Disruptiveness of COVID-19: Differences in Course Engagement, Self-Appraisal, and Learning

Teresa M. Ober (ID)
Ying Cheng
Matthew F. Carter
Cheng Liu (ID)
University of Notre Dame

We investigated how the transition to remote instruction during the COVID-19 pandemic affected students' engagement, self-appraisals, and learning in advanced placement (AP) Statistics courses. Participants included 681 (M_{age} =16.7 years, SD_{age} =.90; %female=55.4) students enrolled in the course during 2017–2018 (N=266), 2018–2019 (N=200), and the pandemic-affected 2019–2020 (N=215) school years. Students enrolled during the pandemic-affected year reported a greater improvement in affective engagement but a decrease in cognitive engagement in the spring semester relative to a previous year. Females enrolled in the pandemic-affected year experienced a greater negative change in affective and behavioral engagement. Students enrolled during the pandemic-affected year reported a greater decrease in their anticipated AP exam scores and received lower scores on a practice exam aligned with the AP exam compared to a prior year. Although students were resilient in some respects, their self-appraisal and learning appear to have been negatively affected by pandemic circumstances.

Keywords: *adolescence, advanced placement, computerized testing, COVID-19, disparities, engagement, high schools, item response theory, learning, learning environments, longitudinal studies, mathematics education, propsentiy-scores, regression analyses, self-appraisal, statistics education*

Introduction

The COVID-19 pandemic has undoubtedly had a direct impact on physical and material aspects of learning in school settings. Widespread school closures in response to the COVID-19 pandemic are estimated to have affected at least 50 million students enrolled in K–12 school-based settings in the United States beginning in March 2020 (UNESCO, 2021). Most school closures within the United States began shortly after COVID-19 was declared a pandemic on March 11 and a national emergency on March 13, 2020 (Federal Register, 2020), as state governments began issuing specific orders (e.g., State of Indiana 2020a, 2020b). Instructors and students rapidly transitioned from in-person to remote and online teaching and learning environments. The impact of such a rapid transition and its long-term effect on factors associated with student learning are still largely unknown, given the unprecedented nature of the circumstances. Not only the modality of instruction but, in many cases, the scope of content within the course itself substantially differed from previous years, given the need to reprioritize (Reimers & Schleicher, 2020). Such swift and drastic instructional and curricular changes created obvious limitations in the use of psychoeducational assessments to measure student learning

in a valid and reliable way that enabled comparisons with previous academic years (Stifel et al., 2020).

Compounding these difficulties in capturing useful information that could be used for monitoring student learning, the rippling impact of the pandemic itself is thought to have exacerbated existing social inequalities (Dorn et al., 2020). Recent evidence has shown a negative impact on the social, emotional, physical, and mental health of K–12 students, with students from minoritized communities likely to have been more negatiely affected (McKune et al., 2021). As such, to understand the impact of the COVID-19 pandemic, we must consider not only differences in learning outcomes but also differences in motivation-related factors, such as engagement and self-efficacy. It is important to evaluate whether differences in such factors are emerging for students from historically marginalized groups, as accumulating evidence suggests a widening of disparities induced by the pandemic circumstances (Darling-Aduana et al., 2022).

Impact of COVID-19

Engagement. It is reasonable to expect that students would and could not be as engaged in learning as they had been in

previous academic years under the stressful circumstances of COVID-19. Such academic engagement is thought to comprise three dimensions related to affective (i.e., interest and motivation), behavioral (i.e., observable participation), and cognitive (i.e., appraisals, self-concept, and self-confidence) processes (Fredricks, 2011). Research regarding undergraduate students' attitudinal changes during the pandemic suggests marked decreases in students' engagement in coursework, perceptions of their own present academic success, and goals related to academic achievement (Daniels et al., 2021). Disengagement is associated with diminished learning outcomes in online settings (Sun & Rueda, 2012). Even under non-pandemic circumstances, K–12 students who have never taken online courses may especially struggle to remain engaged and succeed in the courses and beyond (Hart et al., 2019).

Self-Efficacy. Aside from students' engagement in learning having been affected by school closures during the COVID-19 pandemic, other factors associated with learning are likely to have been negatively affected. Students' self-efficacy and ability to monitor their own performance are likely to have suffered. *Self-efficacy* refers to someone's beliefs about abilities to produce desired levels of performance in activities or events that have some effect on their lives (Bandura, 1994). Some undergraduate students tended to express diminished self-efficacy by reporting a decrease in their academic achievement at the onset of the pandemic (Daniels et al., 2021). Although past research has found that students are remarkably accurate in predicting their performance on AP exams (Ober, Hong, et al., 2022), their ability to do so accurately under normal circumstances appears to necessitate that they be provided standards to calibrate their performance against (Nederhand et al., 2019). During the pandemic, many instructors adapted their approaches to instruction, assessment, and provision of constructive feedback to prioritize other needs, such as students' well-being (UNESCO, 2021). As such, students may not have had sufficient opportunities to self-monitor and -evaluate their learning. The ability to self-monitor one's learning has long been regarded as a critical factor that contributes to academic achievement within a self-regulated learning framework (Zimmerman, 1990). Some evidence indicates that the association between self-efficacy and academic performance is reciprocal, suggesting that initial differences in self-efficacy may contribute to widening disparities in academic achievement (Talsma et al., 2018). Students during the pandemic-affected year are likely to not only have had lower self-efficacy but also may have been less accurate in self-monitoring and appraising their learning, given the limited opportunities for social interaction and feedback with teachers and classmates (Tannert & Gröschner, 2021).

Learning. Growing evidence indicates that the constraints on the context of emergency online and remote learning affected not only students' engagement and ability to accurately self-appraise their performance but also their ability to learn and make measurable progress (Engzell et al., 2021). Given the associations between engagement, self-monitoring, and learning (Lei et al., 2018), evidence of a decrement in one aspect is likely to foreshadow a decrease in the overall quality of students' learning experiences. However, due to existing social inequalities, students may have experienced the impact of the pandemic with greater severity than others (Van de Werfhorst, 2021). Some have pointed to evidence of a "digital divide" as a factor that could have affected access to online educational resources amid the transition to remote learning (Bacher-Hicks et al., 2021) and, thus, opportunities to learn (Lai & Widmar, 2021). The *digital divide* refers to the unequal access or use of digital technologies within society (Van Dijk, 2017). Past research has found that such differences in access and use based on gender (Elena-Bucea et al., 2020), race/ethnicity (Yoon et al., 2020), and socioeconomic status (Harris et al., 2017) affect K–12 students' educational opportunities, even under "normal" pre-pandemic circumstances (Dolan, 2016). Aside from the access to online educational resources for learning, school closures themselves may have cut students off from other critical resources necessary to support their mental and physical health and overall well-being (Martin & Sorensen, 2020; Pattison et al., 2021). These combined factors contributed to the "twin pandemic" in the United States, such that minoritized individuals, particularly those so-categorized based on race/ethnicity and socioeconomic status, experienced greater hardship during the pandemic due to inequalities related to differences in structural and societal factors as access to health resources (Chen et al., 2021).

Student Learning Within Advanced Placement Statistics

There is widespread concern that students did not achieve learning gains during the 2019–2020 COVID-19 pandemic-affected year comparable to grade-level norms established in previous years (Daniel, 2020) and that the circumstances also furthered existing disparities in the quality of education (Lewin, 2020). Yet it is difficult to quantify the extent of the loss of learning opportunities due to the challenges in assessing student progress (Jankowski, 2020). Examining differences in student learning because of the COVID-19 pandemic requires a comparison across instructional contexts that otherwise maintain some level of consistency. Advanced placement (AP) courses may provide a naturalistic context in which making comparisons between the pandemic-affected year and prior years not only is feasible but also may be especially informative. Across the United States, high school students who take AP classes are taught in a way that prepares them to take the cumulative AP exam, which is a standardized exam administered nationally. Given that there are certain requirements for AP courses to be authorized (CollegeBoard, 2022a), teachers are often provided

general (e.g., syllabus guide) and subject-specific (e.g., summer institutes, online workshops) teaching resources, and the course culminates in a standardized exam (CollegeBoard, 2022b), some consistency is to be expected in the course material across sections of the courses taught at different schools. Therefore, it may be possible to draw comparisons across multiple AP classes in the same subject area because the curriculum is likely to be consistent, even despite other contextual differences.

Students who enroll in AP courses are likely to be incentivized to complete the AP exam because they may receive college credit if they achieve a satisfactory grade on it. As such, changes in AP exam participation could forecast later trends in college enrollment (Bowers & Zhou, 2019). Students' AP exam scores in math subject areas are strong predictors of postsecondary achievement (Ackerman et al., 2013). Achievement in math-related high school courses is also associated with other factors, such as higher postgraduate income (Goodman, 2019). Completing applied math courses such as those in statistics may be particularly beneficial for students who otherwise may not be oriented toward math or quantitative reasoning (Gottfried et al., 2014).

Even though the availability of AP programs is already relatively limited in schools serving predominantly African American or Hispanic/Latinx communities (Xu et al., 2021), changes in participation in the AP exam during the pandemic-affected year appear to have affected students from these communities more so than others. Current exam participation rates indicate that during the 2020 administration of the AP exams, participation waned across all subjects (−7% average change between 2019 and 2020; CollegeBoard, 2020a) but decreased especially among Black/African American (−16% average change between 2019 and 2020 across all subjects within the group) and Hispanic/Latinx (−13% average change between 2019 and 2020 across all subjects within the group) students (CollegeBoard, 2019, 2020b). In the context of science, technology, engineering, and mathematics (STEM) education, such students were considered underserved even prior to the onset of the pandemic (National Science Foundation, 2019). *Underrepresented minority* (URM) status is typically defined as belonging to a group whose number is substantially below the comparable figure for scientists and engineers who are not considered a racial or ethnic minority (U.S. Code, 2011). Further examining differences in factors related to student achievement within an AP Statistics course between the pandemic-affected 2019–2020 year and preceding academic years, as well as how such differences affected underserved students, could provide insight about future trends in students' postsecondary educational attainment.

Participation in certain AP course subjects may be especially informative about long-term enrollment trends. Students in AP courses in subject areas that require quantitative reasoning have been found to be one of the strongest predictors of persistence in a STEM major in college (Ackerman et al., 2013). Statistics and data science education is a subject area increasingly valued across multiple theoretical and applied disciplines (Ben-Zvi & Garfield, 2008; Wise, 2020). Students who enroll in statistics courses are likely to have a range of academic and professional interests, given that such coursework renders knowledge and skills for critically assessing data and statistical arguments to make educated judgments. Considering changes during the pandemic-affected year within the context of AP Statistics is specific enough to allow for comparisons with previous academic years yet likely to attract a cohort of students with varied academic and professional interests, making it possible to generalize beyond the context of statistics education.

Research Aims

We were interested in determining whether and the extent to which students' engagement in the course, self-appraisal of learning, and learning outcome in the context of AP Statistics was negatively affected in the 2019–2020 academic year after the transition to remote online instruction relative to previous years. We were specifically interested in understanding whether students who are typically underserved in STEM (i.e., female and/or URM students) are more likely to have been negatively affected. To pursue this aim, we examined changes in end-of-year outcomes related to these factors during the pandemic-affected 2019–2020 academic year in comparison to two previous academic years (i.e., 2017–2018, 2018–2019) through the following research questions.

Baseline Differences Between Cohorts in Engagement and Self-Appraisal

- **RQ1:** Did students' "pre-pandemic" (beginning of the academic year until February to early March) engagement (based on self-report and behavioral indicators) and self-appraisal in the pandemic-affected year differ relative to the prior academic year(s) at comparable time points?

Cohort Differences in Change From Baseline to End-of-Year in Engagement and Self-Appraisal

- **RQ2: (a)** Did the change in engagement (based on self-report and a behavioral indicator) and self-appraisal between pre- (February to early March) and post-pandemic (April to May) periods differ in the pandemic-affected cohort compared to the prior academic year(s)? **(b)** To what extent is change moderated by gender, URM status, or their interaction?

End-of-Year Differences in Self-Appraisal and Learning

- **RQ3: (a)** Did the accuracy and bias of students' end-of-year self-appraisals differ in the pandemic-affected cohort when compared to the preceding year(s)? **(b)** Do the accuracy and bias of students' end-of-year self-appraisals vary based on gender, URM status, or their interaction?
- **RQ4: (a)** Did students' end-of-year knowledge of statistics differ in the pandemic-affected cohort compared to a prior academic year? **(b)** Does students' learning vary based on gender, URM status, or their interaction?

Considering the literature reviewed previously, we had several expectations. First, we anticipated a greater decrease in engagement during the "post-pandemic" period at the end of the 2019–2020 academic year relative to the preceding years. Given that the AP exam format and content differed from preceding years, not to mention the impact of the sheer stress induced by the circumstances, we also anticipated that students' self-appraisal of their anticipated AP exam scores would show a greater decrease from the beginning to the end of the spring semester in the pandemic-affected year compared to a previous year and would be less accurate and more susceptible to bias than previous years. Because the format and content for the AP exam differed so dramatically from previous years (CollegeBoard, 2020a), we were limited in drawing conclusions about student learning based solely on their scores on the AP exam. However, we were able to make comparisons in students' ability based on their scores on a series of computerized assignments and one comprehensive practice exam that was aligned with the pre-pandemic general AP Statistics exam framework (CollegeBoard, 2010). We anticipated a difference in estimates of student learning, with students tending to receive lower scores in the 2019–2020 academic year compared with previous cohorts.

Methods

Participants

The sample consisted of students enrolled in AP Statistics (N_{total}=681) in seven high schools located in the state of Indiana (United States) during the 2017–2018 (*N*=266), 2018–2019 (*N*=200), and 2019–2020 (*N*=215) academic years. School leadership and AP Statistics teachers were invited to take part in the study through contacts and outreach to local schools. Students enrolled in classrooms taught by participating AP Statistics teachers provided consent and assent documentation prior to study enrollment. To be included in the analytic sample, participants also had to provide demographic information on a baseline survey, complete at least one of the first two online assignments administered before the spring semester, and not withdraw from the course. These criteria were set to ensure that the

sample consisted of students who were at least minimally engaged in the course before the period in the academic year corresponding with onset of the pandemic.

Similar to national participation in the AP Statistics exam in corresponding years (CollegeBoard, 2018, 2019, 2020b; see Table 1), the sample had more females (55.4%) than males (44.6%). Participants were ages 14–18 (*Mean age*=16.72 years, *SD*=.90) and identified as White/European American (66.6%), Asian/Asian American (19.7%), Black/African American (6.1%), Hispanic/Latinx (5.6%), Hawaiian/Pacific Islander (0.3%), and more than one race/ethnicity (7.8%). About 10.7% qualified for free or reduced-price lunches.

Measures

Across each academic year, data were collected at multiple time points to reflect students' engagement in the course, their appraisals of their scores on the AP exam, as well as their learning as measured by scores on a computerized practice exam. Figure 1 shows the timeline of the data collection during the academic year for each of the three cohorts of students. The descriptive statistics for each measure are reported separately for each cohort and in aggregate in Table 2. Some self-report measures were not administered in the 2018–2019 academic year (cohort 2) due to other research priorities. Additional information is available in the online repository associated with this manuscript (see Ober & Cheng, 2022).

Engagement (Cohorts 1 and 3)

Self-Reported Engagement (Cohorts 1 and 3). Micro-engagement—that is, engagement within the context of the AP Statistics course—was measured based on students' self-reporting using the Scale of Student Engagement in Statistics (Whitney et al., 2019). The Scale of Student Engagement in Statistics consists of 24 Likert-type items, with eight items each reflecting the affective, behavioral, and cognitive dimensions of engagement. Responses were provided by using a 5-point Likert-type scale indicating the extent to which participants agreed with the statement (1=*Strongly Disagree*, . . ., 5=*Strongly Agree*). The scale had acceptable reliability within each domain on the first occasion of measurement across all three cohorts (affective: *McDonald's* ω=.92; behavioral: *McDonald's* ω=.82; cognitive: *McDonald's* ω=.79) and had acceptable test-retest reliability based on the correlation of scale scores over time based on data collected during a year not affected by the pandemic (cohort 1; affective: *r*=.83; behavioral: *r*=.80; cognitive: *r*=.59).

Online Behavioral Indicators of Engagement (Cohorts 1, 2, and 3). As an indicator reflecting students' actual behavioral engagement within the context of the course, we derived a variable of students' use of one of the features of the online assessment platform. *Results checks* reflect the average number of times the user navigated to their score

TABLE 1
Demographic Information

	Study sample								U.S. AP Statistics exam participation		
	Cohort 1 (2017–2018)		Cohort 2 (2018–2019)		Cohort 3 (2019–2020)		Combined		2018	2019	2020
Total sample size (*N*)	266		200		215		681		212,626	209,569	178,886
Variable	*N*	*%*	*N*	*%*	*N*	*%*	*N*	*%*	*%*	*%*	*%*
Gender											
Male	118	44.4	79	39.5	107	49.8	304	44.6	47.4	47.3	47.5
Female	148	55.6	121	60.5	108	50.2	377	55.4	52.6	52.7	52.4
Age (years)									N/A	N/A	N/A
≤ 15	26	17.3	18	9.0	30	14.1	94	13.8	-	-	-
16	37	13.9	32	16.0	57	26.6	126	18.5	-	-	-
17	153	57.5	114	57.0	89	41.6	356	52.3	-	-	-
≥ 18	30	11.3	36	18.0	38	17.8	104	15.3	-	-	-
Race/Ethnicity											
American Indian / Alaskan Native	1	0.4	1	0.5	0	0.0	2	0.3	0.2	0.2	0.3
Asian/Asian American	61	22.9	44	22.0	19	8.8	124	18.2	18.2	19.1	19.7
Black/African American	12	4.5	19	9.5	7	3.2	38	5.6	5.2	5.1	4.0
Hawaiian/Pacific Islander	1	0.4	1	0.5	0	0.0	2	0.3	0.2	0.1	0.1
White/European American	157	59.0	106	53.0	155	72.1	418	61.4	54.4	53.3	53.9
Hispanic/Latinx	11	4.1	9	4.5	15	7.0	35	5.1	15.7	16.0	14.9
Other	2	0.8	3	1.5	1	0.4	4	0.6	-	-	-
Multiracial	20	7.5	16	8.0	17	7.9	53	7.8	4.5	4.5	4.7
Prefer not to respond or did not respond	1	0.4	1	0.5	1	0.4	3	0.4	1.7	2.0	2.3
Eligible for free/reduced-priced lunch									N/A	N/A	N/A
Yes	30	11.3	15	7.5	27	12.6	72	10.6	-	-	-
No	235	88.3	175	87.5	182	85.0	603	88.5	-	-	-
Prefer not to respond	1	0.4	10	5.0	5	2.3	6	0.8	-	-	-
Highest education of parent/guardian									N/A	N/A	N/A
Did not finish high school	10	3.8	3	1.6	10	4.7	23	3.4	-	-	-
High school diploma or GED	12	4.5	4	2.1	14	6.5	30	4.5	-	-	-
Attended some college; no degree	7	2.6	6	3.1	13	6.0	26	3.9	-	-	-
Associate degree (AA, AS, etc.)	4	1.5	8	4.2	11	5.1	23	3.4	-	-	-
Bachelor's degree (B.A., B.S., etc.)	89	33.5	62	32.3	85	39.5	236	35.1	-	-	-
Master's degree (M.A., M.S., etc.)	78	29.3	60	31.3	53	24.7	191	28.4	-	-	-
Doctoral or professional degree (Ph.D., J.D., M.D., etc.)	66	24.8	49	25.5	29	13.5	144	21.4	-	-	-
AP score											
No scores – Did not take	9	3.4	4	2.0	20	9.3	33	4.9	N/A	N/A	N/A
Scores received	257		196		195		648				
1	10	3.9	6	3.1	19	9.7	35	5.4	14.0	14.7	16.2
2	33	12.8	19	9.7	32	16.4	84	13.0	21.1	18.4	20.7
3	60	23.4	57	29.1	52	26.7	169	26.1	25.1	26.6	23.1
4	63	24.5	48	24.5	47	24.1	158	24.4	16.1	19.3	21.7
5	91	35.4	66	33.7	45	23.1	202	31.2	23.8	21.0	18.3

Note. AP = advanced placement; GED = general equivalency diploma.

reports showing information about their performance on the assignments. Within the pages containing information about their scores on the assignments, students could view their mastery of certain attributes of statistics knowledge,
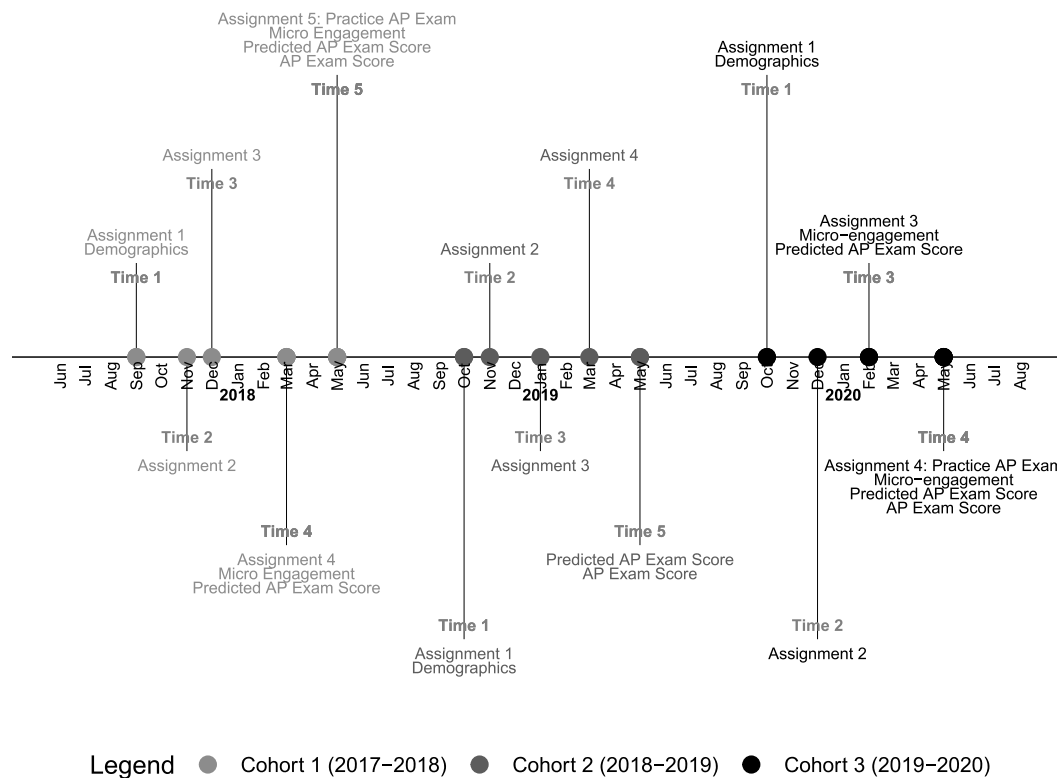
FIGURE 1. *Timeline of the administration of each measure.*

assignment scores, scores on individual items within an assignment, as well as step-by-step answer solutions for items within an assignment. Two variables were derived: one reflecting the average number of times the student checked their results on the first two assignments (administered between September and December for all three cohorts) and the other reflecting the number of checks for the practice AP exam (administered in May for cohorts 1 and 3). Previously conducted analyses of the data had shown this number to be significantly associated with multiple assessment formats even after controlling for self-reported engagement (Ober, Hong, et al., 2021).

Self-Appraisal (Cohorts 1 and 3)

Information about students' self-appraisals of their learning was derived from their AP exam scores and from the scores they self-reported anticipating receiving on the AP exam. Each of these variables is described in the paragraphs that follow.

Advanced Placement Exam Scores. The AP Statistics exam is administered nationally on one day in May each year in a given academic year to all students and is scored such that values range from 1 (lowest) to 5 (highest). Within the combined sample (cohorts 1, 2, and 3), scores ranged between 1 and 5 (*Mean*=3.63, *SD*=1.20, *Median*=4, *Mode*=5). Students

enrolled in certain AP programs experienced changes to the exam content during the May 2020 (cohort 3) administration of the exam. Students taking an AP Statistics course were prepared for a standardized and nationally administered in-person AP exam, yet given constraints related to social distancing, an online exam was administered instead, consisting of new question styles and updated content, resulting in less comprehensive coverage with the removal of specific topics (e.g., chi-square test, regression inference; Chu, 2020). Therefore, AP exam scores in 2020 (cohort 3), although included in our analysis, should not be directly compared to scores in previous years.

Predicted Advanced Placement Exam Score. Along with completing an assignment, students were asked the following question to gauge how well they believed that they would perform on the AP exam: "Whether you plan to take it or not, what do you predict your score will be on the AP Statistics exam?" Students enrolled in the course in the 2017–2018 (cohort 1) and 2019–2020 (cohort 3) academic years were asked this question on two occasions: in early spring (late February or early March) and again in May. Students enrolled in the 2018–2019 (cohort 2) academic year were also asked this question, but only once, in May. The response options ranged between 1 and 5, reflecting the scoring outcomes of the actual AP exam. The polychoric

TABLE 2

Descriptive Statistics for Predictors and Outcomes (N=681)

	Cohort 1 (2017–2018)						Cohort 2 (2018–2019)						Cohort 3 (2019–2020)						Combined sample					
	N	*M*	*SD*	*Median*	*Min*	*Max*	*N*	*M*	*SD*	*Median*	*Min*	*Max*	*N*	*M*	*SD*	*Median*	*Min*	*Max*	*N*	*M*	*SD*	*Median*	*Min*	*Max*
Engagement																								
Self-reported																								
Affective (~late Feb/early March)	202	3.33	0.78	3.38	1.00	5.00	NA						161	3.33	0.88	3.38	1.13	5.00	363	3.33	0.83	3.38	1.00	5.00
Behavioral (~late Feb/early March)	202	3.61	0.70	3.63	1.00	5.00	NA						161	3.85	0.65	3.88	1.50	5.00	363	3.71	0.69	3.75	1.00	5.00
Cognitive (~late Feb/early March)	202	3.17	0.56	3.25	1.13	5.00	NA						161	3.58	0.64	3.63	1.75	5.00	363	3.35	0.63	3.38	1.13	5.00
Affective (May)	200	3.32	0.85	3.38	1.00	5.00	NA						129	3.31	0.85	3.38	1.00	5.00	329	3.32	0.85	3.38	1.00	5.00
Behavioral (May)	200	3.56	0.75	3.63	1.00	5.00	NA						129	3.81	0.58	3.88	1.38	5.00	329	3.65	0.70	3.75	1.00	5.00
Cognitive (May)	200	3.28	0.71	3.31	1.00	5.00	NA						129	3.55	0.60	3.63	1.88	4.88	329	3.39	0.69	3.50	1.00	5.00
Behavioral engagement																								
Avg. results checks (Sept–Dec)	266	2.44	1.86	2.00	0.00	13.00	200	3.13	2.25	2.50	0.00	13.00	215	3.12	3.74	2.00	0.00	28.50	681	2.86	2.71	2.00	0.00	28.50
Results checks (March–May)	266	1.39	2.00	1.00	0.00	17.00	200	1.28	1.07	1.00	0.00	7.00	215	1.53	4.07	1.00	0.00	48.00	681	1.40	2.67	1.00	0.00	48.00
Self-appraisal																								
Predicted AP exam score																								
February	171	3.80	0.87	4.00	1.00	5.00	NA						145	3.60	0.98	4.00	1.00	5.00	316	3.71	0.93	4.00	1.00	5.00
May	186	3.82	0.94	4.00	1.00	5.00	152	3.80	0.96	4.00	1.00	5.00	117	3.50	0.96	4.00	1.00	5.00	455	3.73	0.96	4.00	1.00	5.00
Change (February–May)	128	0.01	0.62	0.00	-1.00	2.00	NA						107	-0.17	0.54	0.00	-2.00	2.00	235	-0.07	0.59	0.00	-2.00	2.00
Learning outcomes																								
Assignment scores (Sept–Dec)	265	0.12	0.76	0.17	-1.87	2.06	198	0.49	0.72	0.49	-1.63	2.06	213	-0.18	0.63	-0.22	-1.69	1.62	676	0.13	0.76	0.14	-1.87	2.06
Practice AP exam score (May)	236	0.24	0.74	0.26	-2.29	1.87	183	0.24	0.63	0.23	-1.29	1.64	168	-0.12	0.74	-0.10	-1.76	1.67	587	0.14	0.73	0.15	-2.29	1.87
AP exam score	257	3.75	1.18	4.00	1.00	5.00	196	3.76	1.11	4.00	1.00	5.00	195	3.34	1.27	3.00	1.00	5.00	648	3.63	1.20	4.00	1.00	5.00
Final class grade (%)	135	0.88	0.08	0.89	0.67	1.04	200	0.85	0.09	0.87	0.53	1.01	215	0.88	0.08	0.89	0.49	1.01	550	0.87	0.09	0.88	0.49	1.04

Note. AP = advanced placement.

7

correlation between the two predicted scores was .77 among students in cohort 1.

Accuracy and Bias Estimates. Two separate estimates were derived from all three cohorts to reflect aspects of students' self-appraisals of their performance on the AP exam. *Accuracy* was defined as the absolute value of the distance of student's predicted scores from their actual AP exam scores (*M*= 1.04, *SD*= 1.03). *Bias* was defined as the over- or under-prediction and consisted of three levels: underconfident (negative estimate), no bias (0), and overconfident (positive estimate). In this case, the tendency toward underconfidence was specifically examined and found to be present among 46.1% of students in the sample combined across all three cohorts.

Learning Measures

Cumulative Assignment Scores (Cohorts 1, 2, and 3). Between September and early December in each academic year, students completed at least two online sectional assignments on AP Statistics content. Students across each cohort were given the same or a comparable set of questions in each assignment, although not in the same order. Each assignment consisted of no more than 30 multiple-choice and blank-field items similar in format and aligned with the content of the AP Statistics test framework (CollegeBoard, 2010). Although all the assignments were graded for effort and completion, performance on the assignments was calculated based on the cumulative data available from students' past work on the assignments up until that point. Students' scores on each assignment were computed based on a Rasch model, a type of item response theory (IRT) model. Larger scores indicate better performance, and scores with greater absolute values are indicative of greater deviation from the expected population mean. Internal consistency was calculated across a total of 179 unique items (*Mean* number of items completed by each student=40.61, *SD*=13.88, *Median*=38) administered in assignments that were acceptable (*McDonald's* ω=.97). Cumulative assignment scores were calculated from an aggregation of items administered across all assignments a student completed. This variable was not used as a learning outcome but rather to compare and control for baseline differences between cohorts.

Practice Advanced Placement Exam Scores (Cohorts 1 and 3). Students enrolled in the 2017–2018 (cohort 1) and 2019–2020 (cohort 3) academic years completed one practice exam administered through the online assessment platform in May. The practice exam was a comprehensive assessment that was developed by the research team in consultation with content experts, mimicking the content coverage of the actual AP Statistics exam. Students completed the test for completion credit in their AP Statistics course and were allowed to complete it only once. Students' IRT scores on the practice exam were computed in a similar fashion to

the homework scores, were also based on a Rasch model, and ranged between −2.29 and 1.87 (*Mean*=.14, *SD*=.73), with higher scores indicative of better performance. The reliability across all abilities was acceptable (*McDonald's* ω=.93), based on the same method used to calculate the reliability of the cumulative assignment score.

Procedure

This study was preregistered (Ober, 2021, February 17). Additional details can be found in the Online Supplemental Materials about the methods and results of data cleaning (Appendix A), attrition analysis (Appendix B), multiple imputation (Appendix C), and calculation of the propensity scores (Appendix D).

Analytic Plan

From preliminary analysis, we found evidence of missingness in several outcomes that differed between cohorts; across each cohort, missingness was associated with students' assignment scores, and thus the assignment score was used for multiple imputations (see Online Supplemental Materials). We therefore attempted to account for differences in the composition of the cohorts through propensity score-adjusted regression. Figure 2 provides a flowchart of the analyses conducted and how they relate to the research questions and measures previously described. The analyses were conducted in several phases, all within the *R* statistical environment version 4.1.0 (R Core Team, 2021).

As noted, considering that differences were found in the demographic composition of cohorts, we subsequently conducted propensity-score adjusted regression. First, we conducted analyses to determine whether engagement and self-appraisals differed across the three cohorts during the "pre-pandemic" period, between late February and early March (**RQ1**). Differences in the change score (i.e., difference-in-difference) for engagement and self-appraisal were then examined as a function of cohort (**RQ2a**). We then examined whether end-of-year self-appraisals among students enrolled in the pandemic-affected academic year differed from those of students enrolled in a prior academic year (**RQ3a**). Next, we examined the extent to which knowledge of statistics among students enrolled in the pandemic-affected academic year differed from that of students enrolled in a prior academic year (**RQ4a**).

For each outcome variable, analyses were also conducted to determine whether after controlling for students' schools there were differences in the outcome variable based on gender (female=1) and URM status (URM=1) in the pandemic-affected cohort (cohort 3) using imputed data (**RQ2b**, **RQ3b**, **RQ4b**). Although we would have preferred to also include a predictor reflecting the students' socioeconomic status (i.e., eligibility for free/reduced-price lunch), preliminary analyses revealed that it was not independent of URM status
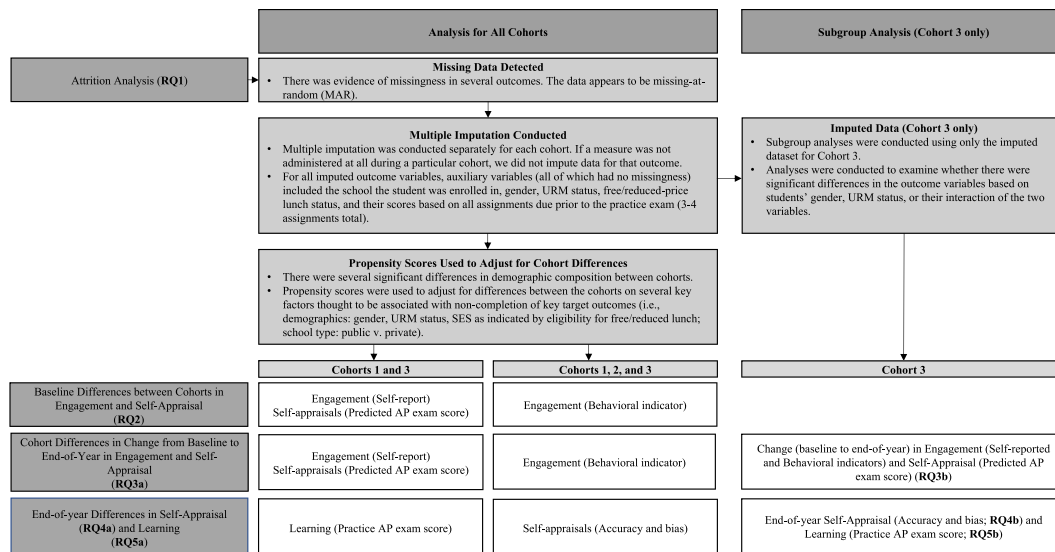
FIGURE 2. *Schematic of analytic procedure.*

(χ^2(df=1)=28.46, p<.001). As such, we focused on the latter, particularly as it would allow us to consider possible interactive effects with gender. Given that this analysis was focused solely on students enrolled in the pandemic-affected year (cohort 3), propensity scores were not used. Analyses with continuous outcome variables (i.e., change score for engagement, self-appraisals of performance on the AP exam, accuracy of score predictions, mock AP scores) were conducted as separate ANOVAs, while analysis for the dichotomously coded outcome variable (i.e., bias of score predictions) involved a logistic regression with the binomial link function.

Results

We examined the impact of COVID-19 on students' engagement, self-appraisal, and learning outcomes within the context of AP Statistics during the 2019–2020 academic year relative to previous cohorts. Furthermore, we were interested in determining whether students who are typically underserved in STEM (i.e., females, URM students) were more likely to have been negatively affected. Table 3 provides a summary of the main results of the analyses, with more details provided in the paragraphs that follow. Where data involved a comparison across multiple cohorts, the propensity score weighted data were used.

Baseline Differences Between Cohorts

We first examined cohort differences in students' engagement, self-appraisal, and assignment scores measured during a baseline "pre-pandemic" period before the transition to online and remote learning would have occurred for students enrolled in 2020 (**RQ1**). Table 4 shows the results of the propensity score adjusted regression models for this analysis.

Engagement

Self-Reported. During the 2017–2018 (cohort 1) and 2019–2020 (cohort 3) academic years, a measure of micro-engagement was administered in the early spring semester (late February/early March). Results did not reveal a significant difference between cohorts in affective engagement (β=.01, *SE*=.09, *p*=.824). However, we did find a significant difference in behavioral (β= −.20, *SE*=.07, *p*<.001) and cognitive engagement (β= −.33, *SE*=.06, *p*<.001). For the dimensions of behavioral and cognitive engagement, students in cohort 3 (behavioral: *Marginal M*= 3.87, *SE*=.05; cognitive: *Marginal M*=3.58, *SE*=.05) reported on average higher levels of engagement than those in cohort 1 (behavioral: *Marginal M*=3.60, *SE*=.05; cognitive: *Marginal M*=3.16, *SE*=.04) early in the respective academic year.

Behavioral Indicator. As a baseline behavioral indicator of engagement, we derived the average number of times the student checked the results page on the first two assignments between September to December of the academic year in which they were enrolled. We used the number of times the student checked the results page on the final assignment as an end-of-year behavioral indicator of engagement. Unlike the previous analysis on the micro-engagement dimensions, which relied on data from the pandemic-affected cohort and only one prior cohort, in the analysis of the behavioral indicators of engagement, we were also able to include data from all three cohorts. The results did not suggest a significant difference between the pandemic-affected year (cohort 3) compared to the 2018–2019 year (cohort 2; β=.008, *SE*=.26, *p*=.882). However, students in the pandemic-affected year (*Marginal M*=3.05, *SE*=.26) appeared to check their scores more frequently than students in the 2017–2018

TABLE 3

Summary of Main Findings

	Data sources						Findings				
	Cohort 1 (2017–2018)		Cohort 2 (2018–2019)		Cohort 3 (2019–2020)		Attrition (see Online Supplemental Materials)	Baseline differences between cohorts (RQ1)	Cohort differences in change (beginning-to-end of year) (RQ2a)	End-of-year differences between cohorts (RQ3a, RQ4a)	Subgroup differences in cohort 3 (RQ2b, RQ3b, RQ4b)
Outcomes	Baseline	End-of-year	Baseline	End-of-year	Baseline	End-of-year					
Engagement											
Self-reported	✓	✓		✓	✓	✓	- No differences between cohorts 3 and 1 in completion of baseline survey. - Cohort 3 students were less likely to complete end-of-year survey than Cohort 1 students.	- Cohort 3 students reported higher levels of behavioral and cognitive engagement than students in Cohort 1. - No differences in affective engagement.	- Cohort 3 students had a greater improvement in affective engagement than Cohort 1 students. - Cohort 3 students had a greater reduction in cognitive engagement than Cohort 1 students. - No differences found in behavioral engagement.	-	- Female students had a greater negative change in affective and behavioral engagement than non-female classmates. - No other demographic variables explained differences in change.
Behavioral indicator from process data	✓	✓	✓	✓	✓	✓	-	- Cohort 3 students appeared to check the results page more frequently than students in Cohort 2. - No differences between Cohort 3 and 1.	- No differences between Cohorts 1, 2, and 3.	-	- No demographic variables explained differences in change.
Self-appraisals											
AP exam score		✓		✓		✓	- Cohort 3 students were less likely to complete the AP exam Cohort 1 or 2 students.	-	-	-	-
Predicted AP exam score	✓	✓	✓	✓	✓	✓	-	- Cohort 1 students had on average slightly higher predicted scores at baseline.	- Cohort 3 students had a more negative change in their estimates of AP exam scores than Cohort 1 students.	-	- No demographic variables explained differences in change.
Accuracy of prediction		✓		✓		✓	-	-	-	- Cohort 3 students were less accuracy than Cohort 2 students. - No differences between Cohort 1 and 3.	- No demographic variables explained differences in students' accuracy of their score predictions.
Bias of prediction (overconfidence)		✓		✓		✓	-	-	-	- No differences between Cohorts 1, 2 and 3.	- No demographic variables explained differences in students' overconfidence bias of their score predictions.
Learning											
Assignment score	✓		✓		✓		-	- Cohort 3 students tended to receive lower scores than students in Cohort 1 or 2.	-	-	-
Practice AP exam score		✓		✓		✓	- Cohort 3 students were less likely to complete the practice AP exam Cohort 1 students.	-	-	- Cohort 3 students received lower scores than Cohort 1 students.	- No demographic variables explained differences in practice AP score.

Note. AP = advanced placement.

TABLE 4

*Results of Propensity Score Adjusted Regression Predicting Baseline and Change Scores for Engagement, Self-Efficacy, and Learning (**RQ1, RQ2a**)*

Cohort	Self-report: engagement (cohort 1 vs. 3 only; N=478)												Behavioral indicator: engagement (cohort 1, 2, 3; N=676)				Predicted score: self-appraisal (cohort 1 vs. 3 only; N=481)				Assignment scores: learning (cohort 1, 2, 3; N=676)			
	Affective				Behavioral				Cognitive															
	β	SE	t	p	β	SE	t	p	β	SE	t	p	β	SE	t	p	β	SE	t	p	β	SE	t	p
Baseline																								
Cohort 1 (2017–2018) = 1	0.012	0.089	0.223	.824	−0.195	0.071	−3.853	<.001	−0.329	0.063	−6.476	<.001	−0.112	0.259	−2.156	.031	0.11	0.104	2.000	0.046	0.196	0.066	4.943	<.001
Cohort 2 (2018–2019) = 1	-	-	-	-	-	-	-	-	-	-	-	-	0.008	0.260	0.148	.882	-	-	-	-	0.415	0.067	10.568	<.001
Change (post-baseline)																								
Cohort 1 (2017–2018) = 1	−0.357	0.068	−8.237	<.001	−0.054	0.05	−1.221	.223	0.091	0.063	2.189	.029	0.073	0.313	1.345	.179	0.17	0.15	4.129	<.001	-	-	-	-
Cohort 2 (2018–2019) = 1	-	-	-	-	-	-	-	-	-	-	-	-	−0.039	0.314	−0.745	.456	-	-	-	-	-	-	-	-

Note. Sample sizes reflect the analytic sample following imputation.

11

TABLE 5

Results of ANOVA Examining Subgroup Differences in Changes (Post-Baseline) in Engagement, Self-Efficacy, and Learning (Cohort 3 Only; N=215) (**RQ2b**)

| Predictor | Self-report: engagement | | | | | | | | | | | | Behavioral indicator: engagement | | | | Predicted score: self-appraisal | | | |
| | Affective | | | | Behavioral | | | | Cognitive | | | | | | | | | | | |
	df	F	p	$\eta^2_{partial}$	df	F	p	$\eta^2_{partial}$	df	F	p	$\eta^2_{partial}$	df	F	p	$\eta^2_{partial}$	df	F	p	$\eta^2_{partial}$
School	5	3.673	.003	0.074	5	1.233	.295	0.019	5	0.902	.481	0.015	5	1.413	.221	0.039	5	2.344	.043	0.031
Female (=1)	1	7.817	.006	0.034	1	5.075	.025	0.021	1	4.857	.029	0.027	1	0.460	.498	0.002	1	0.012	.915	<.001
URM (=1)	1	0.538	.464	0.003	1	1.610	.206	0.008	1	3.103	.080	0.015	1	0.067	.797	<.001	1	1.651	.200	0.008
Female (=1) * URM (=1)	1	1.339	.249	0.006	1	0.071	.791	<.001	1	2.775	.097	0.013	1	4.128	.044	0.020	1	0.613	.435	0.003
Residuals	206				206				206				206				206			

Note. Sample sizes reflect the analytic sample following imputation of cohort 3 responses.

year (cohort 1; *Marginal M*=3.09, *SE*=.17) at the beginning of the academic year (β=−.11, *SE*=.26, *p*=.031).

Self-Appraisal

We next sought to determine whether there were any apparent differences in students' self-appraisal of their AP exam scores. In the 2017–2018 and the 2019–2020 academic years, students were asked to estimate their score on the AP exam in the early spring semester (late February/early March). Students in the 2017–2018 year (cohort 1; *Marginal M*=3.81, *SE*=.06) had slightly higher predictions about their performance than students in the 2019–2020 year (cohort 3; *Marginal M*=3.60, *SE*=.08; β=.11, *SE*=.10, *p*<.05).

Assignment Scores

For a final baseline comparison, we wanted to determine whether students' assignment scores significantly differed between the three cohorts. We found that students in the pandemic-affected year (*Marginal M*=−.18, *SE*=.04) tended to perform worse on the first two assignments than students in the prior cohorts (cohort 1: *Marginal M*=.12, *SE*=.04; β=.20, *SE*=.07, *p*<.001; cohort 2: *Marginal M*=.50, *SE*=.05; β=.42, *SE*=.07, *p*<.001).

Cohort Differences in Change From Baseline to End-of-Year

We next examined whether there were cohort differences in students' change in engagement and self-appraisal of projected AP exam scores prior to (i.e., up until late February or early March) and following (i.e., after late April or early May) the transition to remote online teaching during the pandemic-affected 2019–2020 academic year (**RQ2a**). For this portion of the analysis, Table 4 also shows the results of the propensity score adjusted regression. Looking at the differences in the pre- and post-pandemic scores in the 2019–2020 academic year alone does not tell us whether the

difference is pandemic-related, because it could be due to natural growth or decline. Fortunately, the previous academic years can serve as a control. By comparing the pre- and post-pandemic difference in 2019–2020 against changes in previous years at comparable time points, we can get a more accurate understanding of the impact of the pandemic. A change score, reflecting the difference between end-of-spring semester responses (i.e., "post-pandemic") compared to the mid-spring semester responses (i.e., "pre-pandemic"), was calculated for each student's self-reported engagement in the course, their cumulative use of the results checking feature, and their predicted AP exam scores. We further wanted to determine the extent to which any change from pre- to post-pandemic differed based on gender and URM status among students enrolled in the 2019–2020 academic year (**RQ2b**). Table 5 shows the results of this subgroup analysis.

Engagement

Self-Reported. Separate analyses were conducted to determine whether change scores in each of the three dimensions of engagement differed between students enrolled in the 2017–2018 (cohort 1) and the pandemic-affected 2019–2020 (cohort 3) years. These analyses did not reveal a significant difference between the two cohorts in behavioral engagement (β=.05, *SE*=.05, *p*=.223). However, we did find significant cohort differences in the change from the beginning to the end of the spring semester in terms of affective (β=.36, *SE*=.07, *p*<.001) and cognitive (β=.09, *SE*=.06, *p*=.029) engagement. Although students in the pandemic-affected year (cohort 3) generally had improved affective engagement (*Marginal M*$_{change}$=.49, *SE*$_{change}$=.06), students in a prior year (cohort 1) tended to have decreased affective engagement (*Marginal M*$_{change}$=−.08, *SE*$_{change}$=.04). By contrast, students in the pandemic-affected year (*Marginal M*$_{change}$=−.20, *SE*$_{change}$=.03) appeared to show a greater reduction in cognitive engagement than students in a prior year (*Marginal M*$_{change}$=−.07, *SE*$_{change}$=.05).

TABLE 6
*Results of Propensity Score Adjusted Regression Predicting End-of-Year Outcomes (**RQ3a**, **RQ4a***)

Cohort	Predicted score (cohort 1, 2, 3; N=681)								Practice AP exam score: learning (cohort 1 vs. 3 only; N=478)			
	Accuracy				Underconfidence							
	β	SE	t	p	B	SE	t	p	β	SE	t	p
Assignment score (baseline)									0.680	0.034	18.05	<.001
Cohort 1 (2017–2018) = 1	−0.002	0.088	−0.048	.961	−0.211	0.187	−1.127	.260	0.064	0.049	2.073	.039
Cohort 2 (2018–2019) = 1	−0.100	0.094	−2.155	.032	−0.320	0.213	−1.504	.133	-	-	-	-

Note. Sample sizes reflect the analytic sample following imputation. Given that accuracy was measured by the absolute distance of the prediction from the actual score, a negative coefficient reflects that the cohort was overall more accurate estimate in estimating their performance.

We then examined differences in the change between demographic groups of students in the pandemic-affected year only (cohort 3), using the imputed data. ANOVA models were evaluated to determine whether gender (female=1), URM status (URM=1), or an interaction of the two variables explained any change in students' engagement. We found that female students had a significantly greater negative change in affective ($F(1,206)=7.82$, $p=.006$, $\eta_{partial}^2=.034$; $M_{diff}=-.287$, 95% CI=−.50 : −.07) and behavioral ($F(1,206)=5.08$, $p=.025$, $\eta_{partial}^2=.021$; $M_{diff}=-.142$, 95% CI=−.27 : −.01) engagement. In terms of cognitive engagement, female students appeared to have a less negative change than male students ($F(1,206)=4.86$, $p=.029$, $\eta_{partial}^2=.034$; $M_{diff}=.136$, 95% CI=.01 : .27). No other demographic variables explained significant change in students' cognitive engagement or the other two dimensions of engagement.

Behavioral Indicator. We calculated a difference score based on the number of times students checked the results page for the final assignment minus the average number of checks of the results page for all other assignments that had been administered previously. We were interested in seeing whether students were more or less likely to make use of this feature in the assessment platform after transitioning to online and remote teaching due to the pandemic relative to previous semesters. The findings did not provide evidence of a significant difference between cohort 3 with respect to either cohort 1 ($\beta=.07$, $SE=.31$, $p=.179$) or cohort 2 ($\beta=-.04$, $SE=.31$, $p=.456$). Among students enrolled in the pandemic-affected year, none of the demographic variables appeared to moderate the change in students' average checks of the results page ($p > .05$).

Self-Appraisal

Students in both the 2017–2018 (cohort 1) and 2019–2020 (cohort 3) years were asked to predict their score on the AP exam around early March and again in May prior to taking the AP exam. We were interested in determining whether there was a change in students' predictions between the two cohorts. It was anticipated that students in cohort 3 may experience a greater reduction in their expectations of their scores on the AP exam. As anticipated, we found that students in the pandemic-affected year ($Marginal\ M_{change}=-.42$, $SE_{change}=.09$) tended to have a more negative change in their estimates of AP exam scores than students in a prior year ($Marginal\ M_{change}=.16$, $SE_{change}=.10$; $\beta=.17$, $SE=.15$, $p<.001$). None of the demographic variables appeared to moderate the change in students' predicted AP exam scores among students in the pandemic-affected year ($p>.05$).

End-of-Year Differences

Self-Appraisal

Next, we wanted to determine whether students were more or less accurate or biased in predicting their scores on the AP exam in the pandemic-affected year (**RQ3a**; see Table 6). We also investigated whether there were subgroup differences (**RQ3b**; see Table 7). Two separate sets of analyses were conducted: one for accuracy and one for bias.

Accuracy. The accuracy values reflected the distance between actual and predicted AP exam scores. Poisson regression was used, given that the distribution most closely reflected this distribution. Students in the pandemic-affected year (cohort 3) appeared to be less accurate than students in the 2018—2019 year ($\beta=-.10$, $SE=.09$, $p=.032$), although there appeared to be no significant difference between students in the pandemic-affected year and 2017–2018 ($\beta=-.002$, $SE=.09$, $p=.961$). However, overall, students in the pandemic-affected year ($Marginal\ M=1.13$, $SE=.07$) tended to have greater differences between their estimated and actual scores than students in both prior years (cohort 1: $Marginal\ M=1.12$, $SE=.07$; cohort 2: $Marginal\ M=.90$, $SE=.08$). Thus, despite efforts to reduce the scope of the exam content, students in the pandemic-affected cohort may have struggled in estimating their performance on the exam, given the novelty of the

TABLE 7

Results of ANOVA Examining Subgroup Differences in End-of-Year Self-Appraisals and Learning Outcomes (Cohort 3 Only; N=215)
(RQ3b, RQ4b)

| | Predicted score | | | | | | | Practice AP exam score: learning | | | |
| | Accuracy | | | | Underconfidence | | | | | | |
Predictor	df	F	p	$\eta^2_{partial}$	df	Deviance	p	df	F	p	$\eta^2_{partial}$
School	5	2.504	.032	0.060	5	11.404	.044	5	6.959	<.001	0.131
Female (=1)	1	1.964	.163	0.008	1	0.056	.812	1	0.405	.525	0.002
URM (=1)	1	1.041	.309	0.005	1	2.058	.151	1	0.377	.540	0.002
Female (=1) * URM (=1)	1	0.494	.483	0.002	1	0.023	.881	1	0.932	.336	0.005
Residuals	206				206			206			

Note. Sample sizes reflect the analytic sample following imputation of cohort 3 responses.

exam content and format. Using imputed data from students in the pandemic-affected year only, we wanted to determine whether, after controlling for the effect of school, the accuracy of students' predictions varied based on gender, URM status, or the interaction of the two. There was no evidence that any of these background factors explained variation in the accuracy of students' predictions (*p*>.05).

Bias. We next examined differences in students' bias in terms of underconfidence (*1=underconfident, 0=not underconfident*). We were interested only in examining underconfidence, given that we anticipated that students in the pandemic-affected year would likely have a lower morale than students in the previous academic years, given the weighty context of the pandemic. Students in the pandemic-affected year did not appear to be significantly more underconfident than students in either prior cohort (cohort 1: *B*=−.21, *SE*=.19, *p*=.260; cohort 2: *B*=−.32, *SE*=.21, *p*=.133). The marginal means suggest that approximately 51.0% of students in the pandemic-affected year were underconfident in their self-appraisal relative to actual scores, compared to only 45.7% (cohort 1) and 43.0% (cohort 2). After controlling for the effect of school, we again did not find any evidence that the likelihood of being unconfident was associated with students' gender, URM, or the interaction of the two in the pandemic-affected cohort only (*p*>.05).

Learning

We were subsequently interested in examining whether students' IRT scores on the practice AP exam, an indication of learning, differed between students enrolled in the pandemic-affected year and the 2017–2018 year (cohort 1) (**RQ4a**; see Table 6). Given that we had found baseline differences in students' assignments scores derived from their answers to the first two assignments, we included that score as a covariate. Even while controlling for students' "pre-pandemic" assignment scores, the results indicated

significant differences in students' practice exam scores between the two cohorts (β=.06, *SE*=.05, *p*=.039). Students in the pandemic-affected 2019–2020 year (*Marginal M*=−.17, *SE*=.04) tended to receive lower scores on the practice exam than students in the 2017–2018 year (*Marginal M*=.14, *SE*=.04). We subsequently examined whether there were subgroup differences (**RQ4b**; see Table 7). Controlling for the effect of school, there was no evidence that practice exam scores varied based on gender, URM status, or the interaction of the two among students in the pandemic-affected year (*p*>.05)

Discussion

There presently remains a lack of understanding of the impact of the COVID-19 pandemic on student learning during the 2019–2020 academic year. There has been a great deal of speculation that the year marks a year of "learning loss" (Dorn et al., 2020; Pokhrel & Chhetri, 2021), yet systematically proving that is difficult because many annually administered standardized assessments were either substantially modified—in form, content, or both (as in the case of the AP Statistics exam)—or outright canceled. The present study used data collected in years preceding the pandemic-affected 2019–2020 year to compare differences in students' engagement, their self-appraisals of learning via their predicted AP exam scores, and their IRT scores on a practice aligned with content on the AP Statistic test framework. Acknowledging that the pandemic affected subgroups of students differently, we further sought to understand whether female students and students from underrepresented racial/ethnic minority groups in STEM were more or less at risk.

Given the constraints around online instruction, the hastiness in the transition to remote learning, and the inability to communicate as freely with teachers and peers, research has found some evidence that these perceptions transpired as changes in student engagement during COVID-19 (Chiu, 2021; Ober, Brodsky, et al., 2021). Students were less likely

to complete end-of-year assignments in cohort 3. With the concerns of the pandemic looming, daily stressors combined with a lack of accountability in completing schoolwork, students may have felt less pressure to complete assignments. We also found that students were less likely to participate in the AP exam, possibly for similar reasons of accountability or because they felt unprepared for the novel exam format or experienced untenable technical problems (Strauss, 2020).

Students may not have completed the assignments or the AP exam during the 2019–2020 year because they may not have felt as confident in their ability to perform well. Students in the 2019–2020 academic year reported a significantly more negative change in their cognitive engagement in the course between early March and May compared with students in a preceding year; however, they appeared to be significantly more affectively engaged (i.e., motivated to learn course material), at least based on self-reporting. Amid the stressful circumstances surrounding the pandemic, students enrolled in the course during the 2019–2020 academic year may have, indeed, felt that they were more motivated than previous cohorts, given that any effort dedicated to academics during this time may have felt like an extreme amount. However, greater perceived engagement may not necessarily always result in learning. Certain forms of affective engagement have been linked with a tendency toward academic burnout (Wang et al., 2015), which predicts lower academic achievement (Madigan & Curran, 2021). Alternatively, students may have perceived that they were more engaged, given that many teachers' reported lower expectations of what constituted an acceptable level of engagement in an effort to accommodate the extenuating circumstances (Huck & Zhang, 2021; Ober, Carter, et al., 2022). Although we did not find a significant difference in change based on URM status among students enrolled in the 2019–2020 academic year, we did find that female students tended to report a greater decrease in affective and behavioral engagement, although male students had a greater decrease in cognitive engagement. These findings may underscore gender differences among adolescent students in perceived engagement and support in online learning settings (see Korlat et al., 2021; Lessard & Puhl, 2021; Tate & Warschauer, 2022).

We also found that students in the 2019–2020 academic year tended to have a more substantial decrease in their self-appraisal of their performance on the AP exam from mid- to late spring compared to a preceding year. Although a previous investigation found that students enrolled in AP Statistics were relatively accurate in predicting their performance (Ober, Hong, et al., 2022), students in the pandemic-affected year may have experienced a drop in their self-efficacy brought about by environmental stressors (Camfield et al., 2021) or because they simply lacked confidence in completing an AP exam in an online format with changed content, which would have been unfamiliar. As such, it may have been difficult for students to gauge the likelihood of performing well on the exam because existing materials that could be used as self-assessment tools differed so dramatically from the actual exam. There was, however, no evidence of significant differences between students based on gender or URM status with respect to changes in self-appraisal.

The decrease in students' self-appraisal during the spring 2020 semester may have reflected legitimate concerns that affected their ability to learn. Emerging evidence suggests that students' learning in core subject areas was, indeed, negatively affected by school closures and other pandemic-related circumstances (National Assessment of Educational Progress, 2022a, 2022b). With respect to learning, after controlling for scores based on assignments submitted in the "pre-pandemic" fall semester, we found evidence that students tended to receive lower IRT scores on the practice exam compared with students enrolled in a prior academic year. Although this result by itself is not definitive evidence of a "learning loss," it does suggest that students during the pandemic-affected year tended not to achieve the same level of proficiency in their statistics knowledge as students in a preceding year. This finding comes even after attempting to adjust for baseline differences between cohorts to make them more comparable.

Implications

The findings here suggest several notable implications. First, students enrolled in the pandemic-affected 2019–2020 year performed significantly worse on the practice exam than students in the 2017–2018 year. These findings are consistent with trends for students enrolled in U.S. K–12 schools (Goldhaber et al., 2022; National Assessment of Educational Progress, 2022a, 2022b). Providing instructional opportunities to postsecondary students affected by the pandemic to reengage and relearn content they would have had the chance to master during the 2019–2020 academic year may be one way of addressing potential knowledge gaps.

Second, the findings point toward a conspicuous lack of differences in engagement in the expected direction relative to previous cohorts. This result may be taken as a sign that students may have been remarkably resilient even during what is likely the direst public health crisis they had ever experienced before and potentially in their lifetime. Even though there may be reason to be optimistic in this regard, further inquiry is clearly warranted, given the states for these students, especially given that our sample was limited to students enrolled in an AP course.

Third, although we expected to find gender and race/ethnicity subgroup differences among the pandemic-affected cohort of students, in the present study, we largely did not find any evidence of such effects. Given the associations between social and cultural factors and the quality of students' educational experiences during the pandemic (Goudeau et al., 2021), we believed that it was important to investigate potential differences. In some dimensions of

engagement, males and females in the pandemic-affected year appeared to experience change differently, although in terms of self-appraisal (in relative change and accuracy/bias) and practice exam scores, we did not observe a difference. Further work should nevertheless consider intersectional differences, particularly among a larger and more diverse sample of students.

Fourth, there remains a vexing methodological challenge in trying to estimate changes in learning among students who are more likely to have missing data. Even though the format of the practice exam remained largely the same between cohorts, our ability to draw comparisons in learning was mired by the drop in participation during the pandemic-affected year. In the present study, we therefore opted to impute missing data and use propensity score adjusted regression. Yet, as we describe in more detail in the next section, this solution was by no means unbiased (Cham & West, 2016).

Limitations

Despite the implications of the present findings, there are nevertheless several important limitations to acknowledge. The present study is limited to the sample from which the data were collected. Participants in the study were enrolled in an AP course and thus were unlikely to reflect the broader U.S. population of high school students. Students pursuing advanced coursework not only tend to be more academically motivated but also tend to reflect a demographic of students from families with higher income status (Judson et al., 2019). During the pandemic, students from financially secure households were less likely to have firsthand experience of the global health and economic repercussions of the pandemic (Van Lancker & Parolin, 2020). We also note that the proportion of students across all three cohorts who would meet the National Science Foundation's definition of URM in STEM is still relatively small. This appears to be a problem with AP Statistics course/exam participation more widely (see Table 1) and thus is not exclusively an issue with the study sample. Whether the present findings would be reflected in a more diverse sample deserves further inquiry.

Another limitation of the present study is that data were likely collected from students who were, in general, more engaged in the course. As is the case with many studies attempting to understand the full scope of the "twin pandemics"—that of COVID-19 and of the social inequalities that surfaced in its wake—on student learning, there seems to be an ever-present challenge of handling missing data (Kuhfeld et al., 2020). Using complete data in such a context may produce results that do not adequately reflect students who are more at risk for underachievement due to a general lack of participation or for withdrawing from the course (Suzuki et al., 2022). That the percentage of students who completed the practice exam was significantly lower in the 2019–2020 academic year relative to the 2 preceding years speaks to the

difficulty in estimating the effect of COVID-19 and school closures on student learning. We attempted to handle this issue by using multiple imputation after noting an missing at random (MAR) missing data mechanism, given that it is preferable to complete case analysis under such conditions, although not also bias-free (Hughes et al., 2019).

We attempted to balance the cohorts by using propensity score adjustments based on certain key baseline covariates. However, we note that there was considerable room for improving the balance between cohorts. In the present study, similarities between cohorts were likely affected because one school enrolled in the 2017–2018 year did not participate in the pandemic-affected 2019–2020 year, thus likely affecting the composition of the cohorts. We stratified based on the school factor, given that most schools (five of the seven) participated in all three academic years and thus we could use stratification to match students accordingly to mitigate bias due to school-level demographic and instructional confounders. However, this method of stratification did not guarantee that all students within the strata had the same or similar propensity scores and ensured that strata were not of equal size, which could have introduced bias (Adelson et al., 2017).

Conclusions

Understanding how engagement, self-efficacy, and learning differed in the pandemic-affected year has implications for the immediate and distant future of education. Consistent with emerging evidence (Goldhaber et al., 2022; National Assessment of Educational Progress, 2022a, 2022b), students in the pandemic-affected 2019–2020 academic year performed worse on an end-of-year assessment than students in previous years. Although we anticipated that student engagement would overall decrease significantly among students in the pandemic-affected school year, we did not find consistent evidence to support this. Instead, the findings may be an indication of the resilience of students. Students enrolled in the pandemic-affected academic year also tended to be significantly less accurate and slightly (although not significantly) more underconfident in their self-appraisal on the AP exam relative to their actual scores. These findings suggest how disorienting the pandemic-affected year was for many students, even students enrolled in AP courses who may tend to be more academically motivated.

Although the impact on students' long-term opportunities for learning is yet to be more fully understood, this study charts progress toward an understanding of the impact of the COVID-19 pandemic and school closures on students' learning. Despite the overwhelmingly negative effect of the pandemic, there may be reason to regard it as a disruption that could ultimately lead to the reimagining of current educational practices that reinforce inequalities and, thus, toward more equitable approaches to promote student learning

(García-Morales et al., 2021; Iyengar, 2020). We hope that the present findings provide some understanding of the scope of the pandemic's immediate influence on student engagement, self-appraisal, and learning.

ORCID iDs

Teresa M. Ober [iD] https://orcid.org/0000-0001-9698-9543

Cheng Liu [iD] https://orcid.org/0000-0002-2787-1653

References

Ackerman, P. L., Kanfer, R., & Beier, M. E. (2013). Trait complex, cognitive ability, and domain knowledge predictors of baccalaureate success, STEM persistence, and gender differences. *Journal of Educational Psychology*, *105*(3), 911–927. https://doi.org/10.1037/a0032338

Adelson, J. L., McCoach, D. B., Rogers, H. J., Adelson, J. A., & Sauer, T. M. (2017). Developing and applying the propensity score to make causal inferences: Variable selection and stratification. *Frontiers in Psychology*, *8*, 1413. https://doi.org/10.3389/fpsyg.2017.01413

Bacher-Hicks, A., Goodman, J., & Mulhern, C. (2021). Inequality in household adaptation to schooling shocks: Covid-induced online learning engagement in real time. *Journal of Public Economics*, *193*, 104345. https://doi.org/10.1016/j.jpubeco.2020.104345

Bandura, A. (1994). Self-efficacy. In V. S. Ramachaudran (Ed.), *Encyclopedia of human behavior* (*vol. 4*, pp. 71–81). Academic Press.

Ben-Zvi, D., & Garfield, J. (2008). Introducing the emerging discipline of statistics education. *School Science and Mathematics*, *108*(8), 355–361. https://doi.org/10.1111/j.1949-8594.2008.tb17850.x

Bowers, A. J., & Zhou, X. (2019). Receiver operating characteristic (ROC) area under the curve (AUC): A diagnostic measure for evaluating the accuracy of predictors of education outcomes. *Journal of Education for Students Placed at Risk (JESPAR)*, *24*(1), 20–46. https://doi.org/10.1080/10824669.2018.1523734

Camfield, E. K., Schiller, N. R., & Land, K. M. (2021). Nipped in the bud: COVID-19 reveals the malleability of STEM student self-efficacy. *CBE—Life Sciences Education*, *20*(2), ar25. https://doi.org/10.1187/cbe.20-09-0206

Cham, H., & West, S. G. (2016). Propensity score analysis with missing data. *Psychological Methods*, *21*(3), 427–445. https://doi.org/10.1037/met0000076

Chen, C. Y. C., Byrne, E., & Vélez, T. (2021). Impact of the 2020 pandemic of COVID-19 on families with school-aged children in the United States: Roles of income level and race. *Journal of Family Issues*, *43*(3). https://doi.org/10.1177/0192513X21994153

Chiu, T. K. (2021). Applying the self-determination theory (SDT) to explain student engagement in online learning during the COVID-19 pandemic. *Journal of Research on Technology in Education*, 1–17. https://doi.org/10.1080/15391523.2021.1891998

Chu, S. (2020). *The revised AP Statistics exam for 2020*. The University Network. https://www.tun.com/blog/the-revised-ap-statistics-exam-for-2020/

CollegeBoard. (2010). *AP Statistics: Course and exam description*. https://apcentral.collegeboard.org/pdf/ap-statistics-course-and-exam-description.pdf

CollegeBoard. (2018). *National and state summary reports for 2018*. https://research.collegeboard.org/programs/ap/data/archived/ap-2018

CollegeBoard. (2019). *National and state summary reports for 2019*. https://research.collegeboard.org/programs/ap/data/archived/ap-2019

CollegeBoard. (2020a). *AP coronavirus updates: AP Statistics*. https://apcoronavirusupdates.collegeboard.org/educators/taking-the-exam/statistics

CollegeBoard. (2020b). *National and state summary reports for 2020*. https://research.collegeboard.org/programs/ap/data/participation/ap-2020

CollegeBoard. (2022a). *AP central: AP course audit*. https://apcentral.collegeboard.org/courses/ap-course-audit

CollegeBoard. (2022b). *AP central: AP Statistics*. https://apcentral.collegeboard.org/courses/ap-statistics/professional-learning

Daniel, J. (2020). Education and the COVID-19 pandemic. *Prospects*, *49*(1), 91–96. https://doi.org/10.1007/s11125-020-09464-3

Daniels, L. M., Goegan, L. D., & Parker, P. C. (2021). The impact of COVID-19 triggered changes to instruction and assessment on university students' self-reported motivation, engagement and perceptions. *Social Psychology of Education*, *24*(1), 299–318. https://doi.org/10.1007/s11218-021-09612-3

Darling-Aduana, J., Woodyard, H. T., Sass, T., & Barry, S. S. (2022). *Learning-mode choice, student engagement, and achievement growth during the COVID-19 pandemic*. National Center for the Analysis of Longitudinal Data in Education Research. https://doi.org/10.26300/jxcj-gs73

Dolan, J. E. (2016). Splicing the divide: A review of research on the evolving digital divide among K–12 students. *Journal of Research on Technology in Education*, *48*(1), 16–37. https://doi.org/10.1080/15391523.2015.1103147

Dorn, E., Hancock, B., Sarakatsannis, J., & Viruleg, E. (2020, December 8). *COVID-19 and learning loss—Disparities grow and students need help*. McKinsey and Company. https://www.mckinsey.com/industries/public-and-social-sector/our-insights

/covid-19-and-learning-loss-disparities-grow-and-students-need-help

Elena-Bucea, A., Cruz-Jesus, F., Oliveira, T., & Coelho, P. S. (2020). Assessing the role of age, education, gender and income on the digital divide: Evidence for the European Union. *Information Systems Frontiers*, 1–15. https://doi.org/10.1007/s10796-020-10012-9

Engzell, P., Frey, A., & Verhagen, M. D. (2021). Learning loss due to school closures during the COVID-19 pandemic. *Proceedings of the National Academy of Sciences*, *118*(17). https://doi.org/10.1073/pnas.2022376118

Federal Register. (2020). Declaring a national emergency concerning the novel coronavirus disease (COVID-19) outbreak. https://www.federalregister.gov/d/2020-05794

Fredricks, J. A. (2011). Engagement in school and out-of-school contexts: A multidimensional view of engagement. *Theory Into Practice*, *50*(4), 327–335. https://doi.org/10.1080/00405841.2011.607401

García-Morales, V. J., Garrido-Moreno, A., & Martín-Rojas, R. (2021). The transformation of higher education after the COVID disruption: Emerging challenges in an online learning scenario. *Frontiers in Psychology*, *12*, 196. https://doi.org/10.3389/fpsyg.2021.616059

Goldhaber, D., Kane, T., McEachin, A., Morton, E., Patterson, T., & Staiger, D. (2022). *The consequences of remote and hybrid instruction during the pandemic*. Center for Education Policy Research, Harvard University. https://cepr.harvard.edu/files/cepr/files/5-4.pdf?m=1651690491

Goodman, J. (2019). The labor of division: Returns to compulsory high school math coursework. *Journal of Labor Economics*, *37*(4), 1141–1182. https://doi.org/10.1086/703135

Gottfried, M. A., Bozick, R., & Srinivasan, S. V. (2014). Beyond academic math: The role of applied STEM course taking in high school. *Teachers College Record*, *116*(7), 1–35. https://doi.org/10.1177/016146811411600703

Goudeau, S., Sanrey, C., Stanczak, A., Manstead, A., & Darnon, C. (2021). Why lockdown and distance learning during the COVID-19 pandemic are likely to increase the social class achievement gap. *Nature Human Behaviour*, *5*(10), 1273–1281. https://doi.org/10.1038/s41562-021-01212-7

Harris, C., Straker, L., & Pollock, C. (2017). A socioeconomic related "digital divide" exists in how, not if, young people use computers. *PloS ONE*, *12*(3), e0175011.

Hart, C. M., Berger, D., Jacob, B., Loeb, S., & Hill, M. (2019). Online learning, offline outcomes: Online course taking and high school student performance. *AERA Open*, *5*(1), 2332858419832852. https://doi.org/10.1177/2332858419832852

Huck, C., & Zhang, J. (2021). Effects of the COVID-19 Pandemic on K–12 education: A systematic literature review. *New Waves-Educational Research and Development Journal*, *24*(1), 53–84. https://eric.ed.gov/?id=EJ1308731

Hughes, R. A., Heron, J., Sterne, J. A., & Tilling, K. (2019). Accounting for missing data in statistical analyses: multiple imputation is not always the answer. *International Journal of Epidemiology*, *48*(4), 1294–1304. https://doi.org/10.1093/ije/dyz032

Iyengar, R. (2020). Education as the path to a sustainable recovery from COVID-19. *Prospects*, *49*(1), 77–80. https://doi.org/10.1007/s11125-020-09488-9

Jankowski, N. A. (2020). Guideposts for assessment during COVID-19. *Assessment Update*, *32*(4), 10–11. https://doi.org/10.1002/au.30222

Judson, E., Bowers, N. L., & Glassmeyer, K. (2019). Recruiting and encouraging students to complete advanced placement science and math courses and exams. *Journal for the Education of the Gifted*, *42*(3), 243–265. https://doi.org/10.1177/0162353219855679

Korlat, S., Kollmayer, M., Holzer, J., Lüftenegger, M., Pelikan, E. R., Schober, B., & Spiel, C. (2021). Gender differences in digital learning during COVID-19: Competence beliefs, intrinsic value, learning engagement, and perceived teacher support. *Frontiers in Psychology*, 12, 637776. https://doi.org/10.3389/fpsyg.2021.637776

Kuhfeld, M., Soland, J., Tarasawa, B., Johnson, A., Ruzek, E., & Liu, J. (2020). Projecting the potential impact of COVID-19 school closures on academic achievement. *Educational Researcher*, *49*(8), 549–565. https://doi.org/10.3102/0013189X20965918

Lai, J., & Widmar, N. O. (2021). Revisiting the digital divide in the COVID-19 era. *Applied Economic Perspectives and Policy*, *43*(1), 458–464. https://doi.org/10.1002/aepp.13104

Lei, H., Cui, Y., & Zhou, W. (2018). Relationships between student engagement and academic achievement: A meta-analysis. *Social Behavior and Personality: An International Journal*, *46*(3), 517–528. https://doi.org/10.2224/sbp.7054

Lessard, L. M., & Puhl, R. M. (2021). Adolescent academic worries amid COVID-19 and perspectives on pandemic-related changes in teacher and peer relations. *School Psychology*, *36*(5), 285–292. https://doi.org/10.1037/spq0000443

Lewin, K. M. (2020). Contingent reflections on coronavirus and priorities for educational planning and development. *Prospects*, *49*, 17–24. https://doi.org/10.1007/s11125-020-09480-3

Madigan, D. J., & Curran, T. (2021). Does burnout affect academic achievement? A meta-analysis of over 100,000 students. *Educational Psychology Review*, *33*(2), 387–405. https://doi.org/10.1007/s10648-020-09533-1

Martin, E. G., & Sorensen, L. C. (2020, June). Protecting the health of vulnerable children and adolescents during COVID-19–related K–12 school closures in the US. *JAMA Health Forum*, *1*(6), e200724–e200724. https://doi.org/10.1001/jamahealthforum.2020.0724

McKune, S. L., Acosta, D., Diaz, N., Brittain, K., Joyce-Beaulieu, D., Maurelli, A. T., & Nelson, E. J. (2021). Psychosocial health of school-aged children during the initial COVID-19 safer-at-home school mandates in Florida: A cross-sectional study. *BMC Public Health*, *21*(1), 1–11. https://doi.org/10.1186/s12889-021-10540-2

National Assessment of Educational Progress. (2022a). *NAEP report card: 2022 NAEP mathematics assessment*. https://www.nationsreportcard.gov/highlights/mathematics/2022/

National Assessment of Educational Progress. (2022b). *NAEP report card: 2022 NAEP reading assessment*. https://www.nationsreportcard.gov/highlights/reading/2022/

National Science Foundation. (2019). *Women, minorities, and persons with disabilities in science and engineering*. National Center for Science and Engineering Statistics Report. https://ncses.nsf.gov/pubs/nsf19304

Nederhand, M. L., Tabbers, H. K., & Rikers, R. M. (2019). Learning to calibrate: Providing standards to improve calibration accuracy

for different performance levels. *Applied Cognitive Psychology*, *33*(6), 1068–1079. https://doi.org/10.1002/acp.3548

Ober, T. (2021, February 17). *Disruptiveness of COVID-19 on learning AP Statistics* [Pre-registration]. Open Science Framework. https://doi.org/10.17605/osf.io/7d9ac

Ober, T. M., Brodsky, J. E., Lodhi, A., & Brooks, P. J. (2021). How did introductory psychology students experience the transition to online instruction amidst the COVID-19 outbreak in New York City? *Scholarship of Teaching and Learning in Psychology*. Advance online publication.

Ober, T. M., Carter, M. F., Coggins, M. R., Filonczuk, A., Kim, C., Hong, M. R., & Cheng, Y. (2022). Adaptation to remote teaching during spring 2020 amidst COVID-19: Perspectives of AP Statistics teachers. *Computers in the Schools*, *39*(4), 342–372. https://doi.org/10.1080/07380569.2022.2090764

Ober, T., & Cheng, Y. (2022, December 16). *Disruptiveness of COVID-19: Differences in course engagement, self-appraisal, and learning* [Repository]. Open Science Framework. https://doi.org/10.17605/osf.io/t8zbd

Ober, T., Cheng, Y., Carter, M., & Liu, C. (2021, December 3). *Disruptiveness of COVID-19: Differences in course engagement, self-appraisal, and learning.* https://doi.org/10.31234/osf.io/b2pxd

Ober, T. M., Hong, M. R., Carter, M. F., Brodersen, A. S., Rebouças-Ju, D., Liu, C., & Cheng, Y. (2022). Are high school students accurate in predicting their AP exam scores? *Assessment in Education: Principles, Policy, and Practice*, 1–24. https://doi.org/10.1080/0969594X.2022.2037508

Ober, T. M., Hong, M. R., Rebouças-Ju, D., Carter, M. F., Liu, C., & Cheng, Y. (2021). Linking self-report and process data to performance as measured by different assessment types. *Computers and Education*, *167*, 104188. https://doi.org/10.1016/j.compedu.2021.104188

Pattison, K. L., Hoke, A. M., Schaefer, E. W., Alter, J., & Sekhar, D. L. (2021). National survey of school employees: COVID-19, school reopening, and student wellness. *Journal of School Health*, *91*(5), 376–383. https://doi.org/10.1111/josh.13010

Pokhrel, S., & Chhetri, R. (2021). A literature review on impact of COVID-19 pandemic on teaching and learning. *Higher Education for the Future*, *8*(1), 133–141. https://doi.org/10.1177/2347631120983481

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/.

Reimers, F. M., & Schleicher, A. (2020). *A framework to guide an education response to the COVID-19 pandemic of 2020*. OECD. https://oecd.dam-broadcast.com/pm_7379_126_126988-t63lxosohs.pdf

State of Indiana. (2020a). *Executive order 20-05: Helping Hoosiers during the public health emergency declared for the Coronavirus disease 2019 outbreak*. https://www.in.gov/gov/files/EO_20-05.pdf

State of Indiana. (2020b). *Executive order 20-16: Education matters and COVID-19*. https://www.in.gov/gov/files/Executive%20Order%2020-16%20Education.pdf

Stifel, S. W., Feinberg, D. K., Zhang, Y., Chan, M. K., & Wagle, R. (2020). Assessment during the COVID-19 Pandemic: Ethical, legal, and safety considerations moving forward. *School Psychology Review*, *49*(4), 438–452. https://doi.org/10.1080/2372966X.2020.1844549

Strauss, V. (May 15, 2020). College Board says new online AP tests are going well—but students report big problems. *Washington Post*. https://www.washingtonpost.com/education/2020/05/15/college-board-says-new-online-ap-tests-are-going-well-students-report-big-problems/

Sun, J. C. Y., & Rueda, R. (2012). Situational interest, computer self-efficacy and self-regulation: Their impact on student engagement in distance education. *British Journal of Educational Technology*, *43*(2), 191–204. https://doi.org/10.1111/j.1467-8535.2010.01157.x

Suzuki, H., Hong, M. R., Ober, T. M., & Cheng, Y. (2022). Prediction of differential performance between Advanced Placement exam scores and class grades using machine learning. *Frontiers in Education*, *7*. https://doi.org/10.3389/feduc.2022.1007779

Talsma, K., Schüz, B., Schwarzer, R., & Norris, K. (2018). I believe, therefore I achieve (and vice versa): A meta-analytic cross-lagged panel analysis of self-efficacy and academic performance. *Learning and Individual Differences*, *61*, 136–150. https://doi.org/10.1016/j.lindif.2017.11.015

Tannert, S., & Gröschner, A. (2021). Joy of distance learning? How student self-efficacy and emotions relate to social support and school environment. *European Educational Research Journal*, *20*(4), 498–519. https://doi.org/10.1177/14749041211024784

Tate, T., & Warschauer, M. (2022). Equity in online learning. *Educational Psychologist*, 1–15. https://doi.org/10.1080/00461520.2022.2062597

UNESCO. (2021). *Education: From disruption to recovery*. https://en.unesco.org/covid19/educationresponse

U.S. Code. (2011). Title 20 - Education, Chapter 28 - Higher Education Resources and Student Assistance, Subchapter III - Institutional Aid, Part E - Minority science and engineering improvement program, subpart 3 - administrative and general provisions, §1067k. Definitions. https://www.govinfo.gov/content/pkg/USCODE-2011-title20/html/USCODE-2011-title20-chap28-subchapIII-partE-subpart3-sec1067k.htm

Van de Werfhorst, H. G. (2021). Inequality in learning is a major concern after school closures. *Proceedings of the National Academy of Sciences*, *118*(20). https://doi.org/10.1073/pnas.2105243118

Van Dijk, J. A. (2017). Digital divide: Impact of access. *International Encyclopedia of Media Effects*, 1–11. https://doi.org/10.1002/9781118783764.wbieme0043

Van Lancker, W., & Parolin, Z. (2020). COVID-19, school closures, and child poverty: A social crisis in the making. *Lancet Public Health*, *5*(5), e243–e244. https://doi.org/10.1016/S2468-2667(20)30084-0

Wang, M. T., Chow, A., Hofkens, T., & Salmela-Aro, K. (2015). The trajectories of student emotional engagement and school burnout with academic and psychological development: Findings from Finnish adolescents. *Learning and Instruction*, *36*, 57–65. https://doi.org/10.1016/j.learninstruc.2014.11.004

Whitney, B. M., Cheng, Y., Brodersen, A. S., & Hong, M. R. (2019). The Scale of Student Engagement in Statistics: Development and initial validation. *Journal of Psychoeducational Assessment*, *37*(5), 553–565. https://doi.org/10.1177/0734282918769983

Wise, A. F. (2020). Educating data scientists and data literate citizens for a new generation of data. *Journal of the Learning Sciences*, *29*(1), 165–181. https://doi.org/10.1080/10508406.2019.1705678

Xu, D., Solanki, S., & Fink, J. (2021). College acceleration for all? Mapping racial gaps in Advanced Placement and dual enrollment participation. *American Educational Research Journal*, *58*(5), 954–992. https://doi.org/10.3102/0002831221991138

Yoon, H., Jang, Y., Vaughan, P. W., & Garcia, M. (2020). Older adults' Internet use for health information: Digital divide by race/ethnicity and socioeconomic status. *Journal of Applied Gerontology*, *39*(1), 105–110. https://doi.org/10.1177/0733464818770772

Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist*, *25*(1), 3–17. https://doi.org/10.1207/s15326985ep2501_2

Authors

TERESA M. OBER, PH.D., is an associate research scientist at Educational Testing Service (ETS) and a research affiliate collaborating with members of Learning Analytics and Measurement in Behavioral Sciences (LAMBS) Lab at the University of Notre Dame. Dr. Ober's work examines engagement and learning in technology-mediated environments and individual differences in applied educational contexts.

YING CHENG, PH.D., is a professor in the Department of Psychology, Fellow of the Institute for Educational Initiatives, associate director of the Lucy Family Institute for Data and Society, and director of the Learning Analytics and Measurement in Behavioral Sciences (LAMBS) lab at the University of Notre Dame. Dr. Cheng's background and expertise are in educational and psychological measurement, with a focus on item response theory, computerized adaptive testing, and cognitive diagnostic assessment.

MATTHEW F. CARTER is a former lab manager for the Learning Analytics and Measurement in Behavioral Science (LAMBS) Lab at the University of Notre Dame. Mr. Carter is currently pursuing research in biostatistics and public health.

CHENG LIU, PH.D., is the lead data scientist for the Center for Social Science Research (CSSR) in the Lucy Family Institute for Data & Society and concurrent research assistant professor of psychology at the University of Notre Dame. Dr. Liu 's experience in quantitative research and the infrastructure development of online education platforms, his research interests include educational data mining and education technology.