

Presenting a Validated Mid-Semester Evaluation of College Teaching to Improve Online Teaching

Virginia L. Byrne

*Department of Advanced Studies, Leadership, and Policy
School of Education and Urban Studies
Morgan State University*

Alice E. Donlan

*Teaching and Learning Transformation Center
University of Maryland, College Park*

Abstract

Mid-semester formative evaluations of college teaching are a promising, low-cost solution to providing online instructors with in-the-moment feedback to improve their online teaching practices. However, existing instruments suffer from issues of validity and bias, and fail to align with evidence-based strategies. This paper presented and psychometrically validated a research-based Mid-Semester Evaluation of College Teaching for Online Instructors (MSECT-O) among 170 undergraduate students in seven online courses. Pilot study results demonstrated that the MSECT-O is a valid and reliable tool for online educators seeking to improve their online teaching and classroom climate.

Keywords: online teaching, student evaluation of teaching, formative feedback, formative evaluation of teaching, classroom climate

Byrne, V.L., & Donlan, A.E. (2020). Presenting a validated mid-semester evaluation of college teaching to improve online teaching. *Online Learning*, 24(2), 94-110.
<https://doi.org/10.24059/olj.v24i2.2126>

Correspondence concerning this article should be addressed to Virginia Byrne,
virginia.l.byrne@gmail.com.

Presenting a Validated Mid-Semester Evaluation of College Teaching to Improve Online Teaching

Each semester, more faculty and graduate instructors are asked to teach an online course without formal preparation in effective online teaching (e.g., Borup & Evmenova, 2019). To support these instructors, universities often provide course templates and support (Hill, 2012) that tend to focus on course design instead of implementation (Thomas & Graham, 2019) and fail to ground effective online teaching in broader teaching theory and research. Often, faculty development programs (e.g., coaching, workshops) are adopted to address this disconnect and train new online instructors (Meyer & Murrell, 2014). These programs, however, require time and resources (Meyer, 2014) that many universities cannot offer at the scale of online learning growth. Further, instructors may not have time to learn about online teaching before the semester begins. Instead, they often need tools to help them learn how to effectively teach online *while* they are teaching online (e.g., Darby, 2019).

A low-cost approach for supporting online instructors is to gather mid-semester formative feedback from students, tailored to evidence-based teaching practices. Formative feedback from students can help instructors refine their practices mid-semester and improve students' learning experiences (Hampton & Reiser, 2004; Overall & Marsh, 1979). Formative feedback—which takes place during the process and provides guidance for improvement—is an evidence-based strategy for improving performance (Hattie & Timperley, 2007). Unlike traditional end-of-semester evaluations, mid-semester evaluations (MSEs) provide instructors with feedback while they still have time to refine their teaching, which can be beneficial for addressing issues before the high-stakes end-of-semester evaluations (Cohen, 1980; Costello, Weldon, & Brunner, 2002). MSE feedback has the potential to provide just-in-time feedback to new online instructors; however, the field lacks an evidence-based and valid MSE instrument for online students to reflect on the online instruction. The purpose of this paper is to present a brief, literature-based MSE for online instructors and demonstrate its validity as a measure of college-level online teaching.

Review of Relevant Literature

This study is framed by Hattie and Timperley's (2007) work on the importance of timely feedback for improving teaching and learning. Whereas Hattie and Timperley focus on the importance of feedback from the instructor to improve students' performance, this study focuses on the importance of students' feedback to improve teachers' performance. As instructors learn how to teach online, in-the-moment feedback from students can help them refine their practice during the semester.

Student Evaluations of Teaching

American public institutions evaluate the quality of teaching with multiple methods but most commonly by an end-of-semester student evaluation of teaching questionnaire (SET; McKeachie, 1997; Spooren, Brockx, & Mortelmans, 2013). SETs collect systematic, cross-campus, longitudinal data of student perspectives on teaching and course quality that can be compared across semesters, courses, instructors, and mediums (e.g., face-to-face vs. online teaching; Carle, 2009). SETs afford the university with aggregate statistics of student satisfaction, an attractive metric both for financial and student recruitment. The validity of SETs, however, is critiqued in the literature (Spooren et al., 2013). Most notably, critics say SET responses are biased

because of instructor identity, the data are too general to inform faculty, and items are not developed using evidence-based psychometrics (e.g., Hammonds, Mariano, Ammons, & Chambers, 2017).

Most institutions develop their own SET instrument, but these measures are limited because of their lack of validation or alignment with education literature, and, as with all surveys, are only as strong as the literature used to create them (Hammonds et al., 2017). Most traditional student evaluations fail to reliably measure the multidimensionality of teaching (Knol, Dolan, Mellenbergh, & van der Maas, 2016), and instead load onto only one dimension of teaching (e.g., student satisfaction; Marsh, 1987). Teaching effectiveness, however, is more validly and accurately measured as a multi-dimensional factor, which is not simply satisfaction (Donlan & Byrne, 2020; Ambrose, Bridges, DiPietro, Movett, & Normal, 2010). For example, researchers such as Bangert (2008) pose that online teaching effectiveness consists of multiple factors including active learning.

End-of-semester SETs are also the primary method used for gathering institution-wide student evaluations of online teaching (Byrne, 2018). SETs for online courses are often developed by the institution and include a mix of Likert scale and open-ended core questions used for all in-person and online classes (Hammonds et al., 2017). However, just like evaluations of face-to-face courses, most evaluations for online courses were not developed using evidence-based psychometrics (Hammonds et al., 2017). Exceptions include the costly eSIR, a questionnaire developed by Educational Testing Services (Liu, 2012), which was recently discontinued by the company. Additionally, Gómez-Rey, Barbera, and Fernández-Navarro (2016) adapted the Online Learning Consortium scorecard into a student-facing evaluation. However, at 36 items, its length may hinder students' completion of the survey. Validated, evidence-based surveys enable institutions to ask students meaningful questions about the multidimensional factors of effective online teaching and focus on constructs meaningful to online teachers. Unfortunately, summative end-of-semester feedback cannot be implemented until the following semester, which fails to benefit the students completing the survey.

Mid-Semester Student Evaluations of Teaching

Formative feedback is often used to improve teaching (Overall & Marsh, 1979) and identify issues mid-semester before the end-of-semester evaluations are collected. Mid-semester evaluations (MSE) allow instructors to collect actionable and timely feedback from students in time to implement changes in the current semester (Berridge, Penney, & Wells, 2012; Costello et al., 2002). By collecting feedback mid-semester, instructors have time to react to feedback and make changes that could improve their end-of-semester evaluations, a critical measure for promotion and tenure. Thus, instructors have both intrinsic and extrinsic motivations to improve their teaching and the student experience. The use of a formative feedback process has been found to inform the use of better teaching practices (e.g., Hampton & Reiser, 2004) and can contribute to improved student satisfaction with the course (Costello et al., 2002; Overall & Marsh, 1979). MSEs have been found to be particularly beneficial for new instructors (Hampton & Reiser, 2004).

Bias in Student Evaluations

Existing student evaluations of teaching have been found to be biased measures of instruction dependent on student-level variables (Spooren, Vandermoere, Vanderstraeten, & Pepermans, 2017). For example, students tend to provide more positive evaluations when they expect to earn a higher grade or if the course is an elective, rather than a required course (Ting,

2000). Additionally, older students (e.g., juniors and seniors) are more likely to provide higher evaluations than younger students (Spooren, 2010). Despite the bias in student evaluations of teaching, Thomas, Graham, and Piña (2018) note that they “capture a critical perspective on online instructor behaviors that may be missed otherwise” (p. 6). Thus, an evidence-based formative feedback tool is needed that provides new online instructors with less biased student feedback on their online teaching.

Mid-Semester Evaluations in Online Teaching

Online instructors gather MSE data in different ways. The most discussed in the literature are instructor-made online surveys that collect anonymous open-ended feedback from students (Peterson, 2016). While these opportunities to provide feedback are appreciated by students, repeatedly generating and writing out ideas for course improvements is perceived to be tedious by students (Winchester & Winchester, 2012). Additionally, instructors often adapt in-person formative feedback processes to the online environment such as Berridge et al. (2012) and O’Neil-Hixon, Long, and Block (2017) who both adapted the Small Groups Instructional Diagnosis (SGID; Coffman, 1998) process for online courses. During an online SGID, the instructor identifies a colleague or faculty developer to repeatedly communicate with students via email to collect their thoughts on the teacher’s effectiveness using open-ended prompts such as “What suggestions do you have for this class?” The responses are then synthesized into a report and recommendations for the instructor. This process is both labor-intensive and unfocused, with students providing feedback on topics unrelated to online teaching effectiveness such as technical issues or their desire for face-to-face time with the instructor (Berridge et al., 2012). Considering these limitations, researchers like Thomas et al. (2018) and Walker (2005) suggest the use of formative quantitative evaluations with constructs specific to aspects of the online course which the teacher can control. The field, however, lacks a literature-based instrument that gathers valid and reliable mid-semester feedback from online students with questions that they can answer and provides useful feedback to the instructor on the multiple dimensions of online teaching.

Fearless Teaching Framework

This study is part of a larger research project on mid-semester evaluations based on the University of Maryland’s Fearless Teaching Framework (Donlan, Loughlin, & Byrne, 2019; Donlan & Byrne, 2020), a literature-based model for effective college teaching organized into four dimensions: classroom *climate*, course *content*, teaching *practices*, and learning *assessments*. These four dimensions emerged from education theory and empirical literature as being strong predictors of student engagement, motivation, and success (e.g., Lave & Wenger, 1991; Wigfield & Eccles, 2000) and are classroom aspects that the instructor can control (as opposed to the classroom design, student demographics, etc.).

The Fearless Teaching Framework poses that in courses with a positive course *climate*, the instructor supports students’ learning and designs the curriculum to be challenging, yet accessible, to all students (Morin, Marsh, Nagengast, & Scalas, 2014). Effective course *content* is meaningfully relevant to students’ lives and interests, both personal and professional, developmentally appropriate for their existing knowledge level, and aligned with learning objectives (Howard, 2001; Lave & Wenger, 1991). Positive teaching *practices* are intentionally organized, aligned with active learning research, and connected with prior knowledge (Alexander & Winne, 2006; Wentzel & Brophy, 2014). Effective instructors communicate high and clear expectations of student work and engagement (Online Learning Consortium, 2019). *Assessments*

are most effective when they are aligned with learning objectives, are communicated clearly (e.g., via a rubric), and do not contribute to unnecessary student stress (Wass, Timmermans, Harland, & McLean, 2018).

Building on the extensive literature review and expert validation that informed the Framework, the SET and MSE literatures were reviewed to develop the Mid-Semester Evaluation for College Teaching (MSECT; Donlan & Byrne, 2020). The original Mid-Semester Evaluation for College Teaching (MSECT) was designed to capture students' evaluations of effective face-to-face teaching and provide instructors with reliable and actionable feedback to improve their teaching (Donlan & Byrne, 2020). In partnership with education and faculty development experts, the original 13-item MSECT was designed to be applicable to all instructors from all departments who teach all types of courses. When piloted among 29 instructors and 1,350 undergraduate students, the MSECT items convincingly loaded onto the four latent factors of the Fearless Teaching Framework (all items loaded onto one of the four factors at or above the .40 threshold). That study provided sufficient evidence that the MSECT is a valid formative evaluation instrument of effective in-person teaching (Donlan & Byrne, 2020).

After reviewing the literature on student evaluations of online teaching, however, the MSECT items developed and validated for a face-to-face course may not provide sufficient feedback for an online instructor and thus warrant an online-specific MSECT instrument (Byrne, 2018). The purpose of this paper is to present the design and validation of a pilot study of the MSECT for Online Instructors (i.e., the MSECT-O). Additionally, this paper serves to explore the extent to which the MSECT-O feedback is susceptible to documented trends in student bias of teaching evaluations.

Research Questions

1. In alignment with the Fearless Teaching Framework, to what extent do latent climate, content, practice, and assessment factors fit the underlying structure of the data?
2. To what extent does the MSECT-O data differ by student-reported variables?
3. To what extent does the MSECT-O data differ by course-level variables?

Methods

Instrument Development

Through a three-step process, the original MSECT instrument was amended to gather feedback about online teaching. First, the authors reviewed the original items and determined two that were too specific for a face-to-face classroom and thus were rewritten to be applicable for an online course. Second, experts in college teaching and online teaching reviewed the new MSECT-O instrument and provided feedback on the phrasing and relevance. After incorporating this feedback, a pilot 13-item MSECT-O instrument was developed. Third, the MSECT-O was administered to the students of seven online instructors to collect data to confirm the factor structure and validity, and potentially reduce the number items (Table 3 presents the items in the reduced 12-item MSECT-O).

The MSECT-O items used a 6-point Likert scale (1 = "Strongly Disagree" to 6 = "Strongly Agree") and were based on the four dimensions of the Fearless Teaching Framework (Donlan, Loughlin, & Byrne, 2019): climate, content, practices, and assessments. The initial MSECT-O

consistent of 13 items (reduced to 12 items in the analyses presented below) that load onto four factors:

1. Climate is a three-item measure of the extent to which the instructor fosters a classroom environment is inclusive and positive ($\alpha = .85$). Items include, “My instructor creates an inclusive learning environment where everyone is welcome.”
2. Content is a three-item measure of the extent to which the instructor conveys the relevance and connection of the course content (three-item scale $\alpha = .74$). Items include “I have the prior knowledge necessary to be successful in this course.”
3. Practice is a three-item measure of the extent to which the instructor enacts active and engaging activities to teach the content ($\alpha = .86$). Items include “During online classes, this course includes activities other than watching recorded lectures.”
4. Assessment is a three-item measure of the extent to which the instructor designs graded assignments that are fair and aligned with the learning objectives ($\alpha = .84$). Items include “My instructor provides me with timely feedback on my work.” The full instrument can be found in Table 3.

Sample

In fall 2018, seven online instructors at a large research university in the mid-Atlantic agreed to participate in the study as unpaid volunteers. Instructors were interested in getting student feedback for their professional development. The participating students were enrolled in one of the seven online courses taught by a participating instructor including an introductory level journalism course, a lower level agriculture course, two upper level humanities and social sciences courses, and three upper level STEM courses. 170 undergraduate students completed a mid-semester evaluation of teaching about their online instructor. Of the students, 13 (7.65%) identified as first year students, 22 (12.94%) were sophomores, 43 (25.29%) were juniors, 91 (53.53%) were seniors (see demographic and enrollment information in Table 1; 41.18% men, 55.88% women). This sample is demographically representative of the University of Maryland’s student body. The University’s Institutional Review Board approved this study.

Table 1

Student Demographic Information

Variable	Group	N	Sample (%)
Student Gender	Men	70	41.18%
	Other	2	1.18%
	Women	95	55.88%
Student Year	First Year	13	7.65%
	Sophomore	22	12.94%
	Junior	43	25.29%
	Senior	91	53.53%

Variable	Group	N	Sample (%)
Student Race/ Ethnicity	American Indian and Indigenous	1	0.59%
	Asian, Asian-American, or Pacific Islander	27	15.88%
	Black or African American	34	20.00%
	Latinx	15	8.82%
	Multi	14	8.24%
	Other	1	0.59%
	Prefer Not to Respond or Missing	16	10.59%
Course Discipline	White	60	35.29%
	Agriculture (1 course)	18	10.59%
	Humanities & Social Sciences (2 courses)	45	26.47%
	Journalism (1 course)	57	33.53%
	STEM (3 courses)	50	29.41%

Data Collection

In fall 2018, online instructors from all departments, ranks, and course types (including tenured faculty, tenure-track faculty, professional-track faculty, and graduate instructors) were recruited to participate in this study via email. Seven online instructors agreed to participate. Prior to agreeing to participate, instructors were informed of the Fearless Teaching Framework and the MSECT-O instrument, and that their participation would not impact their employment. Midway through the semester, participating instructors were provided with a link to the MSECT-O instrument to distribute to students. Instructors were asked to collect data after students had completed and received feedback on one major assignment (e.g., a midterm exam or project).

Students of the seven participating online instructors completed an online survey which included the 13 original MSECT-O items and questions about their course and demographic information (e.g., year in school, expected course grade, the degree to which the course is a degree program requirement). Instructors distributed the online survey via email or through the learning management system messaging platform on or around the eighth week of the sixteenth week semester and provided students with three business days to complete the instrument. The survey was designed to be mobile- and tablet-compatible so that students could complete the evaluation even when they were away from their laptop or desktop computer.

Before completing the survey, students were informed that their participation would not impact their employment with the institution, nor would it impact their grade. The Informed Consent form stated that the evaluation results would be shared only with the research team, who would aggregate and anonymize the results before sharing them with the instructors. The results would be used to study the instrument itself and provide instructors with formative feedback.

Data Analysis

Because preliminary normality testing (Shapiro & Wilk, 1965) provided evidence that the sampling distribution for the MSECT-O items was statistically different from a normal distribution ($p < .00$), robust nonparametric versions of traditional statistical methods in SPSS 24 and a robust maximum likelihood estimator (MLR) in MPlus 8.0 were adopted. To address the first research question, Confirmatory Factor Analyses (CFA) in MPlus was conducted to determine the fit of our four-factor solution. CFA assessed the extent to which the data aligned with the theoretically-

expected model based on the Fearless Teaching Framework. Because the MSECT-O items were developed to reflect the four parts of the Fearless Teaching Framework (Donlan, Loughlin, & Byrne, 2019; Kline, 2011), a confirmatory rather than exploratory factor analysis was appropriate. That is, whether the factor structure was aligned with the data was being confirmed, not rather than exploring different potential factor structures. A CFA was used to compare a four-factor model to a one-factor model to assess the extent to which this short scale could measure four subscales, as opposed to one omnibus latent construct. Output included loadings, factor covariance statistics, and multiple fit indices that aided in determining good model fit: a non-statistically significant chi-square test statistic, a comparative fit index (CFI) and Tucker Lewis index (TLI) greater than .95, a standardized root mean square residual (SRMR) value lower than .08, and a root mean square error of approximation (RMSEA) score lower than .06 (Hu & Bentler, 1999; Thompson, 2004). The model with the lower Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) values was seen to have a better fit. Factor replicability was determined by an H-index value below .80 (Hammer, 2016; Hancock & Mueller, 2001) and reliability was assessed using Cronbach's alpha (Raykov & Marcoulides, 2011).

To answer the remaining questions, robust nonparametric methods were adopted, such as: Spearman rank-order correlations (the nonparametric version of the Pearson correlation test), the Kruskal-Wallis test (the nonparametric version of the one-way ANOVA), and the Mann-Whitney U test (the nonparametric version of the unpaired Students' sample *t*-test; Byrne, 2017). The effect sizes for the significant Mann-Whitney U tests were calculated by hand to interpret the percent of the variance of the dependent variable is explained by the independent variable. These analyses assessed the extent to which students' responses differed by students' demographic variables, which may indicate some bias in the MSECT scales.

Results

Research Question 1. In alignment with the theoretical frame, to what extent do latent climate, content, practice, and assessment factors fit the underlying structure of the data? Two CFAs were conducted: one with one factor and another with four factors based on the Fearless Teaching Framework (see Table 2). During analyses, the modification indices indicated that allowing one item (Content3: "The instructor communicated clear learning outcomes for this course") to crossload onto the climate and assessment factors as well as the content factor would improve model fit. To address this crossloading issue, the item Content3 was removed in the final analysis. We suggest that Content3 is not used in future assessments of online instruction because it does not provide unique information to the instructor within the Fearless Teaching Framework. The one-factor model had poor model fit while the reduced four-factor model had good model fit (see Tables 2 and 3).

Table 2

Comparison of Confirmatory Factor Analysis Models

Models	χ^2	df	p	AIC	BIC	RMSEA	RMSEA CI		CFI	TLI	SRMR
							Lower	Upper			
Original Model, One Factor	176.01	65	<.000	5670.15	5792.90	0.10	0.08	0.12	0.88	0.85	0.06
Original Model, Four Factor	115.47	59	<.000	5573.38	5715.01	0.08	0.05	0.10	0.94	0.92	0.05
Reduced Model, One Factor	150.30	54	<.000	5271.89	5385.20	0.10	0.08	0.12	0.88	0.85	0.06
Reduced Model, Four Factor	74.95	48	0.008	5158.28	5290.47	0.06	0.03	0.08	0.97	0.95	0.04

To statistically compare the reduced one- and four-factor models and determine the model with best fit, a χ^2 difference test was conducted to compares the fit of nested latent models (Kline, 2011).

$$\chi_D^2 = \chi_{One-Factor}^2 - \chi_{Four-Factor}^2 = 150.30 - 74.95 = 75.35$$

$$df_D = df_{One-Factor} - df_{Four-Factor} = 54 - 48 = 6$$

The solution was retained where $\chi_D^2(6) = 75.35$, $p < .01$, indicating that the reduced four-factor model fits the data better than a one-factor model (reduced one-factor model: $\chi^2(54) = 150.30$, $p < .00$; AIC = 5271.89; BIC = 5385.20; RMSEA = 0.10; CFI = 0.88; TLI = 0.85, SRMR = 0.06; reduced four-factor model: $\chi^2(48) = 74.95$, $p = .008$; AIC = 5158.28; BIC = 5290.47; RMSEA = 0.06; CFI = 0.97; TLI = 0.95, SRMR = 0.04). Table 4 presents the correlations and descriptive statistics for the individual items.

Table 3

CFA Reduced Model Standardized Factor Loadings and Replicability Scores (H)

Item	CFA factor loadings (SE)				
	Model 1	Model 2			
	Single Factor H = .94	Climate Factor H = .89	Content Factor H = .85	Practice Factor H = .88	Assess. Factor H = .88
Climate1. My instructor creates an online classroom that is supportive for learning.	.86 (.03)	.92 (.02)			
Climate2. My instructor makes the class accessible to students with many different needs.	.71 (.06)	.74 (.06)			
Climate3. My instructor creates an inclusive learning environment where everyone is welcome.	.72 (.06)	.80 (.06)			
Content1. The content in this course is relevant to me academically, personally, and/or professionally.	.52 (.08)		.62 (.07)		
Content2. The instructor helps make the content of this course interesting.	.81 (.03)		.91 (.04)		
Content3. I have the prior knowledge necessary to be successful in this course.	.45 (.09)		.51 (.09)		
Practices1. During online classes, this course includes in-class activities other than lecture.	.69 (.06)			.70 (.06)	
Practices2. My instructor helps me understand new content by connecting it to things I already understand.	.84 (.03)			.89 (.02)	
Practices3. My instructor motivates me to put effort into the course.	.82 (.03)			.84 (.04)	
Assessment1. The assessments (e.g., quizzes, exams, papers) in this course are graded fairly.	.76 (.05)				.90 (.03)
Assessment2. My instructor provides me with timely feedback on my work.	.65 (.07)				.69 (.07)
Assessment3. The expectations for the assignments are clear.	.71 (.06)				.84 (.05)

H = Factor replicability scores

Note. The item Content3: “The instructor communicated clear learning outcomes for this course” was removed from the final instrument. It should not be used to assess online instruction.

Table 4

Spearman Correlations, Means, and Standard Deviations for Survey Items

Items	1	2	3	4	5	6	7	8	9	10	11	12	Mean	SD	N
1. Climate1	1												4.92	1.13	168
2. Climate2	.62**	1											4.91	1.04	151
3. Climate3	.71**	.62**	1										5.16	0.93	160
4. Content1	.50**	.43**	.55**	1									4.69	1.27	166
5. Content2	.71**	.60**	.56**	.57**	1								4.41	1.39	167
6. Content3	.38**	.39**	.32**	.41**	.40**	1							4.52	1.25	164
7. Practices1	.52**	.50**	.41**	.40**	.53**	.29**	1						5.02	1.14	166
8. Practices2	.66**	.55**	.48**	.44**	.69**	.42**	.59**	1					4.47	1.31	161
9. Practices3	.64**	.53**	.56**	.44**	.66**	.30**	.50**	.75**	1				4.52	1.32	159
10. Assessment1	.58**	.45**	.43**	.30**	.51**	.30**	.55**	.63**	.56**	1			4.71	1.30	167
11. Assessment2	.53**	.44**	.48**	.32**	.40**	.22**	.46**	.47**	.52**	.62**	1		5.23	0.95	166
12. Assessment3	.58**	.43**	.47**	.36**	.49**	.32**	.49**	.56**	.54**	.74**	.64**	1	4.95	1.17	167

** Correlation is significant at the 0.01 level (2-tailed).

Table 5

Spearman Correlations and Descriptive Statistics of Aggregate Factors and Student-reported Variables

Variables	1	2	3	4	5	6	7	Mean	SD	N
1. CLIMATE	1							5.04	0.87	149
2. CONTENT	.69**	1						4.54	1.06	162
3. PRACTICE	.74**	.66**	1					4.68	1.13	154
4. ASSESSMENT	.68**	.47**	.71**	1				4.97	1.00	165
5. Year	-.06	-.04	-.11	-.10	1			3.25	0.96	169
6. Expected Grade	.36**	.27**	.30**	.35**	-0.01	1		3.62	0.53	166
7. Required	.17*	.17*	.07	.05	-.02	-.05	1	4.09	1.79	158

** Correlation is significant at the 0.01 level (two-tailed).

* Correlation is significant at the 0.05 level (two-tailed).

Research Question 2. To what extent does the MSECT-O data differ by student reported variables? Students' year in school was not significantly correlated with any of the MSECT-O constructs (see Table 5). This means students provided comparable feedback regardless of their year in college. Next, students' expected grade had a moderate, positive, statistically significant correlation with all four of the constructs which indicates that as expected grade increases so does the degree to which their feedback is positive.

Research Question 3. To what extent does the MSECT-O data differ by course-level variables? As presented in Table 6, the extent to which students identified the course as being a requirement for their major or minor was not correlated with the MSECT-O constructs of practice and assessment. Indicating the course was a degree requirement, however, had a weak, positive correlation with the course climate and content. In other words, students provided more positive climate and content feedback if they perceived the course to be a requirement.

Finally, a Kruskal-Wallis H test provided evidence that there was not a significant difference in MSECT-O responses (across all four factors) between the difference course disciplines, $p > 0.05$. This suggests that, for example, students provided comparable feedback to STEM and to journalism instructors.

Discussion

In summary, this study found evidence that the MSECT-O is a valid and reliable instrument for gathering formative feedback from undergraduate students about online teaching practices. The CFA produced evidence that the reduced MSECT-O instrument is a valuable tool because it provides online instructors with formative feedback specific to the four factors of teaching effectiveness outlined in the Fearless Teaching Framework. With the MSECT-O, online instructors can collect feedback from their students across four dimensions and use the Fearless Teaching Framework to improve their teaching during the semester the data was collected. Institutions or instructors interested in the emerging practice of online MSEs (Thomas, Graham, & Piña, 2018), can use the free and brief MSECT-O instrument with confidence in its validity and reliability. The MSECT-O fills a gap in the field of online student evaluations because it is grounded in the education literature, designed to collect less biased formative feedback, and was developed with rigorous psychometrics.

The MSECT-O responses were generally consistent among students' year in school, their reason for taking the course (i.e., if it was a requirement), and the discipline of the course. The findings provide support for the notion that instructors from all disciplines can feel confident using MSECT-O to collect mid-semester feedback regardless of if their online class is in STEM versus the humanities. However, identifying the course as a requirement had a weak, positive correlation with a more positive evaluation of the class content and climate, i.e., that students in elective courses provided slightly more negative feedback to their instructors about the classroom climate and content. These results differ from that of Ting (2000) and suggests that more research is needed to understand the relationship between students' perceptions of required courses and the evaluations of the classroom climate and content.

Finally, the findings align with those of other student evaluation researchers (e.g., Spooen et al., 2017) that as students' expected course grade increases so does their positive evaluation of the instructor's teaching. While this correlation is only moderate, it is still significant and an area

of further investigation. future analyses will include a qualitative study of students' comments to explore the types of feedback they provide and if the constructiveness differs by course factors or expected grade.

Limitations

Limitations should be considered when interpreting the findings. First, the instructors who participated opted in, and therefore may be more open to student feedback than instructors who did not participate. Critics of SETs may not value or trust student feedback because the existing student-facing instruments ask questions that either students cannot meaningfully answer (e.g., regarding technology choices that require knowledge of the Learning Management System students might not have) or that instructors do not find valuable to improving their teaching (Peterson, 2016). In response to this perception, MSET-O measures were built to be clear and actionable.

Second, the effect of instructor-level variables such as instructor race and gender on students' responses were not explored because of the low number of instructors that participated. Future research will explore issues of bias by purposefully sampling a pool of instructors from diverse backgrounds.

Finally, although participating courses spanned disciplines, all student responses were from a single university. Therefore, further replication could assess the extent to which the measure provides robust information at other higher education contexts, such as small teaching colleges, and community colleges.

Conclusion

As more faculty, staff, and graduate students move their courses online, the MSET-O provides a useful and valid way to gather student input for improving the course climate, content, practices, and assessments. An important contribution of this paper is providing an instrument that gathers formative feedback for online instructors about multiple aspects of their teaching, including how they foster an inclusive online classroom climate. As more instructors move online, faculty developers can provide tools like the MSET-O to aid instructors in evaluating if they are providing an inclusive and equitable learning experience for all students and how they can further foster a climate in which all students feel supported.

References

- Alexander, P. A., & Winne, P. H. (2006). *Handbook of educational psychology* (2nd ed.). Erlbaum.
- Ambrose, S., Bridges, M., DiPietro, M., Lovett, M., & Norman, M. (2010). *How learning works: Seven research-based principles for smart teaching*. Jossey-Bass.
- Bangert, A. W. (2008). The development and validation of the student evaluation of online teaching effectiveness. *Computers in the Schools, 25*(1–2), 25–47. doi:10.1080/07380560802157717
- Berridge, G. G., Penney, S., & Wells, J. (2012). eFACT: Formative assessment of classroom teaching for online classes. *Turkish Online Journal of Distance Education, 13*(April), 119–131.
- Borup, J., & Evmenova, A. S. (2019). The effectiveness of professional development in overcoming obstacles to effective online instruction in a college of education. *Online Learning Journal, 23*(2), 1–20. doi:10.24059/olj.v23i2.1468
- Byrne, D. (2017). *Publishing your medical research* (2nd ed.). Wolters Kluwer.
- Byrne, V. L. (2018, April). *Student online course evaluation: How institutional researchers evaluate high-quality online teaching*. [Paper presentation]. American Educational Research Association Annual Meeting, New York, NY, United States.
- Carle, A. C. (2009). Evaluating college students' evaluations of a professor's teaching effectiveness across time and instruction mode (online vs. face-to-face) using a multilevel growth modeling approach. *Computers and Education, 53*(2), 429–435. doi:10.1016/j.compedu.2009.03.001
- Coffman, S. J. (1998). Small group instructional evaluations across disciplines. *College Teaching, 46*(3), 106–111.
- Cohen, P. A. (1980). Effectiveness of student-rating feedback for improving college instruction: A meta-analysis of findings. *Research in Higher Education, 13*(4), 321–41. doi:10.1007/BF00976252.
- Costello, M. L., Weldon, A. & Brunner, P. (2002). Reaction cards as a formative evaluation tool: Students' perceptions of how their use impacted classes. *Assessment and Evaluation in Higher Education, 27*(1), 23–33.
- Darby, F. (2019). *How to be a better online teacher*. The Chronicle of Higher Education. <https://www.chronicle.com/interactives/advice-online-teaching>
- Donlan, A. E., Loughlin, S., & Byrne, V. L. (2019). The Fearless Teaching Framework: A model to synthesize foundational education research for university instructors. *To Improve the Academy: A Journal of Educational Development 38*(1), 33–49. doi:10.1002/tia2.20087
- Donlan, A. E., & Byrne, V. L. (2020). Confirming the factor structure of a research-based mid-semester evaluation for college teaching. *Journal of Psychoeducational Assessment*. doi:10.1177/0734282920903165

- Gómez-Rey, P., Barbera, E., & Fernández-Navarro, F. (2016). Measuring teachers and learners' perceptions of the quality of their online learning experience. *Distance Education, 37*(2), 146–163. doi:10.1080/01587919.2016.1184396
- Hammer, J. (2016). *Construct replicability calculator: A Microsoft Excel-based tool to calculate the Hancock and Mueller (2001) H index*. <http://DrJosephHammer.com/>
- Hammonds, F., Mariano, G. J., Ammons, G., & Chambers, S. (2017). Student evaluations of teaching: Improving teaching quality in higher education. *Perspectives: Policy and Practice in Higher Education, 21*(1), 26–33. doi:10.1080/13603108.2016.1227388
- Hampton, S. E., & Reiser, R. A. (2004). Effects of a theory-based feedback and consultation process on instruction and learning in college classrooms. *Research in Higher Education, 45*(5), 497–527.
- Hancock, G. R., & Mueller, R. O. (2001). Rethinking construct reliability within latent variable systems. *Structural Equation Modeling: Present and Future, 195–216*.
- Hattie, J., & Timperley, H. (2007). The power of feedback the meaning of feedback. *Review of Educational Research, 77*(1), 81–112. doi:10.3102/003465430298487
- Hill, P. (2012). Online educational delivery models: A descriptive view. *EDUCAUSE Review, 47*, 84–86. <https://er.educause.edu/articles/2012/11/online-educational-delivery-models--a-descriptive-view>
- Howard, T. C. (2001). Telling their side of the story: African-American students' perceptions of culturally relevant teaching. *The Urban Review: Issues and Ideas in Public Education, 33*(2), 131–49. doi:10.1023/A:1010393224120.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55.
- Kline, R. B. (2011). *Principles and Practice of Structural Equation Modeling* (3rd ed.). Guilford Press.
- Knol, M. H., Dolan, C. V., Mellenbergh, G. J., & van der Maas, H. L. J. (2016). Measuring the quality of university lectures: Development and validation of the instructional skills questionnaire. *PLOS One, 11*(2), 1–21.
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge University Press.
- Liu, O. L. (2012). Student evaluation of instruction: In the new paradigm of distance education. *Research in Higher Education, 53*(4), 471–486. doi:10.1007/s11162-011-9236-1
- Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research, 11*(3), 253–388. doi:10.1016/0883-0355(87)90001-2
- Marsh, H. W., & Roche, L. A. 2000. Effects of grading leniency and low workload on students' evaluations of teaching: Popular myth, bias, validity, or innocent bystanders? *Journal of Educational Psychology, 92*(1), 202–28. doi:10.04.13/0022-0663.92.1.202
- McKeachie, W. J. (1997). Student ratings: The validity of use. *American Psychologist, 52*(11), 1218–1225.

- Meyer, K. (2014). An analysis of the cost and cost-effectiveness of faculty development for online teaching. *Online Learning, 18*(1). doi:10.24059/olj.v18i1.389
- Meyer, K. A., & Murrell, V. S. (2014). A national study of training content and activities for faculty development for online teaching. *Journal of Asynchronous Learning Networks, 18*, 3–18.
- Morin, A. J. S., Marsh, H. W., Nagengast, B., & Scalas, L. F. (2014). Doubly latent multilevel analyses of classroom climate: An illustration. *The Journal of Experimental Education, 82*(2), 143–67. doi:10.1080/00220973.2013.769412.
- O’Neil-Hixon, K., Long, J., & Bock, M. (2017). The eSGID process: How to improve teaching and learning in online graduate courses, *Journal of Effective Teaching, 17*(2), 45–57.
- Online Learning Consortium. (2019). *OLC quality scorecard*.
<https://onlinelearningconsortium.org/consult/olc-quality-scorecard-suite/>
- Overall, J. U., & Marsh, H. W. (1979). Midterm feedback from students: Its relationship to instructional improvement and students’ cognitive and affective outcomes. *Journal of Educational Psychology, 71*(6), 856–65. doi:10.1037/0022-0663.71.6.856.
- Peterson, J. L. (2016). Formative Evaluations in Online Classes. *Journal of Educators Online, 13*(1).
- Raykov, T., & Marcoulides, G. (2011). Classical item analysis using latent variable modeling: A note on a direct evaluation procedure. *Structural Equation Modeling, 18*(2), 315–324.
- Spooren, P. (2010). On the credibility of the judge: A cross-classified multilevel analysis on students’ evaluation of teaching. *Studies in Educational Evaluation, 36*(4), 121–131. doi:10.1016/j.stueduc.2011.02.001
- Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the validity of student evaluation of teaching: The state of the art. *Review of Educational Research, 83*(4), 598–642. doi:10.3102/0034654313496870
- Spooren, P., Vandermoere, F., Vanderstraeten, R., & Pepermans, K. (2017). Exploring high impact scholarship in research on student’s evaluation of teaching (SET). *Educational Research Review, 22*, 129–41. doi:10.1016/j.edurev.2017.09.001
- Thomas, J., Graham, C., & Piña, A. (2018). Current practices of online instructor evaluation in higher education. *Online Journal of Distance Learning Administration, 21*(2), 1–13.
- Thomas, J. E., & Graham, C. R. (2019). Online teaching competencies in observational rubrics: What are institutions evaluating? *Distance Education, 40*(1), 114–32. doi:10.1080/01587919.2018.1553564
- Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. American Psychological Association.
- Ting, K. F. (2000). A multilevel perspective on student ratings of instruction: Lessons from the Chinese experience. *Research in Higher Education, 41*(5), 637–61. doi:10.1023/A:1007075516271
- Walker, S. L. (2005). Modifying formative evaluation techniques for distance education class evaluation. *Turkish Online Journal of Distance Education, 6*(4), 7–11.

- Wass, R., Timmermans, J., Harland, T., & McLean, A. (2018). Annoyance and frustration: Emotional responses to being assessed in higher education. *Active Learning in Higher Education*, 1–13. doi:10.1177/1469787418762462.
- Wentzel, K. R. & Brophy, J. E. (2014). *Motivating Students to Learn* (4th ed.). Routledge.
- Wigfield, A. & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology*, 25(1), 68–81. doi:10.1006/ceps.1999.1015.
- Winchester, M. K., & Winchester, T. M. (2012). If you build it will they come? Exploring the student perspective of weekly student evaluations of teaching. *Assessment & Evaluation in Higher Education*, 37(6), 671–682. doi: 10.0.4.56/02602938.2011.563278