Journal of Online Learning Research (2023) **9**(1), 39-57

Evaluating Online Teaching Self-Efficacy and Effectiveness in Public PK12 Teachers

VIRGINIA L. BYRNE
Morgan State University, USA
virginia.byrne@morgan.edu

DIANE JASS KETELHUT
University of Maryland, College Park, USA
djk@umd.edu

KEYSHAWN MONCRIEFFE
Morgan State University, USA
keyshawn.moncrieffe@morgan.edu

As the COVID-19 pandemic forced schools to close their buildings, public PK12 educators in the US had to pivot to something they had never done before or been trained to do: teach online. While significant research exists on how educators learn to teach online in higher education and specialty PK12 settings, the pandemic provided a novel opportunity to explore how the general teaching population feels about and learns to teach online. We present the adoption of two online teaching self-evaluation instruments and initial findings from an ongoing study of PK12 teachers across 11 public school districts in a MidAtlantic state. Survey data sheds light on how to measure teachers' self-evaluations of their online teaching strategies and self-efficacy.

Keywords: teachers, covid-19, professional development, online teaching, self-efficacy

PURPOSE

The devastating COVID-19 pandemic forced public schools to close and transition all classrooms to what became known as Emergency Remote Teaching (ERT; Hodges et al., 2020). This rapid transition to ERT created a novel opportunity to explore online teaching and learning in PK12 public school classrooms. In particular, this transition was a unique time to explore the attitudes and experiences of seasoned public-school teachers who previously had not taught online, nor had they been trained to teach online. In this paper, we present our methodological approach of measuring PK12 teachers' self-efficacy and self-evaluations of online instruction, including initial reliability testing, as well as initial findings from an ongoing study of a large, multidistrict online teaching professional development program. The purpose of this paper is threefold: to introduce a revised online teaching self-evaluation instrument, which we referred to the Self-Evaluation of Online Teaching for PK12 (SEOT-PK12); to demonstrate the utility of two measures of Online Teaching Self-Efficacy (Robinia, 2008; Robinia & Anderson, 2010); and to present initial findings with these instruments. This paper answers the questions: *To what extent are two self-evaluation instruments of online teaching (that were originally designed for a higher education context) reliable among a sample of ~300 PK12 teachers in the Mid-Atlantic? How are teachers' responses to these instruments related?*

Previous Literature

Our work builds on the existing literature of online teaching evaluations which has traditionally been situated in higher education contexts (e.g., Byrne & Donlan, 2020; Gómez-Rey et al., 2016; Robinia & Anderson, 2010; Thomas et al., 2018), to explore the utility of these self-evaluation instruments in a PK12 setting. This project was motivated by the need to gauge PK12 teachers' online teaching attitudes during the COVID-19 pandemic and to explore the possible impacts of our professional development program. Our team reviewed the published, non-proprietary online teaching self-evaluation instruments, but found them to be nontransferable to a PK12 teacher sample (for example asking questions about course organization and testing that could not be translated to a pre-kindergarten teacher's classroom). We concluded that there is a gap in the literature for brief, general instruments that measures PK12 teachers' adoption of effective online teaching strategies that are neither discipline nor age-group specific.

Of particular interest to us were instruments that measure teachers' self-efficacy for online teaching, i.e., teachers' beliefs that they will succeed as an online teacher in the future (e.g., Robinia & Anderson, 2010). Bandura's (1997) self-efficacy theoretical framework explains and predicts future performance based on current self-perception. While often misunderstood, this perceived self-efficacy does not measure "the number of skills you have, but with what you believe you can do with what you have under a variety of circumstances" (Bandura, 2006, pg. 37). Self-efficacy instruments ask

people to imagine themselves performing successfully in a future, specific domain. The more closely a measurement can specify the teaching task and context, the more likely it is that teachers can predict their future success.

Teaching self-efficacy has been studied significantly (e.g., Glackin & Hohenstein, 2018; Skaalvik & Skaalvik, 2014; Tschannen-Moran et al., 1998), to reveal that it is positively related to teacher behavior and instructional decisions, as well as student outcomes. In other words, teacher self-efficacy (obtained either from past successful teaching experiences or professional learning experiences) is "related to the effort [teachers] invest in teaching, the goals they set, their persistence when things do not go smoothly, and their resilience in the face of setbacks" (Tschannen-Moran & McMaster, 2009, p.228). While still nascent, the research on Online Teaching Self-Efficacy (OTSE) has found that there is a positive relationship with teacher's prior experience with online instruction, participation in online teaching training (Robinia & Anderson, 2010), and sense of school connectedness (Yang, 2021), and a negative relationship with teachers' levels of compassion fatigue and burnout (Yang et al., 2021). Similar to face-to-face self-efficacy, professional development and coaching programs have been found to be positively related to OTSE (Robinia & Anderson, 2010). Our work explores PK12 teachers' OTSE and how their responses relate to their self-evaluation of their online teaching effectiveness.

METHODS

Our study is situated within a broader professional development (PD) program focused on online and hybrid instruction for public school teachers from 11 different districts across the same Mid-Atlantic state. The program was funded by a state grant and the Department of Education and is coordinated by a team of researchers and teacher educators from across the state. The program began in January 2021 and ended in March 2022. Approximately 450 teachers enrolled in this program in January 2021, which consisted of webinars, online discussion boards, and district-level meetings.

Data Collection

The program hosted its first Zoom-based webinar in January 2021. During the live webinar, the research team conducted the pretest survey. Teachers who viewed the January webinar later, as a recording, were encouraged to take the pretest survey on their own. In June 2021, the program hosted the last webinar for the 2021-2022 academic year. Teachers who had participated in the pretest survey were recruited to take a survey after this June webinar which we refer to as the midway survey. Finally, in March 2022 the program hosted the final webinar. The participants were again asked to complete a third survey which we refer to as the posttest. All surveys were voluntary, not affiliated with their completion requirements for the PD program, and had prior IRB approval.

Participants

While 300 people completed the pretest survey, only 90 people completed the midway survey, and 58 people completed the posttest survey. Participants' responses were matched so that we could run paired tests: 83 people fully completed the pretest and midway survey but only 32 people fully completed all three surveys. Table 1 reports on the demographic results of the participants at all three time points to recognize who persisted in the study and who did not. Among the 290 people who completed the pretest survey, the majority self-identified as women (71.88%) and white (88.19%). During the posttest survey we collected information about participants' professional positions, as presented in Table 1. While this posttest sample is only a small fraction of the pretest sample, the majority self-identified as classroom teachers or instructors. The sample reflected all grade levels and years of teaching experience.

Table 1

Participant Demographic Information at all Three Timepoints

		January 2021 (n = 300)		June 2021 (n = 90)		March 2022 (n = 58)	
Variable	Group	N	Percent	N	Percent	N	Percent
Gender Identity	Man	39	15.00%	8	10.81%	2	4.88%
	Prefer Not to Respond	2	0.77%	1	1.35%	1	2.44%
	Woman	219	84.23%	65	87.84%	38	92.68%
Race/Ethnicity	Asian, Asian-American, or Pacific Islander	5	1.67%	0	0.00%	2	4.17%
	Black and/or African American	14	4.867%	4	4.60%	1	2.08%
	Latinx and/or Hispanic	4	1.33%	0	0.00%	1	2.08%
	Multiracial or Mixed Race	6	2.00%	3	3.45%	1	2.08%
	Prefer Not to Respond	2	0.67%	1	1.15%	1	2.08%
	White	269	89.67%	79	90.80%	42	87.50%
Age	18-24	9	3.00%	3	3.45%	0	0.00%
	25-34	81	27.00%	22	25.29%	15	31.25%
	35-44	105	35.00%	38	43.68%	15	31.25%
	45-54	74	24.67%	19	21.84%	15	31.25%
	55-64	25	8.33%	4	4.60%	2	4.17%
	65-74	6	2.00%	1	1.15%	1	2.08%

Table 1, *Continued*

Years of Teaching Experience	1-2	21	7.37%	6	6.74%	1	2.04%
	3-5	31	10.88%	8	8.99%	3	6.12%
	6-10	59	20.70%	19	21.35%	10	20.41%
	11-15	66	23.16%	25	28.09%	12	24.49%
	16-20	53	18.60%	17	19.10%	10	20.41%
	21-25	29	10.18%	7	7.87%	8	16.33%
	26-30	28	9.82%	5	5.62%	4	8.16%
	31-35	4	1.40%	1	1.12%	1	2.04%
	36+	4	1.40%	1	1.12%	0	0.00%
School Level	Elementary	N/A	N/A	N/A	N/A	27	44.26%
	Middle	N/A	N/A	N/A	N/A	18	29.51%
	High	N/A	N/A	N/A	N/A	13	21.31%
	A Teacher of Adults	N/A	N/A	N/A	N/A	3	4.91%
Position	Teacher/ Instructor	N/A	N/A	N/A	N/A	45	71.43%
	Administrator	N/A	N/A	N/A	N/A	2	3.17%
	Specialist	N/A	N/A	N/A	N/A	6	9.52%
	Instructional Coach	N/A	N/A	N/A	N/A	10	15.87%
Prior Formal Training in Online Teaching	Yes	96	30.28%	29	32.22%	17	32.69%
	No	203	64.04%	55	61.11%	32	61.54%
	Unsure	18	5.68%	6	6.67%	3	5.77%

Not at all participants completed all demographics items. Percentages out of the number of responses.

Instruments

The three surveys each consisted of questions about the participants' teaching role and responsibilities, mental health, demographics, online teaching strategies, self-efficacy, and email address. Of particular importance to this paper are three of the instruments included in each survey: the Self-Evaluation of Online Teaching for PK12 (SEOT-PK12) and two 8-item subscales of Online Teaching Self-Efficacy designed by Robinia (2008) that measure *Self-Efficacy for Online Student Engagement* and *Self-Efficacy for Online Instructional Strategies*. We present information about each instrument below.

First, we implemented a redesigned version of the Mid-Semester Evaluation of College Student Teaching (MSECT; Donlan & Byrne, 2020) to serve as a self-evaluation tool for PK12 teachers which we refer to as the Self-Evaluation of Online Teaching for PK12 (SEOT-PK12). The original instrument was designed in alignment with a literature-based framework for effective college teaching that found evidence for four interconnected concepts: classroom *Climate*, active teaching *Practices*, reliable *Assessments*, and relevant course *Content* (Donlan et al., 2019). After developing a first iteration of survey items, we determined the content validity by gathering feedback from a committee of education researchers and expert faculty developers (Donlan & Byrne, 2020). Next, we confirmed the construct validity and the four-aspect structure among a sample of 1,350 undergraduate students in face-to-face courses at a large Predominantly White Institution (PWI) by conducting an exploratory factor analysis on one half of the student sample, and confirmatory factor analysis on the other using MPlus 8.0 (Donlan & Byrne, 2020). We found the four-factor model to be a good fit and recommended that the scale could be used by instructors to assess the four aspects of their face-to-face college teaching. The result was a 13-item survey using a 6-point scale in which 1 represented Strongly Disagree and 6 represented Strongly Agree.

Then, in 2020 we revised the four-factor instrument for an online college course setting (referred to as the Mid-Semester Evaluation of College Student Teaching – Online or MSECT-O) and piloted the instrument among 170 undergraduate students in seven online courses at a large PWI in the Mid-Atlantic (Byrne & Donlan, 2020). We conducted a confirmatory factor analysis in MPlus 8.0 and found evidence that the revised instrument adequately measures the four factors of online teaching effectiveness determined by our evidence-based framework (i.e., Donlan et al., 2019). We also determined factor replicability and reliability based on sufficient H-index values (Hammer, 2016; Hancock & Mueller, 2001) and Cronbach's alpha values (Raykov & Marcoulides, 2011).

To better align with our study participants, we rephrased the MSECT-O item language to be more appropriate for PK12 classroom teachers and to serve as a self-evaluation. The revised survey items can be found in Table 2. Because we were conducting the survey with pre-kindergarten through high school teachers, we chose to omit the items for the course *Content* construct, assuming that, unlike many higher education faculty, some PK12 teachers were not fully in control of deciding what content to cover in their classes. Thus, we collected data using the revised MSECT-O instrument for the constructs of *Climate* (three items), *Practices* (three items), and *Assessments* (three items) which we refer to as the Self-Evaluation of Online Teaching for PK12 (SEOT-PK12).

Table 2

The Self-Evaluation of Online Teaching for PK12 (SEOT-PK12) Instrument

CLIMATE To what extent do you agree with the following statements about your new ONLINE teaching environment? If a question does not apply to your teaching, select "N/A." 1. I have created an online environment that is supportive for learning. 2. I have made my class accessible to students with many different needs. 3. I have created an inclusive learning environment where everyone is welcome.
PRACTICES To what extent do you agree with the following statements about your new ONLINE teaching environment? If a question does not apply to your teaching, select "N/A." 1. During synchronous sessions, my class includes in-class activities other than lecture. 2. I have helped students understand new content by connecting it to things they already understand. 3. I have motivated students to put effort into my class.
ASSESSMENT To what extent do you agree with the following statements about your new ONLINE teaching environment? If a question does not apply to your teaching, select "N/A." 1. The assessments (e.g., quizzes, exams, papers) in my class are graded fairly. 2. I have provided timely feedback on students' work. 3. My expectations for the assignments are clear.

Second, our surveys included two 8-item subscales of Online Teaching Self-Efficacy designed by Robinia (2008) that measured *Self-Efficacy for Online Student Engagement* and *Self-Efficacy for Online Instructional Strategies* but modified with a 100-point sliding scale (1 = nothing; 100 = a great deal). Items include "How much can you do to motivate students who show low interest in online work?" and "How much can you do to get students to believe that they can do well in an online class?" (Robinia, 2008). These scales were developed as part of the larger *Michigan Nurse Educators Sense of Efficacy for Online Teaching* instrument which was based on the *Teacher's Sense of Self-Efficacy Scale* (Tschannen-Moran & Hoy, 2001) but modified to fit an online learning environment. Robinia's aim was to examine the influences on online teacher self-efficacy in a higher education setting. Their prior work (e.g., Robinia, 2008; Robinia & Anderson, 2010) found these scales to be valid measures of educators' online teaching self-efficacy via an expert panel and a reliable measure among 140 nurse educators as

assessed by Cronbach's alpha (*Self-Efficacy for Online Student Engagement* α=.93; *Self-Efficacy for Online Instructional Strategies* α=.94). In 2020, Robinia's (2008) entire instrument was deployed among 351 Chinese school teachers and found it to be a valid and reliable measure (Ma et al., 2021). Otherwise, from our review of the literature, we found that these subscales have not been previously utilized in a PK12 setting in the U.S.

Data analysis

Participants' survey responses were linked by their email address. We assessed the reliability of the constructs of *Self-Efficacy for Online Student Engagement* and *Self-Efficacy for Online Instructional Strategies* and the three constructs of the SEOT-PK12 using Cronbach's alpha across three time points (Raykov & Marcoulides, 2011). Before conducting our analyses, we checked for normality among the pretest survey data using SPSS. A Shapiro-Wilk test of normality provided evidence that the SEOT-PK12 constructs of *Climate, Practices*, and *Assessment* suffered from issues of non-normality ($p < .05$). However, the aggregated self-efficacy items demonstrated tendencies of normality. For these reasons, we conducted robust nonparametric statistics tests in SPSS (Byrne, 2017).

Specifically, to test if there was a difference in teachers' responses between the pretest, midway, and posttest surveys, we adopted the same procedure for each of the constructs of interest. We conducted the nonparametric alternative to the one-way ANOVA with repeated measures called the Friedman Test. Then, to determine if any differences were significant, we ran Wilcoxon signed-rank tests with a Bonferroni adjustment on the different combinations of time points (pretest and midway, pretest and posttest, and midway and posttest). We calculated the Bonferroni adjustment by simply dividing our initial significance level ($p = 0.05$) by the number of tests we ran (3) resulting in $0.05/3 = 0.017$. Thus, our significance level was $p < 0.017$.

We then ran tests of difference by the teachers' school level (i.e., elementary, middle, and high school). First, we removed the teachers of adults because they were such a small population (n=3). We then conducted Kruskal-Wallis H tests to determine if there was a statistical difference in the constructs during any of the time points by school level.

Finally, we ran tests of difference if teachers had or had not previously completed any formal training or coursework related to online teaching. To determine if teachers with prior training answered the questions differently, we ran another series of Kruskal-Wallis H tests.

FINDINGS

For each of the survey instruments, we present results from our reliability and descriptive analyses as well as the tests of difference across the three time periods.

Self-Evaluation of Online Teaching Effectiveness

In Table 2 we present descriptive findings of the responses to the Self-Evaluation of Online Teaching for PK12 (SEOT-PK12) instrument at three timepoints. The participants reported a high average of self-reported adoption of effective online teaching practices across the three constructs (5 to 5.5 out of 6). The participants' self-reported evaluations increased for all three constructs across all three timepoints.

Climate

A Friedman Test of the 28 participants who fully completed the *Climate* instrument at the three timepoints (i.e., the pretest, midway, or posttest survey) found that there was a statistically significant difference in the adoption of effective *Climate* practices, $\chi^2(2) = 6.58$, $p = .04$ (pretest mean rank was 1.64, midway mean rank was 2.25, posttest mean rank was 2.11). Median climate scores were 5.00 (pretest), 5.00 (midway), and 5.67 (posttest). From the Wilcoxon signed-rank tests with a Bonferroni adjustment we determined a significant increase between the pretest and midway surveys ($Z = -2.51$, $p = .012$) and between the pretest and posttest surveys ($Z = -3.29$, $p = .001$). However, there was not a significant difference between the midway and posttest surveys ($Z = -0.45$, $p = .66$).

Practices

A Friedman Test of the 27 participants who completed the *Practices* instrument at the three timepoints found that there was not a statistically significant difference in the adoption of effective online teaching Practices, $\chi^2(2) = 3.90$, $p = .14$ (pretest mean rank was 1.74, midway mean rank was 2.07, posttest mean rank was 2.19). Median *Practices* scores were 5.00 (pretest), 5.33 (midway), and 5.67 (posttest). The Wilcoxon signed-rank tests with a Bonferroni adjustment determined that there was a significant increase between the pretest and posttest surveys ($Z = -2.57$, $p = .010$). However, there was no significant difference (at the .017 level) between the between the pretest and midway surveys ($Z = -1.78$, $p = .08$) or between midway and posttest surveys ($Z = -1.55$, $p = .12$).

Assessment

A Friedman Test of the 26 participants who completed the *Assessment* instrument at the three timepoints found that there was a statistically significant difference in the adoption of effective *Assessment* practices, $\chi^2(2) = 11.69$, $p = .003$ (pretest mean rank was 1.60, midway mean rank was 2.02, posttest mean rank was 2.38). Median *Assessment* scores were 5.17 (pretest), 5.50 (midway), and 5.50 (posttest). From the Wilcoxon signed-rank tests with a Bonferroni adjustment was a significant increase between the pretest and midway surveys ($Z = -3.27$, $p = .001$) and between the pretest and posttest surveys ($Z = -3.69$, $p < .001$). However, there was not a significant difference between the midway and posttest surveys ($Z = -2.24$, $p = .03$).

In conclusion, we found that between the pretest and the posttest surveys teachers reported statistically significant higher evaluations of their use of effective online teaching decisions that promote a positive classroom *Climate* and reliable *Assessments*, but not active teaching Practices.

Reliability

Among the pretest and midway survey data, our tests resulted in Cronbach Alpha values that can be interpreted as acceptable to good reliability (between .75 - .85). However, as presented in Table 3, the results among posttest responses were unacceptably low. These low Alpha values suggest that this instrument may not have been reliable for the posttest data.

Table 3

Descriptive Findings of the Self-Evaluation of Online Teaching for PK12 (SEOT-PK12) Constructs

	Climate			Practices			Assessment		
	January 2021	June 2021	March 2022	January 2021	June 2021	March 2022	January 2021	June 2021	March 2022
N	290	87	50	287	83	50	273	81	50
Alpha	0.76	0.85	0.53	0.74	0.79	0.66	0.74	0.80	0.58
Mean	5.04	5.19	5.45	5.03	5.17	5.46	5.09	5.31	5.49
Median	5.00	5.00	5.67	5.00	5.33	5.67	5.00	5.33	5.67
Standard Deviation	0.67	0.70	0.49	0.71	0.75	0.51	0.72	0.75	0.51
Standard Error	0.39	0.08	0.07	0.04	0.08	0.07	0.04	0.08	0.07

Table 4
Descriptive Statistics of the Two Online Teaching Self-Efficacy Measures

	Self-Efficacy for Online Student Engagement			Self-Efficacy for Online Instructional Strategies		
	January 2021	June 2021	March 2022	January 2021	June 2021	March 2022
N	289	90	58	293	93	58
Alpha	0.88	0.88	0.93	0.90	0.90	0.93
Mean	59.08	69.52	72.90	68.51	78.61	81.22
Median	60.25	70.25	75.00	69.38	81.25	84.38
Standard Deviation	15.99	14.89	15.83	15.81	13.50	14.01
Standard Error	0.94	1.58	2.08	0.92	1.40	1.84

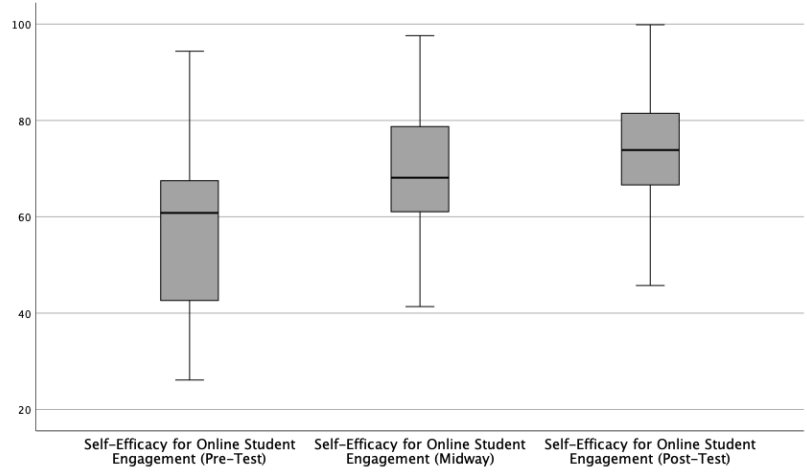
Online Teaching Self-Efficacy

Next, we present the findings of the *Self-Efficacy for Online Student Engagement* and *Self-Efficacy for Online Instructional Strategies* scales (Robinia, 2008; Robinia & Anderson, 2010). During the pretest data collection in January 2021, the participating teachers self-reported moderate levels of self-efficacy for their online teaching (mean *Self-Efficacy for Online Student Engagement* = 59.08 out of 100; mean *Self-Efficacy for Online Instructional Strategies* = 68.51; see Table 3 for descriptive statistics). In other words, the average participant in our sample was 59% out of 100% confident in their ability to meaningfully engage their students in an online classroom and 68% confident in their ability to employ high quality online teaching strategies. In June 2021, the average self-efficacy increased (mean *Self-Efficacy for Online Student Engagement* = 69.52; mean *Self-Efficacy for Online Instructional Strategies* = 78.61) suggesting that teachers were now more confident in their online teaching than they were in January. Then, in March 2022, the average self-efficacy increased even further (mean S*elf-Efficacy for Online Student Engagement* = 72.90; mean *Self-Efficacy for Online Instructional Strategies* = 81.22) suggesting that teachers were now even more confident in their online teaching than they were in January or June 2021.

A nonparametric repeated measures Friedman Test of the 32 participants who completed the survey at the three timepoints (i.e., the pretest, midway, or posttest survey) found that there was a statistically significant difference in both self-reported *Self-Efficacy for Online Student Engagement* ($\chi^2(2) = 21.00$, $p < .001$; pretest mean rank was 1.38, midway mean rank was 2.13,

posttest mean rank was 2.50) and *Self-Efficacy for Online Instructional Strategies* ($\chi^2(2) = 15.73$, $p < .001$; pretest mean rank was 1.48, midway mean rank was 2.05, posttest mean rank was 2.47). To examine if the differences between the three time periods were significant, we conducted post hoc separate Wilcoxon signed-rank tests with a Bonferroni adjustment and found that for both measures there was a significant increase between the pretest and midway surveys (*Self-Efficacy for Online Student Engagement*: Z = -3.61, $p < .001$; *Self-Efficacy for Online Instructional Strategies*: Z = -3.23, $p = .001$) and between the pretest and posttest surveys (*Self-Efficacy for Online Student Engagement*: Z = -4.34, $p < .001$; S*elf-Efficacy for Online Instructional Strategies*: Z = -3.98, $p < .001$). However, there was no significant difference between the midway and posttest surveys (*Self-Efficacy for Online Student Engagement*: Z = -1.90, $p = .06$; *Self-Efficacy for Online Instructional Strategies*: Z = -1.41, $p = .16$). As presented in the boxplots in Figures 4 and 5, the median *Self-Efficacy for Online Student Engagement* values were 60.81 (pretest), 68.13 (midway), and 73.88 (posttest) and the median values were 67.25 (pretest), 81.88 (midway), and 83.81 (posttest).

Reliability. As presented in Table 4, the Cronbach Alpha values demonstrate that both self-efficacy measures were reliable across all three time periods.



Figures 1. Boxplots of Self-Efficacy for Online Student Engagement at Three Timepoints (N = 32).

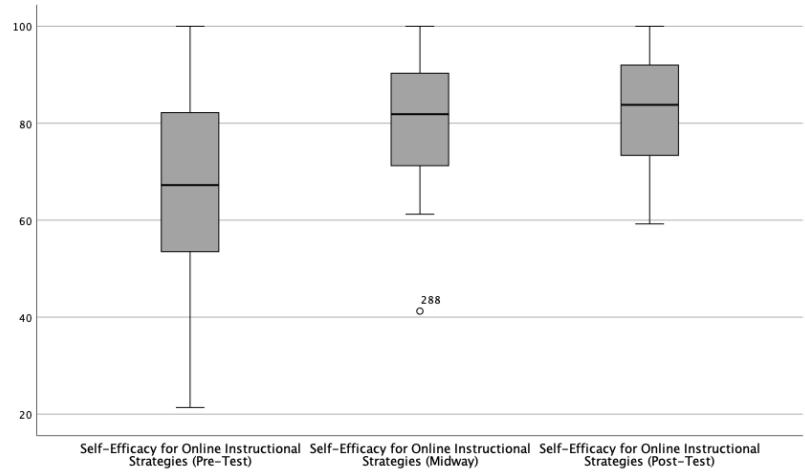


Figure 2. Boxplots of Self-Efficacy for Online Instructional Strategies at Three Timepoints (N = 32).

Relationship between Self-Efficacy and Self-Evaluation Constructs

As presented in Table 5, we calculated the nonparametric Spearman correlation coefficients (rho) for the three SEOT-PK12 and the two self-efficacy constructs at all three-points. First, we found that the repeated measures of Self-Efficacy for Online Student Engagement and Self-Efficacy for Online Instructional Strategies were significantly, positively correlated at all three timepoints ($p < .01$), as expected. In other words, people who reported higher self-efficacy in January 2021 also reported higher self-efficacy in June 2021 and then March 2022. Second, we found that among each timepoint the self-efficacy measures are significantly, positively correlated with the Climate and Practices constructs, suggesting that those with higher self-efficacy for online teaching also have a high self-evaluation of these online teaching practices. However, the self-efficacy measures were not consistently correlated with the Assessment constructs.

Relationship between School Type, Self-Efficacy, and Self-Evaluation

We then ran a series of Kruskal-Wallis H tests to determine if there was a relationship between teacher's grade level and their responses to the three SEOT-PK12 and the two self-efficacy constructs at any of the three time points. None of these tests found statistically significant differences (at the $p<.05$ level) in self-efficacy or the SEOT-PK12 constructs between elementary, middle, and high school teachers at any of the time points.

Relationship between Prior Online Teaching Training, Self-Efficacy, and Self-Evaluation

We ran another series of Kruskal-Wallis H tests to determine if teachers' prior training in online teaching was related to their self-efficacy and self-evaluation. The results were mixed. We found statistically significant differences between teachers who said they had received training in online teaching prior to 2021 and those who said they had not for *Self-Efficacy for Online Student Engagement* pre-test responses (U= 5.05, p = 0.025; mean rank of educators with prior training = 151.87, without prior training = 129.02), *Self-Efficacy for Online Instructional Strategies* pre-test responses (U= 11.38, $p < .001$; mean rank of educators with prior training = 161.61, without prior training = 127.13), and the midway SEOT-PK12 constructs of *Climate* (U = 521.00, p = 0.023; mean rank of educators with prior training = 33.11, without prior training = 45.17), *Assessment* (U = 435.50, p = 0.016; mean rank of educators with prior training = 30.13, without prior training = 42.43), and *Practices* (U = 477.00, p = 0.032; mean rank of educators with prior training = 31.67, without prior training = 42.96). However, we did not find significant relationships between three SEOT-PK12 constructs from the pretest or posttest, or between the self-efficacy measures from the midway or posttest surveys.

DISCUSSION

In this paper, we share findings from the design and implementation of two instruments measuring online teaching attitudes and behaviors among a sample of PK12 public school teachers who had to pivot to online and hybrid teaching during the COVID-19 pandemic. Using the Self-Evaluation of Online Teaching for PK12 (SEOT-PK12) instrument and two scales of self-efficacy (*Self-Efficacy for Online Student Engagement* and *Self-Efficacy for Online Instructional Strategies*), our surveys captured baseline data of teachers new to online teaching (pretest data) as well as how their self-evaluations and self-efficacy changed throughout the program and the pandemic. We found that at the start of our program, teachers on average expressed moderate self-efficacy for online teaching and agreed that they enacted effective online teaching strategies. These self-reports steadily increased as time went on, even though fewer and fewer participants completed our surveys. Between January 2021 (pretest) and March 2022 (posttest), teachers self-reported significantly higher levels of self-efficacy as well as the adoption of online teaching strategies aligned with promoting a more positive online classroom *Climate* and reliable *Assessment*.

From our tests of reliability, we determined that the *Self-Efficacy for Online Student Engagement* and *Self-Efficacy for Online Instructional Strategies* instruments were highly reliable at each timepoint. However, our

Table 5

Spearman Correlation of SEOT-PK12 & Self-Efficacy Constructs at Three Time Points

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Self-Efficacy for Online Student Engagement Pre-Test	Spearman's rho	--														
	N	289														
2. Self-Efficacy for Online Student Engagement Midway	Spearman's rho	.501**	--													
	N	86	90													
3. Self-Efficacy for Online Student Engagement Post-Test	Spearman's rho	.405**	.759**	--												
	N	46	33	58												
4. Self-Efficacy for Online Instructional Strategies Pre-Test	Spearman's rho	.766**	.443**	0.226	--											
	N	282	86	48	293											
5. Self-Efficacy for Online Instructional Strategies Midway	Spearman's rho	.430**	.814**	.676**	.454**	--										
	N	88	90	34	89	93										
6. Self-Efficacy for Online Instructional Strategies Post-Test	Spearman's rho	.470**	.781**	.834**	.430**	.819**	--									
	N	46	33	58	48	34	58									
7. Climate Pre-Test	Spearman's rho	.448**	.313**	0.242	.481**	0.203	.364*	--								
	N	276	86	47	282	89	47	290								
8. Climate Midway	Spearman's rho	0.181	.443**	0.325	0.099	.392**	0.313	.474**	--							
	N	82	84	33	83	87	33	84	87							
9. Climate Post-Test	Spearman's rho	.337*	0.275	.472**	.319*	0.239	.511**	0.271	.485**	--						
	N	39	28	49	42	29	49	41	28	50						
10. Practices Pre-Test	Spearman's rho	.382**	.237*	0.233	.412**	0.119	.334*	.531**	.302**	0.19	--					
	N	275	85	47	281	88	47	284	83	41	287					
11. Practices Midway	Spearman's rho	0.1	.365**	0.08	0.086	.362**	0.128	.311**	.641**	.536**	.408**	--				
	N	78	81	32	81	83	32	81	83	27	80	83				
12. Practices Post-Test	Spearman's rho	0.272	.483**	.473**	.346*	.480**	.596**	0.263	.441*	.780**	0.196	.425*	--			
	N	39	28	49	42	29	49	41	28	50	41	27	50			
13. Assessment Pre-Test	Spearman's rho	.209**	0.083	0.22	.307**	0.112	.302*	.419**	0.178	.364*	.494**	.369**	.422**	--		
	N	262	79	44	267	82	44	271	77	39	271	75	39	273		
14. Assessment Midway	Spearman's rho	0.059	0.269	0.256	0.198	0.275	.288*	0.229	0.353	.571**	0.082	.789**	.653**	.618**	--	
	N	39	28	49	42	29	49	41	28	50	41	27	50	39	50	
15. Assessment Post-Test	Spearman's rho	-0.035	.264*	0.275	0.012	.285**	0.275	.288*	.422**	.468*	.384**	.661**	0.358	.566**	.814**	--
	N	77	78	31	79	81	31	79	81	27	79	79	27	75	27	81

** Correlation is significant at the 0.01 level (2-tailed). * Correlation is significant at the 0.05 level (2-tailed).

finding that the SEOT-PK12 constructs are only reliable at the pretest and midway survey points is interesting. We wonder if the posttest instruments were not reliable because the instrument asks teachers to report on what they are *currently* doing in their online classes and, at the time of the posttest survey, teachers had returned to their physical classrooms. We wonder if the SEOT-PK12 is only reliable when teachers are teaching online or hybrid, but not when they are teaching face-to-face.

We also found that the both self-efficacy values are significantly positively correlated across the three timepoints and are correlated with many of the three SEOT-PK12 constructs. These results are consistent with the existing research on self-efficacy (e.g., Tschannen-Moran et al., 1998) which demonstrates how self-efficacy increases with experience and time on task – topics we elaborate on in our future analyses. We conclude that these significant correlations are an initial validation of the constructs, but further analysis is needed.

Finally, we tested for differences in responses by teachers' school level and prior online teaching training. First, we found that teachers' responses were not related to their school level. We interpret this finding to mean that these instruments can be useful to administrators and professional development designers working with teachers across school levels. Second, we found that teachers' prior participation in formal online teaching training was only partially related to the constructs measured. While this finding could be interpreted in several ways, we posit that the instrument is accessible, in some part, to teachers who are unfamiliar with online education jargon. Meaning that teachers without prior training in online teaching were able to understand and relate the instruments to their own teaching, in some way. More analysis is needed to better understand these findings.

As Corry and Stella (2018) stated, there is a gap in rigorous empirical validation of the relationship between online teacher self-efficacy and online student achievement. This paper moves the field closer to closing this gap by clarifying the reliability and potential utility of these two instruments for a PK12 setting. Future researchers could adopt these instruments to rigorously evaluate the impact of a professional development program on teachers' OTSE. Similarly, education leaders and teacher educators could adopt these instruments to explore trends in teachers' OTSE and self-evaluation of online teaching effectiveness for formative feedback or development plans.

Limitations

Our work is limited in several respects. First, our data suffered from significant participant dropout between the pretest and posttest. This is related to the PD design and district policies which are outside of the scope of

this study. Second, we recognize that self-reported evaluation data suffers from validity issues and desirability bias, but still see value in using these instruments as part of a suite of methods for teachers to document their own growth. Third, while we do not make any causal claims about the relationship between teacher development and their participation in the PD program, more evaluation work is needed.

CONCLUSION

This paper presented initial validation and reliability finding of two instruments among a diverse population of PK12 teachers. We posit that the two instruments were reliable and useful for gauging teachers' self-evaluation of online teaching practices and their self-efficacy for online teaching. Furthermore, we believe our instruments will be useful for other researchers and teacher educators who seek self-reported measures of knowledge, adoption, and growth.

DECLARATIONS

The authors declare no conflicts of interest.
This research was funded in part by a grant from the State of Maryland.
Ethic board permission/human subjects approval for this work
was granted from the University of Maryland.

References

Bandura, A. (1997). *Self-efficacy: The exercise of control.* Macmillan.

Bandura, A. (2006). Guide for constructing self-efficacy scales. *Self-efficacy beliefs of adolescents, 5*(1), 307-337.

Byrne, D. (2017). Publishing your medical research (Second Ed.). Wolters Kluwer.

Byrne, V. L. & Donlan, A. E. (2020). A mid-semester evaluation of college teaching to improve online teaching effectiveness. *Online Learning Journal, 24*(2). doi.org/10.24059/olj.v24i2.2126

Corry, M., & Stella, J. (2018). *Teacher self-efficacy in online education: A review of the literature.* Publisher. https://scholarworks.iupui.edu/handle/1805/21251

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, 38*(7), 866–881. doi.org/10.1177/0734282920903165

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.org/10.1002/tia2.20087

Glackin, M., & Hohenstein, J. (2018). Teachers' self-efficacy: Progressing qualitative analysis. International *Journal of Research and Method in Education, 41*,271–290. https://doi.org/10.1080/1743727X.2017. 1295940.

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*. Accessed from http://DrJoseph-Hammer.com/

Hancock, G R. & Mueller, R. O. (2001). Rethinking construct reliability within latent variable systems. *Structural Equation Modeling: Present and future,* 195-216.

Hodges, C., Moore, S., Lockee, B., Trust, T., & Bond, A. (2020, March 27). The difference between emergency remote teaching and online learning. *EDUCAUSE Review, 3.* https://er.educause.edu/ articles/2020/3/ the-difference-between-emergency-remote-teaching-and-online-learning

Ma, K., Chutiyami, M., Zhang, Y., & Nicoll, S. (2021). Online teaching self-efficacy during COVID-19: Changes, its associated factors and moderators. *Education and Information Technologies, 26*(6), 6675–6697. https://doi.org/10.1007/s10639-021-10486-3

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.

Robinia, K. A. (2008). [DISSERTATION] *Online teaching self-efficacy of nurse faculty teaching in public, accredited nursing programs in the state of Michigan*. Western Michigan University.

Robinia, K. A., & Anderson, M. L. (2010). Online teaching efficacy of nurse faculty. *Journal of Professional Nursing, 26*(3), 168–175. https://doi.org/10.1016/j.profnurs.2010.02.006

Skaalvik, E. M., & Skaalvik, S. (2014). Teacher self-efficacy and perceived autonomy: Relations with teacher engagement, job satisfaction, and emotional exhaustion. *Psychological reports, 114*(1), 68-77. https://doi.org/10.2466/14.02.PR0.114k14w0

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.

Tschannen-Moran, M., & Hoy, A. W. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and teacher education, 17*(7), 783-805. https://doi.org/10.1016/S0742-051X(01)00036-1

Tschannen-Moran, M., Hoy, A. W., & Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research, 68(*2), 202-248.

Tschannen-Moran, M., & McMaster, P. (2009). Sources of self-efficacy: Four professional development formats and their relationship to self-efficacy and implementation of a new teaching strategy. *The Elementary School Journal, 110*(2), 228–245. https://doi.org/10.1086/605771

Yang, C. (2021). Online teaching self-efficacy, social–emotional learning (sel) competencies, and compassion fatigue among educators during the COVID-19 Pandemic. *School Psychology Review, 50*(4), 505–518. https://doi.org/10.1080/2372966X.2021.1903815

Yang, C., Manchanda, S., & Greenstein, J. (2021). Educators' online teaching self-efficacy and compassion fatigue during the COVID-19 pandemic: The dual roles of "connect". *School Psychology, 36*(6), 504–515. https://doi.org/10.1037/spq0000475