INSTRUCTOR AND INSTRUCTIONAL EFFECTS ON STUDENTS' STATISTICS ATTITUDES

CHAO XU
University of Houston - Clear Lake
xuc567920@163.com

MICHELLE PETERS
University of Houston - Clear Lake
petersm@uhcl.edu

SUE BROWN
University of Houston - Clear Lake
browns@uhcl.edu

ABSTRACT

Using data from 23 statistics instructors and 1,924 students across 11 post-secondary institutions in the United States, we employ multilevel covariate adjustment models to quantify the sizes of instructor and instructional effects on students' statistics attitudes. The analysis suggests that changes in students' statistics attitudes vary considerably across statistics instructors. Instructor-associated changes in students' statistics attitudes are positively associated with instructional practices most proximal to tasks involving data as well as with instructors' attitudes toward teaching their statistics classes. Moreover, instructor-associated changes in students' statistics attitudes are positively related to changes in students' expected grades. These findings lend support to previous qualitative findings about links between certain dimensions of teaching practices and students' statistics attitudes.

Keywords: *Statistics education research; SATS-36; Instructor effectiveness; Multilevel modeling*

1. INTRODUCTION

Statistics education is becoming an essential component of higher education, largely due to an increased awareness of the utilitarian value of statistics in everyday life. Most students, however, take only one introductory statistics course while they are in college. This limited exposure to formal statistics instruction poses a great challenge to instructors if the ultimate goal is to enhance student understanding of statistics. The *Guidelines for Assessment and Instruction in Statistics Education* (GAISE; GAISE College Report ASA Revision Committee, 2016) report recognizes the pivotal role that instructors play in developing student learning outcomes. The report includes six recommendations intended to help instructors improve introductory statistics courses. Although the GAISE report makes a clear case that learning outcomes and instructors are intrinsically intertwined, not enough is known about how instructor and instructional effects can be identified based on learning outcomes.

"Instructor effects" are a measure of differences in instructor effectiveness or the extent to which changes in student outcome scores vary across individual instructors after controlling for pre-course scores on the same outcome, classroom peer effects, and a range of student-level covariates in relation to this learning outcome (Nye et al., 2004). The effects of instructors on learning outcomes may be accounted for by instructor characteristics and, more likely, instructional quality. In this sense, instructional effects are defined as the (potentially causal) relationships between instructional practice and learning outcomes (Blazar & Kraft, 2017).

In the present study, we use multilevel covariate adjustment models to examine instructor and instructional effects on students' statistics attitudes. Specifically, we ask the following three questions:

Statistics Education Research Journal, 19(2), 7-26, http://iase-web.org/Publications.php?p=SERJ

1. How much do changes in students' statistics attitudes vary across statistics instructors?
2. How much do self-reported instructional practices account for the instructor-level variation in changes in students' statistics attitudes?
3. Are instructor-associated changes in students' statistics attitudes related to changes in students' expected grades?

We diagram a conceptual overview of this study in Figure 1, which visualizes the three research questions. To answer them, we draw on a large set of data collected by the SATS project (Schau & Emmioğlu, 2012). In addition to student responses to a survey that captures statistics attitudes, this rich data set also includes instructor-reported perceptions of instructional practices collected on a separate instructor questionnaire.

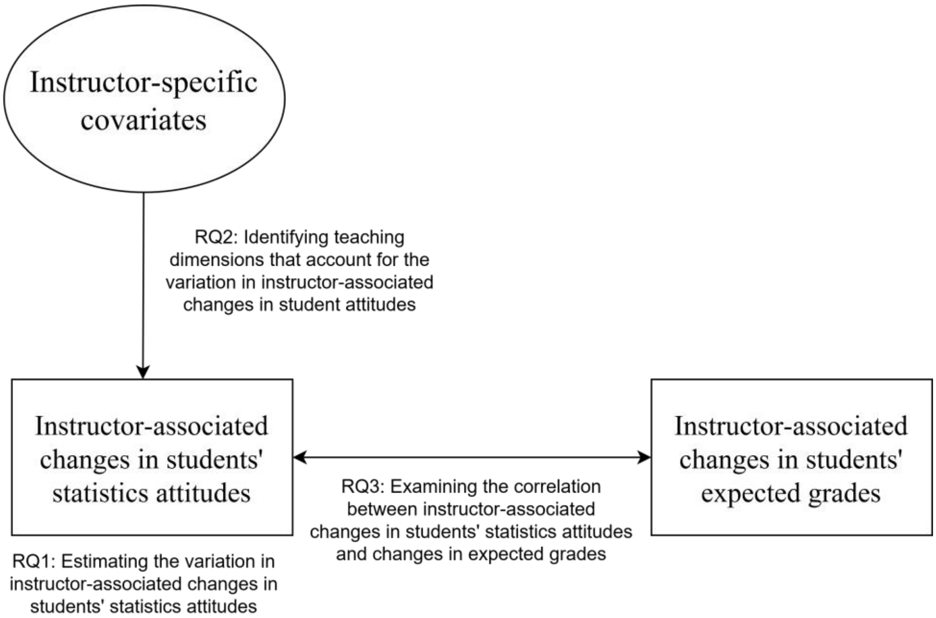

Figure 1. Conceptual overview of the research questions

2. LITERATURE REVIEW

2.1. TEACHER EFFECTS ON STUDENT ATTITUDES

Student learning goes beyond attainment of core academic knowledge and skills. Academic attitudes, as a primary learning outcome, influence how much students learn in school and how willing they are to use what they have learned outside of school (Duckworth et al., 2012). Longitudinal studies have shown that children with more positive academic attitudes in the early years have better long-term outcomes, such as earnings and educational attainment (Chetty et al., 2011). In this regard, instructors should be expected not only to effectively deliver content but also to improve students' academic attitudes.

Pertinent to statistics education, Waters et al. (1988) found that, among 10 course sections taught by part-time faculty or graduate students, students in some sections had higher statistics attitudes at the end of the courses than at the beginning whereas students in others did not. Schau (2003) reported that, among 11 course sections, the section-to-section variability in average attitudes scores became larger towards the end of semesters. This finding indicates that instructors may affect students' statistics attitudes differentially. Schau further identified instructor and instructional characteristics as one determinant of students' statistics attitudes. Drawing on data from interviews with 20 students, Petocz and Reid (2003) examined the relationships between students' beliefs about learning statistics and their conceptions of teaching statistics. Their results suggest the importance of using instructional strategies that tap into student interest in statistics and, more crucially, encourage the broadest levels of learning. Umugiraneza et al. (2018) used a mixed methods approach to study a related problem, but from the teacher's perspective. The researchers found that teachers perceive their ability to improve students' motivation and interest to be an area of most concern. In addition, Martins et al.(2012) proposed that teachers' attitudes toward statistics may influence student attitudes, although, to our knowledge, this hypothesis remains to be tested in the field.

If instructors matter in impacting various dimensions of students' statistics attitudes as the above studies suggest, then researchers need to try to isolate the effects of instructors on student attitudes. Using data from mathematics teachers in middle schools, Ruzek et al. (2015) found small teacher effects on students' motivation, ranging between 0.03 and 0.08 standard deviation (*SD*s). Blazar and Kraft (2017) found that the effects of upper-elementary teachers on students' happiness in class are sizable (0.31 *SD*) and larger than teacher effects on standardized test scores on mathematics (0.18 *SD*). In addition, the extent to which teachers provide emotional support in mathematics classrooms is positively associated with students' self-efficacy in mathematics as well as their happiness in class. Using a set of data from a randomized experiment, Blazar (2018) provides validity evidence for the teacher effectiveness measure explored in the previous study. Drawing on an independent classroom roster randomization experiment, Kraft (2019) provides added information about the sizes of teacher effects on student attitudes, ranging between 0.10 and 0.16 *SD*s.

2.2. MEASURES OF STUDENTS' ATTITUDES TOWARD STATISTICS

Central to the issue of estimating instructor effects on students' academic attitudes is the development of survey instruments that accurately measure this psychological trait and, therefore, capture instructor-associated changes in the outcome data on student attitudes. There are over a dozen survey instruments that have been created to assess students' statistics attitudes. Nolan et al. (2012) found only four of them frequently appearing in peer-reviewed articles: Survey of Attitudes Toward Statistics (SATS)-28 (Schau et al., 1995), SATS-36 (Schau, 2003), Statistics Attitude Survey (SAS; Roberts & Bilderback, 1980), and Attitude Toward Statistics scale (ATS; Wise, 1985). Of these four surveys, the SATS-28 is considered by researchers to have undergone the most rigorous development process (Nolan et al., 2012), as Schau et al. (1995) employed the nominal group technique and involved both students and experts in the development of the SATS-28.

The SATS-36 builds on the SATS-28 and consists of six subscales – *Interest*, *Value*, *Cognitive Competency*, *Difficulty*, *Affect*, and *Effort*. From the standpoint of psychometrics, both versions of the SATS have been examined far more frequently than SAS and ATS, particularly at the survey item level (Persson et al., 2019; Vanhoof et al., 2011; Xu & Schau, 2019). Whereas Vanhoof et al. (2011) suggest a four-factor model with *Difficulty*, *Affect,* and *Cognitive Competency* combined, the other two articles support a six-factor structure with need of minor modifications. Thus, the development and use of the SATS provide a unique opportunity to examine outcomes of instructional interventions directed at improving multiple dimensions of statistics attitudes (e.g., Carlson & Winquist, 2011; Carnell, 2008; Lesser et al., 2016; Paul & Cunnington, 2017; Posner, 2011).

Although the SATS-36 does not link teaching competencies with students' statistics attitudes directly, it does provide theoretical links, which can be formally framed in theories of motivation. Ramirez, Schau, and Emmioğlu (2012) aligned the SATS-36 with Eccles' expectancy-value theory (EVT) *a posteriori* in their conceptualization of the Model of Students' Attitudes Toward Statistics. EVT provides a theoretical framework that ties many dimensions of teaching behaviors to students' development across behavioral, cognitive, and social-emotional domains (Muenks et al., 2018). Using EVT, Hood, Creed, and Neumann (2012) tested the idea that attitudes and expectancies predict students' statistics achievement empirically. Their findings indicate that statistics instructors are likely to foster educational success through using instructional approaches that improve students' statistics attitudes. van Es and Weaver (2018) examined the potential impact of students' expectations for course achievement on their attitudes toward statistics. The researchers found that students' self-reported expected grades are positively correlated with the *Affect* (r = 0.339) and *Cognitive Competency* (r = 0.324) components of statistics attitudes. Their findings imply that students who expect to do well in the course usually have positive feelings toward this discipline and are confident in their capabilities. We are unaware of any attempt to decipher the relationship between instructor-associated changes in students' statistics attitudes and changes in their expected grades.

2.3. CONCEPTUAL FRAMEWORK

Instructional approaches have always been in the spotlight in statistics education. Over the years, increasing emphasis has been placed on the translation of cognitive and developmental research as well as on perspectives from practicing professional statisticians to statistics-learning environments (e.g., Gal, 2002; Wild & Pfannkuch, 1999). As with teaching mathematics, statistics content-specific views of teaching highlight the importance of developing student conceptual understanding of content and accurately assessing their level of understanding (e.g., Chance, 2002; Garfield, 2002). Because statistics, as opposed to mathematics, works with data (Cobb & Moore, 1997; De Veaux et al., 2017), statistics instructors are also expected to use teaching strategies that focus students on higher-order thinking skills through using data in their classes (e.g., Chance et al., 2007; Cobb, 2015; Garfield et al., 2012). The framework provided in the GAISE report maps out these dimensions of instructional practice, which may roughly be organized into pedagogy that focuses on data, assessment, and concepts.

3. METHODS

3.1. DATA

The data used in this study are from students' and instructors' responses obtained from the SATS project. The complete data set includes a total of 3,775 students and 33 instructors, not all of which could be used in this study. SurveyMonkey, a web-based data collection software program, was used to collect the data across the three academic years from the August 2007 fall term through May 2010. Instructors teaching statistics courses in the United States volunteered to ask their students to take the SATS-36. Students responded to the survey during or outside of class within two weeks of the beginning and of the end of their classes. Instructors were also asked to complete a questionnaire providing information such as rank and self-perception of teaching practices at the end of the term. Each year, the SATS Project was approved by a Human Subjects Institutional Review Board (IRB). See Schau and Emmioğlu (2012) for more information.

3.2. SAMPLE

The sample includes course sections with students who (a) were educated in three major types of U.S. institutions (Baccalaureate Colleges, Master's Colleges and Universities, and Doctoral/Research Universities), (b) took introductory statistics courses with either no mathematics prerequisite or with an algebra-only prerequisite (i.e., "service" courses), (c) responded to 32 or more items on the SATS-36 at both pre and posttest, and (d) completed the questions requesting information on age, gender, and previous achievement in mathematics, all of which are necessary for multilevel analyses with covariate adjustment. Applying these criteria to the data set yielded 1,924 students. Of these, 1,751 students (91%) completed all pretest and posttest items. The rest of the students were missing eight or fewer items on both tests but no more than four items on either test. Survey non-responses were deleted or imputed using hot-deck imputation, a technique commonly used for handling survey non-responses (Andridge & Little, 2010); the results from these two approaches were compared. Importantly, imputation does not cause reduced variability nor obvious bias when applied to the outcome data on statistics attitudes (see Appendix Table A1). As a result, imputed data were retained and used for subsequent analyses. Moreover, the course sections with over 40% student response rate were retained.

Instructors included in this study must have reported Master's or higher degree in statistics or a related quantitative field, and have no missing responses to items of primary interest on the teacher questionnaire. Restrictions on both students and instructors result in a final sample of 1,924 students linked with 23 instructors who taught 90 course sections at 11 post-secondary institutions. These institutions included five Baccalaureate Colleges, four Master's Colleges and Universities, and two Doctoral/Research Universities. The final data set is available upon request from the first author.

3.3. MEASURES

Student-specific measure Both pre- and post-course statistics attitudes were measured by the SATS-36. The pre- and post-versions of the SATS-36 contain identical items except for changes in tense. Students receive a mean score on each subscale. The SATS-36 contains 36 seven-point Likert scale items (1 = Strongly Disagree, 4 = Neutral/No opinion, 7 = Strongly Agree). Each of these items belongs to one of six attitude subscales – *Affect* (6 items), *Cognitive Competence* (6 items), *Value* (9 items), *Difficulty* (7 items), *Interest* (4 items) and *Effort* (4 items). The responses to negatively worded items are reversed before scoring. The students who give higher numerical responses to any item have more positive attitudes than those who give lower responses. Students who have higher composite scores on *Difficulty* perceive statistics to be less difficult.

Self-reported biographical information includes gender, age, and levels of prior achievement in mathematics (prior_math). Students' response to prior_math is on a 1 (very poorly) to 7 (verywell) Likert-scale. Descriptive statistics are presented in Table 1. Both pre and post versions of the SATS-36 also include a question on grade expectations (*exgrade*) in the course. The response is measured by letter grades and then converted to numeric values, with letter grades of A+ corresponding to 4.3 and F to 0. The SATS-36 can be acquired through https://www.evaluationandstatistics.com/.

Table 1. Student and instructor characteristics

	Students	Instructors
Age	19.87 (1.41)	
prior_math	6.00 (1.48)	
Female	1,155 (0.60)	19 (0.93)
Rank		
Adjunct (Part Time)		6 (0.26)
Adjunct (Full Time)		2 (0.09)
Assistant Professor		7 (0.31)
Associate Professor		4 (0.17)
Full Professor		4 (0.17)
Degree		
Master's		7 (0.31)
PhD		16 (0.69)
Sample Size	1,924 (*n*)	23 (*N*)

Note. Age and self-reported prior achievement in mathematics (prior_math) are summarized by median and median absoluate deviation (MAD). For the categorical variables, the value in parentheses indicates percentage.

Cronbach's coefficient alpha values for *Affect*, *Cognitive Competence*, *Value*, *Difficulty*, *Interest*, and *Effort* in the pre-version of the SATS-36 are 0.81, 0.84, 0.87, 0.76, 0.89 and 0.81, respectively. The corresponding alpha values in the post-version are 0.85, 0.86, 0.90, 0.79, 0.91 and 0.77 (Schau & Emmioğlu, 2012). As a result, both versions of the SATS-36 exhibit good to excellent internal consistencies.

Instructor-specific measure The original instructor questionnaire consists of over 20 questions. These questions were created to be consistent with the teaching recommendations found in the earlier version of the GAISE report (C. Schau, personal communication, June 29, 2018). We first removed the questions with binary response options (i.e., Yes/No) because the responses to these questions lack variability as the responses are almost unanimous. Of the 11 remaining items, we further remove the *Communication* item because it does not seem to fit well with the description of either the *Data* or *Assessment* domain (see Table 2 for item descriptions), and combining it with the *Concept* domain increased the Cronbach alpha by only 0.03. Table 2 presents the final 10 items whose selection was based on the framework described in Section 2.3.

Table 2. Descriptions of teacher survey items

Domain	Item name	Question for instructors
Concept	StatLit	Emphasizing statistical literacy
	StatThinking	Emphasizing statistical thinking
	TechConcept	Using technology to develop conceptual understanding
	ActLearn	Fostering active learning
Data	TechData	Using technology for data analysis
	DataContext	Using data in a meaningful context
Assessment	EvalStuLearn	Using assessments to evaluate student learning
	ImprStuLearn	Using assessments to improve student learning
Gen_Attitudes	GenThisCourse	I like teaching introductory statistics course.
Spe_Attitudes	ThisCourse	I liked teaching this specific section of this course.

Note. Questions 1–8 start with "When relevant, how often did you incorporate each of the following elements into your class periods and assignments in this course?". These questions are measured on 5-point Likert-scale (1 = Never or almost never; 3 = About half of the time; 5 = Almost every time). The two questions regarding attitudes are measured on a 7-point Likert-scale (1 = Strongly Disagree, 4 = Neither Disagree nor Agree; 7 = Strongly Agree).

Just as we believe that students' attitudes are important in impacting their course experiences, we also believe that instructors' attitudes impact their teaching effectiveness and so students' experiences. The *GenThisCourse* item purports to measure instructors' attitudes toward teaching introductory statistics courses in general. On the other hand, the *ThisCourse* item purports to measure the instructors' attitudes toward teaching their own specific course sections. Including the *ThisCourse* item in this study is based on the premise that effective teachers are expected to possess a wide range of teaching skills beyond the ability to provide instructional support, to such a degree that teachers continually motivate students to achieve the highest levels of learning (Cheng & Zamarro, 2018; Umugiraneza et al., 2018).

The additional items are classified at face value into three teaching domains. Of the eight items, two contain the key word "data" in the naming and thus purport to measure how well instructors integrate data collection and analysis into their classes; two contain the key word "assessment" and are thus intended to measure how often instructors use assessments in teaching statistics; the other four measure the degree to which instructors deepen students' conceptual understanding of statistics through emphasizing statistical thinking and/or implementing active learning approaches. The respective items are combined to create a final score for each domain by averaging the raw responses across the corresponding set of items. The three instructional domains align with the six recommendations put forward by the GAISE report (p. 6). The *Concept* domain corresponds to GAISE recommendations 1, 2, and 4; the *Data* domain to recommendations 3 and 5; the *Assessment* domain to recommendation 6. Therefore, the measure of instructional practice explored in this study has strong substantive validity.

3.4. STATISTICAL ANALYSIS

We first examine the distributions of the student response data (i.e., post-course attitude scores). Normality appears to be an appropriate approximation for the attitude scores collected from some instructors, but much less so for the data from others. The *Effort* data exhibit extreme skewness and very small instructor-to-instructor variability. To correct for non-normality in the response data, a robust method is implemented throughout this study using *robustlmm* package in R (Koller, 2016). The sample R code is available from the first author upon request.

Multilevel modeling is an empirical strategy for dealing with the complexity of group contexts commonly seen in education research settings. The SATS project data clearly exhibit such organizational hierarchy. Students linked with the same individual instructor share similar classroom experiences and, therefore, may be more likely to develop similar statistics attitudes than those taught by different instructors. Multilevel models allow both within- and between-instructor error structures to be specified to account for the potentially large dependencies among students linked with the same instructors (Thomas & Heck, 2001).

Following conventions (e.g., McCaffrey et al., 2004), we begin by specifying a multilevel model with covariate adjustment to estimate instructor effects on statistics attitudes for student *i* linked with instructor *j*:

$$Y_{ij\_post} = \alpha Y_{ij\_pre} + \pi X_{ij} + \varphi \overline{X}^{c} + \omega_s + \tau_j + \varepsilon_{ij} \tag{1}$$

Here, Y_{ij_post} is the outcome of interest (attitude subscale) measured at the end of the course. The model includes a vector of students' characteristics X_{ij} (i.e., gender, age, self-reported previous mathematics achievement), the same X_{ij} averaged at the course section level $\overline{X}^{c}$, and a school type fixed effect ω_s. Here, previous mathematics achievement is another major determinant of statistics attitudes in addition to instructor and instructional characteristics while gender and age were frequently found to account for a fraction of the variation in statistics attitudes (Ramirez et al., 2012). The purpose of controlling for $\overline{X}^{c}$ and ω_s in the model is to remove peer effects across course sections and any effects due to different school types, respectively. As with any other studies in the teacher effectiveness research, we allow only the intercept to vary across instructors. The error structure is of two levels: the instructor-specific error τ_j and idiosyncratic student-specific error ε_{ij}. Both assume normal density with means of zero.

To address the possibility of bias in estimated instructor effects due to the observational nature of the SATS Project data, we also include a prior measure of the corresponding attitude subscale (Y_{ij_pre}). This strategy intends to mitigate nonrandom sorting. The core assumption underlying this strategy is that students' statistics attitudes measured at the beginning of the courses are sufficient to summarize all the factors that cause the disparities in their statistics attitudes up to that point (Ballou et al., 2004). In practice, those factors are often unmeasurable and/or beyond a teacher's control.

We quantify the size of instructor effects by first obtaining the variance of the instructor-specific random error τ_j (i.e., σ^2_τ) via restricted maximum likelihood (REML), which produces both maximally efficient and consistent σ^2_τ estimator. Instructor effects are defined as the standard deviation (*SD*) of the instructor-level random effect (i.e., $\sqrt{\sigma^2_\tau}$), conditional on pre-attitude scores and an array of covariates. We also compute conditional intraclass correlation coefficient (ICCs) based on Equation 1. Lastly, we extract best linear unbiased prediction (BLUP) estimates of random effects that aim to capture changes associated with individual instructors in each outcome. For Equation 1, Y_{ij_post} and Y_{ij_pre} are transformed into *z*-scores so estimated instructor effects are on the same scale as those obtained in other studies.

To investigate instructional effects on students' statistics attitudes, we specify the following multilevel model:

$$Y_{ij\_post} = \beta T_j + \alpha Y_{ij\_pre} + \pi X_{ij} + \varphi \overline{X}^{c} + \omega_s + \tau_j + \varepsilon_{ij} \tag{2}$$

Beyond and above the same components, Equation 1 is modified to further include a vector of the *j*th instructor's self-reported scores on five instructional domains (T_j). An instructor may have taught more than one course section. These scores are averaged across sections taught by this specific instructor and thus serve as self-reported instructional variables. Consequently, the vector of regression coefficients β contains the parameters of primary interest. For Equation 2, Y_{ij_post} and Y_{ij_pre} are kept on the original scale to make the results more interpretable.

4. RESULTS

4.1. INSTRUCTOR SELF-REPORTED MEASURES

The two attitude items (*GenThisCourse* and *ThisCourse*) are kept separate from the other eight items as well as from each other. By definition, the two items were written to measure different yet interrelated dimensions of teaching (see Table 2). This claim is supported by a moderate correlation found between instructors' responses to these two items ($r = 0.42$, $p < 0.05$; see Table 3). Table 3 also presents summary statistics for each of the instructor self-reported measurements.

Table 3. Descriptive statistics for the instructor self-reported measure

	Univariate statistics		Pairwise Spearman's rank correlations				
Domain	Median	MAD	1	2	3	4	5
1. *Gen_Attitudes*	6.64	0.34	1.00				
2. *Spe_Attitudes*	6.00	0.72	0.42*	1.00			
3. *Concept*	4.25	0.62	0.41*	0.28	1.00		
4. *Data*	4.74	0.38	0.47*	0.21	0.56**	1.00	
5. *Assessment*	3.71	1.05	-0.11	-0.19	0.21	0.15	1.00

Note. ***p* < 0.01, **p* < 0.05. MAD stands for median absolute deviation.

The pairwise Spearman's rank correlation between instructors' responses to the *Concept* and *Data* domains is relatively large ($r = 0.56$, $p < 0.01$). This result suggests that *Concept* and *Data* may represent two different but interrelated dimensions of instructional practice. The degree to which these two domains are correlated can be viewed as a reflection of the interactions among instructional domains in some of the six GAISE recommendations. For example, one recommendation—"Use technology to explore concepts and analyze data"—emphasizes the importance of statistics instruction proximal to data and concepts simultaneously. The rest of the pairwise correlations are weak to moderate (see Table 3). As a result, it is reasonable to posit that the instructor self-reported measure has captured five dimensions of teaching practice. Also, the magnitudes of these correlations imply that multicollinearity is not a potential problem in this study.

In addition, we evaluate the reliability for instructor-level measures using Cronbach's coefficient alpha. The alpha values for the *Data*, *Assessment*, and *Concept* domains are 0.735, 0.703, and 0.885, respectively. As a guideline, an alpha of 0.7 or above is acceptable for a measurement scale (Lance et al., 2006). The reliability of single items (i.e., *Gen_Attitudes* and *Spe_Attitudes*) cannot be calculated using Cronbach's alpha. We also fit a three-factor confirmatory analysis model to the eight items depicting the *Data*, *Assessment*, and *Concept* domains. The model fails to converge to a solution. This non-convergence may be caused by the small ratio of (instructor-level) sample size to the number of model parameters (Myers et al., 2011). As a result, we are not able to establish the factorial validity for the instructor-level measures explored in this study.

4.2. INSTRUCTOR EFFECTS ON STUDENTS' STATISTICS ATTITUDES

We check whether nonrandom sorting of students to different types of institutions has detectable associations with students' prior statistics attitudes. Figure 2 presents a visual examination of average prior attitudes scores for each instructor by institution type. The plots show that average prior statistics attitudes are not systematically high or low for instructors at any of the three institution types. Specifically, the instructors at Doctoral/Research Universities did not always have students with the highest (or lowest) pre-test statistics attitudes in their classes. Nor did the instructors at the other two types of institutions. The results indicate that any bias due to nonrandom sorting based on pre-attitude scores may be limited for this study. Thus, we pool data collected from different institution types.

We next examine instructor effects on students' post statistics attitudes and corroborate the presence of instructor effects with two closely related metrics: instructor-to-instructor sample variability and conditional ICC. Here, conditional ICC is defined as the ratio of the teacher-level variation to the total variation comprising the teacher- and student-level variation, after accounting for pre-attitude scores and an array of covariates. The results reported in Table 4 suggest that there is a non-negligible amount of instructor-level sample variability in average post-attitude scores, with standard deviations ranging between 0.24 (*Effort)* and 0.63 (*Interest*). Notably, the instructor-to-instructor variability in average attitudes scores is substantially greater at the posttest than at the pretest.

Conditional ICCs from Equation (1) follow a similar pattern. All but one of these values are greater than the conventionally acceptable cutoff 0.05 (Jak et al., 2014), thereby justifying the use of multilevel models. On the other hand, the ICC value for *Effort* is very low, slightly under 0.02. A low ICC as such indicates that there are very small differences across instructors with regard to student-reported *Effort*.

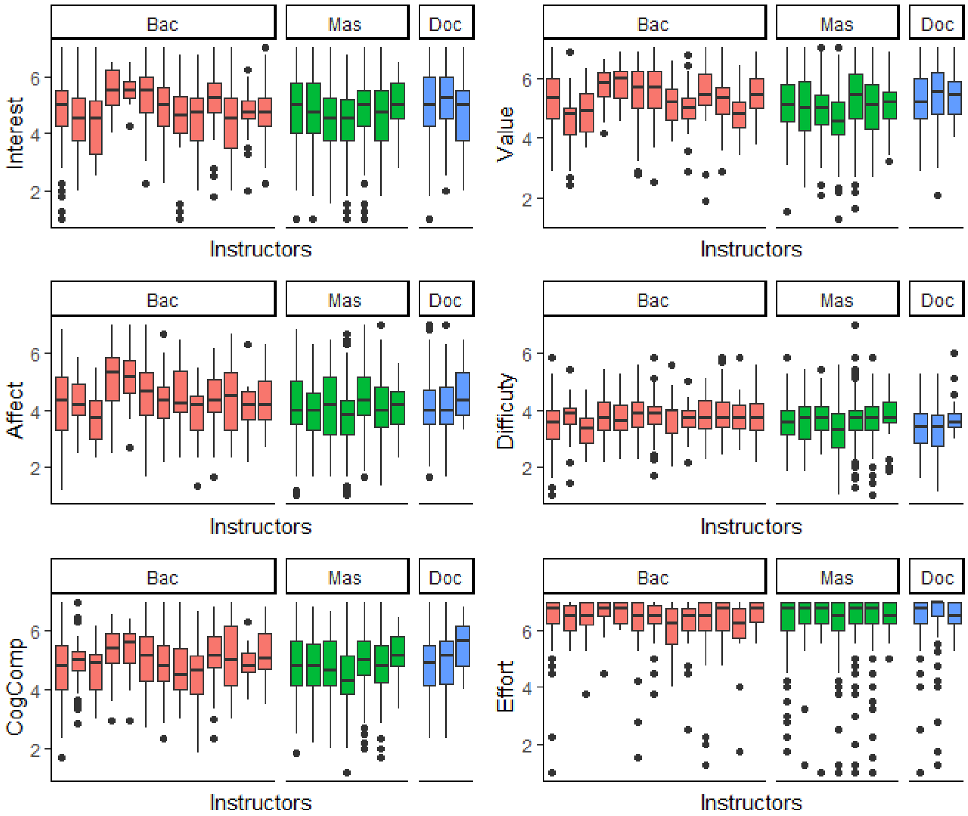

Figure 2. Boxplots of pre-attitude subscale scores for each instructor by school type

Table 4. Sample variability, ICCs, and instructor effects on student outcomes

Outcome	Instructor level scores		Instructor effect	Conditional ICC
	Pre Mean (*SD*)	Post Mean (*SD*)		
Interest	4.81 (0.34)	4.37 (0.63)	0.30	0.14
Affect	4.30 (0.35)	4.32 (0.61)	0.36	0.16
CogComp	4.96 (0.25)	5.05 (0.39)	0.19	0.06
Difficulty	3.65 (0.19)	3.76 (0.25)	0.18	0.06
Value	5.21 (0.32)	4.91 (0.54)	0.34	0.20
Effort	6.38 (0.19)	5.91 (0.24)	0.09	0.02
exgrade	3.35 (0.08)	2.93 (0.28)	0.28	0.12

Technically, the small differences might not be productively modeled as a function of self-reported instructional variables. As a result, the *Effort* subscale is excluded from the subsequent analyses.

Instructor effects on standardized post-attitudes for *Interest* (0.30 *SD*), *Affect* (0.36 *SD*), and *Value* (0.34 *SD*) are comparable in magnitude to teacher effects on upper-elementary school students' Happiness in Class (0.31 *SD*), the highest one reported in Blazar and Kraft (2017). Specifically, our findings suggest that 1 *SD* difference in instructor effectiveness is equivalent to about 1/3 of a *SD* difference in students' *Interest*, *Affect*, and *Value* scores, and about 1/5 of a *SD* difference in students' *Difficulty* and *CogComp* scores. Given the sizable instructor effects observed on statistics attitudes, the next step is thus to seek possible explanations as to what instructional dimensions account for the effects.

4.3. INSTRUCTIONAL EFFECTS ON STUDENTS' STATISTICS ATTITUDES

We first examine whether any of the five dimensions of self-reported instructional practice helps account for the considerable amount of variability in post-course statistics attitudes across instructors using Equation (2). We present the unstandardized estimates of regression coefficients in Table 5, which shows four statistically significant relationships among certain instructional dimensions and statistics attitude components. Instructors' attitudes toward teaching their own course sections (i.e., *Spe_Attitudes*) are positively related to students' interest in statistics (*Interest*) and their feeling about statistics (*Affect*) but not with other attitude components. Specifically, a one-unit increase in instructors' attitudes toward teaching their classes is associated with a 0.32 or 0.28 unit increase in *Interest* or *Affect* score among students in their classes, holding all other independent variables constant. As a point of comparison, instructors' attitudes toward introductory statistics courses (i.e., *Gen_Attitudes*) are not significantly associated with any of the five attitude components.

In addition, the extent to which instructors emphasize data collection and analysis is positively associated with *Value* and *Affect*. Specifically, a one-unit increase in instructors' emphasis on data is associated with a 0.32 or 0.40 unit increase in *Value* or *Affect* score among students in their classes, holding all other independent variables constant. This finding provides empirical support for a data-intensive introduction to statistics courses for non-majors.

Table 5. Instructional effects on students' statistics attitudes (coefficient and standard error)

	Interest	*Difficulty*	*Value*	*CogComp*	*Affect*
Gen_Attitudes	-0.25	-0.10	-0.17	0.01	-0.16
	(0.16)	(0.08)	(0.16)	(0.09)	(0.18)
Spe_Attitudes	0.32*	0.06	0.23	0.08	0.28*
	(0.11)	(0.06)	(0.11)	(0.07)	(0.13)
Concept	-0.07	0.02	-0.14	0.04	-0.15
	(0.14)	(0.08)	(0.15)	(0.09)	(0.17)
Data	0.27	0.14	0.32*	0.15	0.40*
	(0.14)	(0.07)	(0.15)	(0.08)	(0.17)
Assessment	0.16	0.07	0.11	0.04	0.26
	(0.13)	(0.07)	(0.14)	(0.08)	(0.16)

Note. $^*p < 0.05$. The regression coefficient estimates are unstandardized.

We also examine Cohen's f^2, which summarizes the effect size in the multilevel setting, using the formula described in Selya et al. (2012). Cohen's f^2 for the effects of the *Data* domain on *Affect* and *Value* are both approximately 0.02. Accordingly, Cohen's f^2 for the effects of the *Spe_Attitudes* domain on *Affect* and *Interest* are approximately 0.03 and 0.02, respectively. These values correspond to a small effect size (i.e., $f^2 = 0.02$), according to Cohen's conventions. Nonetheless, our findings largely contribute to the predictive validity of the instructor self-reported measure, as instructional practices are anticipated to have the most impact on *Interest*, *Affect*, and *Value* (Kerby & Wroughton, 2017; Schau, 2003).

Given the dearth of multilevel studies on student learning in the literature of statistics education, we also report the regression coefficient estimates on student characteristic variables in the presence of self-reported instructional variables (see Appendix Table A2). Results include slightly higher scores on the *Difficulty*, *Cognitive Competency*, *Affect*, and *Interest* subscales by male students, and better attitudes for students who have higher scores on self-reported previous achievement in mathematics. In addition, student age does not predict statistics attitude scores.

4.4. CORRELATIONS AMONG INSTRUCTOR-ASSOCIATED CHANGES IN STUDENT ATTITUDES AND CHANGES IN EXPECTED GRADES

We extract BLUP estimates of random effects to capture changes in statistics attitudes and expected grades that are associated with individual instructors. Correlating BLUP estimates gives rise to the extent to which instructor-associated changes in students' statistics attitudes are associated with changes in students' expected grades. The largest correlation is found between expected grade and *Difficulty* ($r = 0.81$, $p < 0.001$). The interpretation suggests that the variation in instructor-associated changes in students' *Difficulty* score can account for 66% of the variation in instructor-associated changes in students' expected grades (i.e., r^2). The estimates are 0.79, 0.76, 0.69, and 0.64 ($p < 0.001$) for the correlations between instructor-associated changes in expected grade and changes in *Affect*, *CogComp*, *Value*, and *Interest*, respectively. We illustrate these relationships in Figure 3. Solid lines reflect the magnitude of Pearson correlation coefficients. Although a positive relationship clearly exists between instructor-associated changes in *Value* (or *Interest*) and changes in expected grades, the two variables do not seem to increase at a constant rate; that is, the underlying relationship may be non-linear.

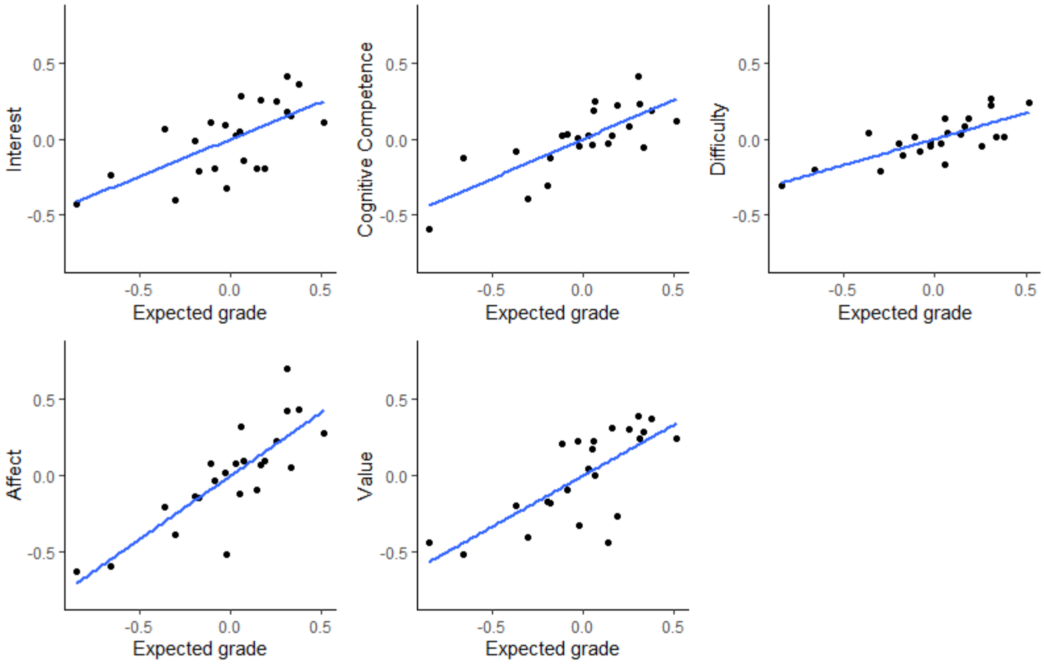

Figure 3. Scatterplots of instructor-associated changes in students' statistics attitudes and changes in expected grades

Examining the results for the correlations reveals that positive instructor-associated gains in students' statistics attitudes are associated with positive gains in students' expected grades. To sound a note of caution, however, the precision in BLUP estimates decreases as the proportion of instructor-level variation becomes smaller. Thus, the correlation derived from BLUP is an imperfect measure of the true relationships between instructor-associated changes in students' statistics attitudes and changes in expected grades. The correlations reported here should be viewed as suggestive rather than conclusive.

5. DISCUSSION

5.1. INSTRUCTOR EFFECTS ON STUDENTS' STATISTICS ATTITUDES

Much of what goes on in statistics education takes place within group contexts. The hierarchical structure of the SATS Project data prompts use of multilevel models to examine instructor effects on students' statistics attitudes. Given the observational nature of the SATS Project data, our multilevel modeling approach relies exclusively on covariate adjustment to reduce the bias assumed to arise from nonrandom sorting. We incorporate a set of student-specific covariates, averaged covariates per course section, school fixed effects, and pre-instruction attitude scores into our models in an effort to account for potential sources of bias, an empirical strategy described in Blazar and Kraft (2017). More critically, recent quasi-experimental and randomized analyses provide strong empirical support to the unbiasedness of teacher effect estimates even in the presence of systematic sorting (Blazar, 2018; Chetty et al., 2014; Kane et al., 2013; Kraft, 2019). Findings from these studies support the validity of our analysis approaches.

Previous studies have implied the presence of instructor effects on certain components of students' statistics attitudes (e.g., Petocz & Reid, 2003; Schau, 2003; Umugiraneza et al., 2018; Waters et al., 1988). Even so, it is hard to determine to what degree empirical evidence supports their claims. We provide supporting quantitative evidence that there are large instructor effects on students' statistics attitudes. The sizes of instructor effects on the five statistics attitude components (*Effort* excluded) range between 0.18 and 0.36 *SD*s; these values generally align with the sizes of teacher effect estimates in Blazar and Kraft (2017), where the authors observe sizable impacts of the upper-elementary school teachers on students' academic attitudes and behaviors (0.14 to 0.31 *SD*s). The interpretation of instructor effects suggests that statistics attitudes for a student assigned to an average instructor might have been between 0.18 and 0.36 *SD*s higher had he/she been assigned to an instructor one *SD* higher on the instructor effectiveness distribution.

5.2. INSTRUCTIONAL EFFECTS ON STUDENTS' STATISTICS ATTITUDES

Researchers have begun to understand how teaching quality contributes differentially to student learning in general (e.g., Blazar & Kraft, 2017) and to the learning of statistics in particular (Chance et al., 2016). We do not find associations between students' statistics attitudes and the *Concept* or *Assessment* domain of instructional practice. We do find evidence that the *Data* domain predicts *Value* and *Affect*. This finding is largely in line with several qualitative observations, which suggest using data-intensive approaches to foster student learning of statistics via links to students' appreciation of the utilitarian value of statistics in their personal and professional lives (Neumann et al., 2013; Songsore & White, 2018). The instructor-associated changes in students' statistics attitudes may thus be viewed as an outcome of group processes directed by instructors who frequently have students engaged in common tasks involving data. Carnell (2008) suggests that instructors may need to do this if their goal is to positively impact students' statistics attitudes, as the author found limiting students to one data collection project was not sufficient to achieve this goal.

We caution against placing too much emphasis on the causal interpretation of our findings on instructional effects. The current multilevel covariate adjustment models can measure instructor-level causal quantities, but only if some extreme assumptions are met (Rubin et al., 2004). One important case in this study is that the participating instructors chose their own instructional practices. Nonrandom assignment of instructional practices may lead to the dependency of random effects on self-reported instructional variables. Dependency as such could be a source of bias in the instructional effect estimates. In practice, it is difficult to judge to what degree the empirical evidence reflects such dependency. As a result, we are not able to assert that the frequent use of data or instructors' positive attitudes toward teaching their classes is the true cause that underlies improvements in students' statistics attitudes.

Nonetheless, the positive associations found in this study have some implications for statistics instruction. For example, counter-intuitive phenomena are common when it comes to statistics as this discipline takes root in probability. Students may develop frustration over learning statistics when they observe those phenomena. When this occurs, statistics instructors are expected not only to deliver concepts with great precision but also to provide motivational support to students who struggle to learn statistics. In this regard, our work adds to the statistics education literature as we have identified associations of instructors' attitudes toward teaching their classes with the *Affect* and *Interest* components of students' statistics attitudes.

This finding makes sense, as instructors who have higher attitudes toward teaching are more likely to hold students' perspective in regard, show responsiveness to student problems, and create a positive classroom climate. A positive classroom climate not only establishes a positive teacher-student relationship that fosters growth in self-efficacy by creating a safe learning environment free of ridicule and criticism, it also creates opportunities for academic challenge and student success in the classroom (Peters, 2013). In this environment, students perceive their instructors as caring and supportive. Those at the high end of the instructor effectiveness distribution, for example, may be more willing to take the time and break a recalcitrant problem into smaller pieces, thereby fostering students' interest and creating positive feelings toward statistics.

Lastly, our analysis provides empirical support for the use of multilevel models with the need to control for student background, as advised by Ballou et al. (2004). Our findings on the relationships between student gender and statistics attitudes are consistent with those reported in early studies conducted in other countries than the United States (Ramirez et al., 2012) and one recent study conducted in the United States (van Es & Weaver, 2018). Note that these studies do not investigate the *Interest* component of student attitudes because they use the SATS-28. Previous studies also suggest that age may relate to student attitudes (and anxiety) in complicated ways (e.g., Coetzee & van der Merwe, 2010; Onwuegbuzie, 2004). The present study does not find significant linear relationships among student age and attitude component scores. Together, it is suggested that the effect of age on statistics attitudes may be idiosyncratic, depending on type of courses, locations, countries, and many other factors. As anticipated from the literature (e.g., Ramirez et al., 2012; Tishkovskaya & Lancaster, 2012), students with more prior experience in mathematics tend to have higher statistics attitudes at the end of the courses.

5.3. CORRELATIONS AMONG INSTRUCTOR-ASSOCIATED CHANGES IN STUDENT ATTITUDES AND CHANGES IN EXPECTED GRADE

Expected grade is a performance measure used in almost all student evaluations of teachers. Average expected grade by a given class is viewed as one determinant of an instructor's evaluation score. For example, McPherson (2006) reported that, with an increase of 1 point in average expected grade, an evaluation score increases by 0.34 point. In this study, we find that positive instructor-associated gains in students' statistics attitudes are associated with positive gains in students' expected grades. This finding implies that higher student evaluation scores may be achieved when instructors do a better job of teaching through improving certain dimensions of students' statistics attitudes. The instructor effectiveness measures explored in this study may be used to orient statistics instructors on the development of new pedagogical skills centered on students' statistics attitudes. This, however, raises the concern that instructors may be able to "buy" positive statistics attitudes or academic attitudes in general (and thus good evaluations) through a policy of easy grading or grade inflation. Additional studies are needed to distinguish between instructor-associated improvements in statistics attitudes due to quality teaching and improvements due to an inclination for easy grading.

5.4. FUTURE DIRECTIONS

Future research may focus on a comprehensive analysis of teaching practice specific to statistics classrooms. This analysis can then be mapped onto different domains of a theoretical framework that in turn helps inform design of instructor-level measures. There are several candidate frameworks. One teaching framework that has been popularized in American K-12 education research but may also be instrumental in higher education is the Classroom Assessment Scoring System (CLASS, Pianta & Hamre, 2009). The CLASS framework consists of three domains: Instructional Supports, Emotional Supports, and Classroom Organization.

Whereas the teacher-report measures we explored in this study elicit important information about teachers' self-reported perception of instructional practice in statistics classrooms, data collected from observational measures can provide added information. Thus, one branch of future research could focus on observations of statistics classrooms as a means of capturing the quality of teachers' instruction. Also, statistics content has become an important component of pre-college school curricula (Usiskin & Hall, 2015). This trend provides a unique opportunity for statistics education researchers to engage in large-scale education science which often requires development of standardized observational assessment of teaching practice with links to various student outcomes.

In addition, we used a single item as a proxy for teachers' emotional support skills as well as classroom management skills. Future studies will benefit from operationalization of these teaching dimensions in relation to statistics classrooms. With reliable and valid instruments, researchers will be able to model the changes in scores on various dimensions of teaching competencies as a function of specific training and supports provided to statistics instructors. Such validated measures are required to determine whether teacher training is effective; namely, whether exposure to such support is predictive of student gains in statistics attitudes and other learning outcomes.

5.5. LIMITATIONS

We recognize that our analyses and results are limited by the SATS project data available to us, which were collected about ten years prior to this analysis. Drawing on a recent data set, we are conducting a study to determine the replicability of the findings regarding instructor effects on student attitudes. The interpretation of our results is further limited by the sample of instructors and students in post-secondary institutions being exclusively from the United States. We would expect teacher effects on student attitudes in other countries and educational systems to be present, as large differences in teacher and teaching quality exist regardless of countries and systems. For example, Umugiraneza et al. (2018) found that schoolteachers in South Africa differ in their willingness to use instructional strategies that enhance students' motivation to learn mathematics and statistics. The differences may be accounted for by a range of factors including age, educational attainment, and teacher training provided to the particular sample of teachers. Therefore, it is not possible to extrapolate the size of teacher effects on students' statistics attitudes to other countries or systems outside the United States.

Second, our results and conclusions only apply to those aspects of the GAISE report that are measured in this study. Other instructor-level measures based on GAISE or different conceptual frameworks might yield additional findings. Furthermore, there are other instruction- and instructor-related variables that are important to students' statistics attitudes but are not collected as a part of the SATS project data. For example, Chance et al. (2016) report that levels of teachers' experience with simulation-based statistics curricula are positively related to students' conceptual understanding of statistics. It is thus of great interest to examine to what extent instructor preparation for teaching predicts gains in students' statistics attitudes in future studies.

Third, the results reported in this study were produced without consideration of student attrition. This limitation highlights an inherent complication in the analysis of quantitative data on student learning outcomes in higher education research; that is, a large number of college students drop the course or simply choose not to take course surveys. It is unclear to what extent the censoring of the post-attitude data impacts our results. To address this issue, future studies need to consider information on course dropouts in the analysis of students' statistics attitudes as a learning outcome. To our knowledge, large-scale data sets with this information are being collected (Chance et al., 2018).

Fourth, an important caveat in the design of this study is that, without data from classroom observations, the degree of consistency between instructor self-reported practices and their actual teaching performances is unclear. By calibrating our results with previous qualitative findings reported in Neumann et al. (2013) and Songsore and White (2018), however, both of which also find some associations between the use of genuine data and certain components of statistics attitudes, we suggest that threats to the validity of the instructor self-reported measure may be limited.

A further limitation is that the SATS Project dataset does not contain student race/ethnicity information and therefore may not accurately reflect the rapidly changing student demography in the United States. Prior studies have shown, however, that there are no differences in statistics attitudes between racial/ethnical groups (Ramirez et al., 2012). Thus, the generalizability of our findings may not be in jeopardy if only information on racial composition is unavailable. One additional concern about the current study is the issue of small instructor-level sample size (N = 23). Future studies designed to replicate our results and/or identify effects of new instructional practices on statistics attitudes should strive to include more instructors.

6. CONCLUSION

Recent research on the multidimensional nature of teaching practice and its implications for education policy creates strong incentives for educators to improve students' academic attitudes (Blazar & Kraft, 2017; Kraft, 2019). Using multilevel modeling approaches with covariate adjustment, we showed that there is large variation in instructor-associated changes in students' statistics attitudes at the end of introductory statistics courses. The Mathematics and Statistics departments in many post-secondary institutions are well-positioned to utilize this evidence because these departments often have latitude in hiring part-time instructors. Staffing flexibility, in terms of hiring faculty and opportunities for professional development, can be a source of differences in teacher effectiveness. If teaching competencies vary considerably across instructors who teach entry-level introductory statistics courses, then reforming teacher preparation and/or training policies may lead to improved student outcomes.

Moreover, our study places emphasis on linking instructor self-reported measures of teaching practice to student-level assessment data. The analysis shows that the large variation in instructor-associated changes in student attitudes can be partly accounted for by how much instructors like teaching their classes as well as by the extent to which instructors use data to engage students in learning statistics. Thus, our analysis provides support to some recent qualitative findings (Neumann et al., 2013; Songsore & White, 2018; Umugiraneza et al., 2018). Lastly, our results suggest that measures of instructor effectiveness in relation to student attitudes may serve as a proxy for student evaluation of teaching performance. The present study emphasizes the centrality of teacher and teaching quality in statistics classrooms and helps inform the necessity of reforming policies on teacher training and professional development oriented toward enhancing pedagogical skills that have the potential to improve students' statistics attitudes.

ACKNOWLEDGEMENTS

We would like to thank Dr. Candace Schau for providing the SATS project data as well as her invaluable insight. We are also grateful to the two anonymous reviewers, associate editor, assistant editor and editor for their constructive criticisms and suggestions. All errors and omissions belong to the first author.

REFERENCES

Andridge, R. R., & Little, R. J. (2010). A review of hot deck imputation for survey non-response. *International Statistical Review, 78*(1), 40–64.
[Online: https://doi.org/10.1111/j.1751-5823.2010.00103.x]

Ballou, D., Sanders, W., & Wright, P. (2004). Controlling for student background in value-added assessment of teachers. *Journal of Educational and Behavioral Statistics, 29*(1), 37–65.
[Online: https://doi.org/10.3102/10769986029001037]

Blazar, D. (2018). Validating teacher effects on students' attitudes and behaviors: Evidence from random assignment of teachers to students. *Education Finance and Policy, 13*(3), 281–309.
[Online: https://doi.org/10.1162/edfp_a_00251]

Blazar, D., & Kraft, M. A. (2017). Teacher and teaching effects on students' attitudes and behaviors. *Educational Evaluation and Policy Analysis, 39*(1), 146–170.
[Online: https://doi.org/10.3102/0162373716670260]

Carlson, K. A., & Winquist, J. R. (2011). Evaluating an active learning approach to teaching introductory statistics: A classroom workbook approach. *Journal of Statistics Education, 19*(1).
[Online: https://doi.org/10.1080/10691898.2011.11889596]

Carnell, L. J. (2008). The effect of a student-designed data collection project on attitudes toward statistics. *Journal of Statistics Education, 16*(1).
[Online: https://doi.org/10.1080/10691898.2008.11889551]

Chance, B. (2002). Components of statistical thinking and implications for instruction and assessment. *Journal of Statistics Education, 10*(3).
[Online: https://doi.org/10.1080/10691898.2002.11910677]

Chance, B., Ben-Zvi, D., Garfield, J., & Medina, E. (2007). The role of technology in improving student learning of statistics. *Technology Innovations in Statistics Education, 1*(1).
[Online: https://escholarship.org/content/qt8sd2t4rr/qt8sd2t4rr.pdf]

Chance, B., Mendoza, S., & Tintle, N. (2018). Student gains in conceptual understanding in introductory statistics with and without a curriculum focused on simulation-based inference, *Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS-10, July, 2018*), Kyoto, Japan. International Statistical Institute.
[Online: http://iase-web.org/icots/10/proceedings/pdfs/ICOTS10_3B2.pdf?1531364254]

Chance, B., Wong, J., & Tintle, N. (2016). Student performance in curricula centered on simulation-based inference: A preliminary report. *Journal of Statistics Education, 24*(3), 114–126.
[Online: https://doi.org/10.1080/10691898.2016.1223529]

Cheng, A., & Zamarro, G. (2018). Measuring teacher non-cognitive skills and its impact on students: Insight from the measures of effective teaching longitudinal database. *Economics of Education Review, 64*, 251–260.
[Online: https://doi.org/10.1016/j.econedurev.2018.03.001]

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., & Yagan, D. (2011). How does your kindergarten classroom affect your earnings? Evidence from Project STAR. *The Quarterly Journal of Economics*, *126*(4), 1593–1660.
[Online: https://doi.org/10.1093/qje/qjr041]

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review, 104*(9), 2593–2632.
[Online: https://doi.org/10.1257/aer.104.9.2593]

Cobb, G. (2015). Mere renovation is too little too late: We need to rethink our undergraduate curriculum from the ground up. *The American Statistician, 69*(4), 266–282.
[Online: https://doi.org/10.1080/00031305.2015.1093029]

Cobb, G. W., & Moore, D. S. (1997). Mathematics, statistics, and teaching. *The American Mathematical Monthly, 104*(9), 801–823.
[Online: https://doi.org/10.1080/00029890.1997.11990723]

Coetzee, S., & van der Merwe, P. (2010). Industrial psychology students' attitudes towards statistics. *SA Journal of Industrial Psychology, 36*(1), 1–8.
[Online: https://doi.org/10.4102/sajip.v36i1.843]

De Veaux, R. D., Agarwal, M., Averett, M., Baumer, B. S., Bray, A., Bressoud, T. C., ... & Kim, A. Y. (2017). Curriculum guidelines for undergraduate programs in data science. *Annual Review of Statistics and its Application, 4*, 15–30.
[Online: https://doi.org/10.1146/annurev-statistics-060116-053930]

Duckworth, A. L., Quinn, P. D., & Tsukayama, E. (2012). What No Child Left Behind leaves behind: The roles of IQ and self-control in predicting standardized achievement test scores and report card grades. *Journal of Educational Psychology, 104*(2), 439–451.
[Online: https://doi.org/10.1037/a0026280]

GAISE College Report ASA Revision Committee (2016). *The Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report*. American Statistical Association.
[Online: https://www.amstat.org/asa/files/pdfs/GAISE/GaiseCollege_Full.pdf]

Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. *International Statistical Review*, *70*(1), 1–25.
[Online: https://doi.org/10.1111/j.1751-5823.2002.tb00336.x]

Garfield, J. (2002). The challenge of developing statistical reasoning. *Journal of Statistics Education, 10*(3).
[Online: https://doi.org/10.1080/10691898.2002.11910676]

Garfield, J., delMas, R., & Zieffler, A. (2012). Developing statistical modelers and thinkers in an introductory, tertiary-level statistics course. *ZDM Mathematics Education, 44*(7), 883–898.
[Online: https://doi.org/10.1007/s11858-012-0447-5]

Hood, M., Creed, P. A., & Neumann, D. L. (2012). Using the expectancy value model of motivation to understand the relationship between student attitudes and achievement in statistics. *Statistics Education Research Journal, 11*(2), 72–85.

[Online: https://iase-web.org/documents/SERJ/SERJ11(2)_Hood.pdf]
Jak, S., Oort, F. J., & Dolan, C. V. (2014). Measurement bias in multilevel data. *Structural Equation Modeling, 21*(1), 31–39.
[Online: https://doi.org/10.1080/10705511.2014.856694]
Kane, T. J., McCaffrey, D. F., Miller, T., & Staiger, D. O. (2013). *Have we identified effective teachers? Validating measures of effective teaching using random assignment*. Bill & Melinda Gates Foundation.
[Online: https://k12education.gatesfoundation.org/download/?Num=2676& filename=MET_Validating_Using_Random_Assignment_Research_Paper.pdf]
Kerby, A. T., & Wroughton, J. R. (2017). When do students' attitudes change? Investigating student attitudes at midterm. *Statistics Education Research Journal, 16*(2), 476–486.
[Online: https://iase-web.org/documents/SERJ/SERJ16(2)_Kerby.pdf]
Koller, M. (2016). robustlmm: An R package for robust estimation of linear mixed-effects models. *Journal of Statistical Software, 75*(6), 1–24.
[Online: https://doi.org/10.18637/jss.v075.i06]
Kraft, M. A. (2019). Teacher effects on complex cognitive skills and social-emotional competencies. *Journal of Human Resources, 54*(1), 1–36.
[Online: https://doi.org/10.3368/jhr.54.1.0916.8265R3]
Lance, C. E., Butts, M. M., & Michels, L. C. (2006). The sources of four commonly reported cutoff criteria: What did they really say?. *Organizational Research Methods, 9*(2), 202–220.
[Online: https://doi.org/10.1177/1094428105284919]
Lesser, L. M., Pearl, D. K., & Weber, J. J. (2016). Assessing fun items' effectiveness in increasing learning of college introductory statistics students: Results of a randomized experiment. *Journal of Statistics Education, 24*(2), 54–62.
[Online: https://doi.org/10.1080/10691898.2016.1190190]
Martins, J. A., Nascimento, M. M., & Estrada, A. (2012). Looking back over their shoulders: A qualitative analysis of Portuguese teachers' attitudes towards statistics. *Statistics Education Research Journal, 11*(2), 26–44.
[Online: https://iase-web.org/documents/SERJ/SERJ11(2)_Martins.pdf]
McCaffrey, D. F., Lockwood, J., Koretz, D., Louis, T. A., & Hamilton, L. (2004). Models for value-added modeling of teacher effects. *Journal of Educational and Behavioral Statistics, 29*(1), 67–101.
[Online: https://doi.org/10.3102/10769986029001067]
McPherson, M. A. (2006). Determinants of how students evaluate teachers. *The Journal of Economic Education, 37*(1), 3–20.
[Online: https://doi.org/10.3200/JECE.37.1.3-20]
Muenks, K., Wigfield, A., & Eccles, J. S. (2018). I can do this! The development and calibration of children's expectations for success and competence beliefs. *Developmental Review, 48*, 24–39.
[Online: https://doi.org/10.1016/j.dr.2018.04.001]
Myers, N. D., Ahn, S., & Jin, Y. (2011). Sample size and power estimates for a confirmatory factor analytic model in exercise and sport: A Monte Carlo approach. *Research Quarterly for Exercise and Sport, 82*(3), 412–423.
[Online: https://doi.org/10.1080/02701367.2011.10599773]
Neumann, D. L., Hood, M., & Neumann, M. M. (2013). Using real-life data when teaching statistics: Student perceptions of this strategy in an introductory statistics course. *Statistics Education Research Journal, 12*(2), 59–70.
[Online: https://iase-web.org/documents/SERJ/SERJ12(2)_Neumann.pdf]
Nolan, M. M., Beran, T., & Hecker, K. G. (2012). Surveys assessing students' attitudes toward statistics: A systematic review of validity and reliability. *Statistics Education Research Journal, 11*(2), 103–123.
[Online: https://iase-web.org/documents/SERJ/SERJ11(2)_Nolan.pdf]
Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis, 26*(3), 237–257.
[Online: https://doi.org/10.3102/01623737026003237]

Onwuegbuzie, A. J. (2004). Academic procrastination and statistics anxiety. *Assessment & Evaluation in Higher Education, 29*(1), 3–19.
[Online: https://doi.org/10.1080/0260293042000160384]

Paul, W., & Cunnington, R. C. (2017). An exploration of student attitudes and satisfaction in a GAISE-influenced introductory statistics course. *Statistics Education Research Journal, 16*(2), 487–510.
[Online: https://iase-web.org/documents/SERJ/SERJ16(2)_Paul.pdf]

Persson, I., Kraus, K., Hansson, L., & Wallentin, F. (2019). Confirming the structure of the Survey of Attitudes Toward Statistics (SATS-36) by Swedish students. *Statistics Education Research Journal*, *18*(1), 83–93.
[Online: https://iase-web.org/documents/SERJ/SERJ18(1)_Persson.pdf]

Peters, M. L. (2013). Examining the relationships among classroom climate, self-efficacy, and achievement in undergraduate mathematics: A multi-level analysis. *International Journal of Science and Mathematics Education*, *11*(2), 459–480.
[Online: https://doi.org/10.1007/s10763-012-9347-y]

Petocz, P., & Reid, A. (2003). Relationships between students' experience of learning statistics and teaching statistics. *Statistics Education Research Journal, 2*(1), 39–53.
[Online: https://iase-web.org/documents/SERJ/SERJ2(1).pdf]

Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher, 38*(2), 109–119.
[Online: https://doi.org/10.3102/0013189X09332374]

Posner, M. A. (2011). The impact of a proficiency-based assessment and reassessment of learning outcomes system on student achievement and attitudes. *Statistics Education Research Journal, 10*(1), 3–15.
[Online: https://iase-web.org/documents/SERJ/SERJ10(1)_Posner.pdf]

Ramirez, C., Schau, C., & Emmioglu, E. (2012). The importance of attitudes in statistics education. *Statistics Education Research Journal, 11*(2), 57–71.
[Online: https://iase-web.org/documents/SERJ/SERJ11(2)_Ramirez.pdf]

Roberts, D. M., & Bilderback, E. W. (1980). Reliability and validity of a statistics attitude survey. *Educational and Psychological Measurement, 40*(1), 235–238.
[Online: https://doi.org/10.1177/001316448004000138]

Rubin, D. B., Stuart, E. A., & Zanutto, E. L. (2004). A potential outcomes view of value-added assessment in education. *Journal of Educational and Behavioral Statistics*, *29*(1), 103–116.
[Online: https://doi.org/10.3102/10769986029001103]

Ruzek, E. A., Domina, T., Conley, A. M., Duncan, G. J., & Karabenick, S. A. (2015). Using value-added models to measure teacher effects on students' motivation and achievement. *The Journal of Early Adolescence, 35*(5–6), 852–882.
[Online: https://doi.org/10.1177/0272431614525260]

Schau, C. (2003, August). *Students' attitudes: The "other" important outcome in statistics education.* Paper presented at the Joint Statistical Meetings, San Francisco, CA.
[Online: http://www.statlit.org/pdf/2003SchauASA.pdf]

Schau, C., & Emmioğlu, E. (2012). Do introductory statistics courses in the United States improve students' attitudes? *Statistics Education Research Journal, 11*(2), 86–95.
[Online: https://iase-web.org/documents/SERJ/SERJ11(2)_Schau.pdf]

Schau, C., Stevens, J., Dauphinee, T. L., & Vecchio, A. D. (1995). The development and validation of the Survey of Attitudes Toward Statistics. *Educational and Psychological Measurement, 55*(5), 868–875.
[Online: https://doi.org/10.1177/0013164495055005022]

Selya, A. S., Rose, J. S., Dierker, L. C., Hedeker, D., & Mermelstein, R. J. (2012). A practical guide to calculating Cohen's f^2, a measure of local effect size, from PROC MIXED. *Frontiers in Psychology, 3*.
[Online: https://doi.org/10.3389/fpsyg.2012.00111]

Songsore, E., & White, B. J. (2018). Students' perceptions of the future relevance of statistics after completing an online introductory statistics course. *Statistics Education Research Journal, 17*(2), 120–140.

[Online: https://iase-web.org/documents/SERJ/SERJ17(2)_Songsore.pdf]

Thomas, S. L., & Heck, R. H. (2001). Analysis of large-scale secondary data in higher education research: Potential perils associated with complex sampling designs. *Research in Higher Education*, *42*(5), 517–540.
[Online: https://doi.org/10.1023/A:1011098109834]

Tishkovskaya, S., & Lancaster, G. A. (2012). Statistical education in the 21st century: A review of challenges, teaching innovations and strategies for reform. *Journal of Statistics Education, 20*(2).
[Online: https://doi.org/10.1080/10691898.2012.11889641]

Umugiraneza, O., Bansilal, S., & North, D. (2018). Examining teachers' perceptions about improving the teaching and learning of mathematics and statistics. *Statistics Education Research Journal*, *17*(2), 239–254.
[Online: https://iase-web.org/documents/SERJ/SERJ17(2)_Umugiraneza.pdf]

Usiskin, Z., & Hall, K. (2015). The relationships between statistics and other subjects in the K-12 curriculum. *Chance, 28*(3), 4–18.
[Online: https://doi.org/10.1080/09332480.2015.1099361]

van Es, C., & Weaver, M. M. (2018). Race, sex, and their influences on introductory statistics education. *Journal of Statistics Education*, *26*(1), 48–54.
[Online: https://doi.org/10.1080/10691898.2018.1434426]

Vanhoof, S., Kuppens, S., Sotos, A. E. C., Verschaffel, L., & Onghena, P. (2011). Measuring statistics attitudes: Structure of the Survey of Attitudes toward Statistics (SATS-36). *Statistics Education Research Journal*, *10*(1), 35–51.
[Online: https://iase-web.org/documents/SERJ/SERJ10(1)_Vanhoof.pdf]

Waters, L., Martelli, T. A., Zakrajsek, T., & Popovich, P. M. (1988). Attitudes toward statistics: An evaluation of multiple measures. *Educational and Psychological Measurement*, *48*(2), 513–516.
[Online: https://doi.org/10.1177/0013164488482026]

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, *67*(3), 223–248.
[Online: https://doi.org/10.1111/j.1751-5823.1999.tb00442.x]

Wise, S. L. (1985). The development and validation of a scale measuring attitudes toward statistics. *Educational and Psychological Measurement*, *45*(2), 401–405.
[Online: https://doi.org/10.1177/001316448504500226]

Xu, C., & Schau, C. (2019). Exploring method effects in the six-factor structure of the Survey of Attitudes Toward Statistics (SATS-36). *Statistics Education Research Journal*, *18*(2), 39–53.
[Online: https://iase-web.org/documents/SERJ/SERJ18(2)_Schau.pdf]

CHAO XU
University of Houston – Clear Lake
Department of Curriculum and Instruction
2700 Bay Area Blvd
Houston, TX 77058

APPENDIX

Table A1. Sample means and standard deviations for missing value-imputed and -deleted data

	Pretest				Posttest			
	Imputed (n = 1924)		Deleted (n = 1751)		Imputed (n = 1924)		Deleted (n = 1751)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Effort	6.42	0.86	6.45	0.87	5.97	1.06	5.96	1.06
Cognitive Competence	4.93	1.01	4.95	1.01	5.07	1.11	5.10	1.10
Affect	4.23	1.09	4.24	1.10	4.38	1.30	4.36	1.30
Difficulty	3.59	0.77	3.54	0.76	3.72	0.91	3.75	0.92
Value	5.21	0.97	5.22	0.98	5.00	1.11	5.03	1.11
Interest	4.86	1.20	4.84	1.21	4.42	1.46	4.39	1.45

Table A2. Relationships between student characteristics and statistics attitudes

	Interest	*Difficulty*	*Value*	*CogComp*	*Affect*
Age	0.02	-0.01	0.02	-0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Male	0.18**	0.08*	0.02	0.13*	0.16*
	(0.05)	(0.03)	(0.05)	(0.05)	(0.07)
prior_math	0.10***	0.05***	0.08***	0.21***	0.17***
	(0.02)	(0.01)	(0.02)	(0.02)	(0.03)

Note. *p < 0.05, **p < 0.01, ***p < 0.001. Regression coefficient estimates are unstandardized.