruction, motivation, academic coursework, and homework on academic achievement. *School Psychology Quarterly*, *7*(3), 207–226.

McKeachie, W., & Svinicki, M. D. (2006). *McKeachie's teaching tips: Strategies, research, and theory for college and university teachers* (12th ed.). Boston, MA: Houghton Mifflin.

Rabin, M., & Schrag, J. L. (1999). First impressions matter: A model of confirmatory bias. *The Quarterly Journal of Economics*, *114*(1), 37–82.

Rosenzweig, P. (2007). *The halo effect... and the eight other business delusions that deceive managers*. New York, NY: Simon and Schuster.

Rule, N. O., Ambady, N., Adams, Jr., R. B., Ozono, H., Nakashima, S., Yoshikawa, S., & Watabe, M. (2010). Polling the face: Prediction and consensus across cultures. *Journal of Personality and Social Psychology*, *98*, 1–15.

Weisbuch, M., Slepian, M. L., Clarke, A., Ambady, N., & Veenstra-VanderWeele, J. (2010). Behavioral stability across time and situations: Nonverbal versus verbal consistency. *Journal of Nonverbal Behavior*, *34*, 43–56.

Authors

PREETI G. SAMUDRA is a Graduate Student in Developmental Psychology, University of Michigan, 530 Church Street, Ann Arbor, MI 48109; preetigs@umich.edu. Her primary research interests include literacy learning and instruction.

INAH MIN is a Graduate Student in Combined Program in Educational Psychology, University of Michigan, 530 Church Street, Ann Arbor, MI 48109; inahpark@umich.edu. Her primary research interests include cross-cultural comparisons in education, learning, and teaching.

KAI S. CORTINA is a Professor of Psychology at the University of Michigan, 530 Church Street, Ann Arbor, MI 48109; schnabel@umich.edu. His primary research interest include learning motivation, cultural differences in educational processes, and quantitative research methods

KEVIN F. MILLER is a Professor of Education and Psychology at the University of Michigan, 610 E University Avenue, Ann Arbor, MI 48109. His primary research interests include classroom processes and cultural and developmental effects on children's learning.