



Reassessing Disparities in Online Learner Student Engagement in Higher Education

Justin Paulsen¹ and Alexander C. McCormick¹

Online learning is the fastest growing segment in U.S. higher education and is increasingly adopted in public and private not-for-profit institutions. While the impact of online learning on educational outcomes is becoming more clear, the literature on its connection with student engagement is sparse. Student engagement measures identify key aspects of the learning process that can improve learning and outcomes like retention and achievement. The few studies investigating the link between online learning and student engagement found positive benefits for online learners compared to face-to-face learners in terms of perceived academic challenge, learning gains, satisfaction, and better study habits. On the other hand, face-to-face learners reported higher levels of environment support, collaborative learning, and faculty interaction. However, these studies did not effectively account for the differences in background characteristics like age, time spent working or caring for dependents, and enrollment status. Further, they did not consider the increasingly large population of students who enroll in both online and face-to-face courses. In our study, we used propensity score matching on the 2015 National Survey of Student Engagement data to account for the disparities in these groups' demographics variables. After matching, we found that some of the previous literature's differences diminish or disappear entirely. This suggests differences in supportive environments and learning strategies have more to do with online student characteristics than learning mode. However, online learning still falls well below other modes in terms of collaborative learning and interaction with faculty.

Keywords: computers and learning; higher education; learning environments; online learning; quasi-experimental analysis; student engagement

Technological innovation and increasing broadband connectivity in the United States have rendered online learning accessible to broader segments of the population. In U.S. higher education, online learning has grown faster than overall enrollment, with recent growth faster among public and private not-for-profit institutions than at private for-profits (Allen & Seaman, 2016). According to the 2017 *Digest of Education Statistics*, in fall 2014, roughly 5.2 million U.S. undergraduates (31% of the total) were enrolled in at least some “distance education” course (nearly all of which is delivered online), including about 2.2 million (13%) who participated exclusively at a distance. The percentages at 2- and 4-year institutions were nearly identical (U.S. Department of Education, 2018).

Empirical evidence of online learning's efficacy is contested. One meta-analysis suggested blended learning (combining face-to-face and online delivery) is associated with the highest learning gains followed by entirely online learning, but critics of the

study suggested those results are not generalizable (Jaggars & Bailey, 2010; Means et al., 2009). Many studies comparing individual classes across online, blended, and face-to-face modalities showed essentially equivalent results in cognitive gains and mixed results in behavioral and emotional engagement (see Cosgrove & Olitsky, 2015; Garratt-Reed et al., 2016; Reece & Butler, 2017; Tseng & Walsh, 2016). Research on massive open online courses (MOOCs) suggests that online learning participation decreases steeply after courses begin (Perna et al., 2014). However, MOOCs represent a special case of online delivery with few enrollment restrictions (Allen & Seaman, 2016), whereas the present study examined online course-taking by bachelor's degree-seeking undergraduates. Especially relevant to our investigation is Jaggars and Xu's (2016) finding that the quality of interactions in online courses, and particularly the

¹Indiana University, Bloomington, IN

frequency and effectiveness of student-faculty interaction, was associated with higher academic performance. The present study considers the effects of online learning on student engagement—that is, participation in empirically supported effective educational practices.

Motivation for the Study

In recent decades, student engagement has emerged as a fruitful framework for understanding the efficacy of students' educational experiences in college (Kuh, 2003; Kuh et al., 2008; McCormick et al., 2013). Integrating elements from Tinto's (1975, 1986) student departure theory, Pace's (1980, 1982) work on the quality of student effort, Astin's (1984) student involvement theory, and Pascarella's (1985) general causal model of college environmental effects, the student engagement perspective builds on decades of research findings about activities, experiences, and environmental features related to desired learning outcomes (McCormick et al., 2013). The student engagement perspective rests on some simple propositions:

- Student learning is related to the time and effort students devote to their studies.
- Students benefit from a collegiate environment that promotes and supports their success.
- Colleges, universities, and individual faculty members can and should promote student success by emphasizing empirically supported effective educational practices in and outside the classroom.

Although the conceptual foundations of student engagement predate online learning (McCormick et al., 2013, documented these conceptual foundations), the key elements of student engagement such as active learning, peer collaboration, and interaction with faculty neither assume nor require face-to-face instruction. For most facets of engagement, there is no inherent reason to expect differences related to course delivery modality. Yet equivalence should not be assumed, and efforts to assess the educational effectiveness of online learning should investigate the impact of delivery modality on student engagement.

Despite the broad acceptance of student engagement and the potentially unique interactions between online learning and student engagement, there has been scant rigorous comparison of student engagement in online and face-to-face contexts. Coates (2007) adapted student engagement in the Australian context to formulate a typology of engagement styles using a survey including items tailored to the online experience, but the intent of the study was not to contrast online learners (OLs) and face-to-face learners (FFLs). Similarly, the community of inquiry framework (Garrison et al., 2010) incorporates engagement constructs and has been used to examine online and blended learning arrangements, but it does not afford comparison to traditional face-to-face instruction.

The National Survey of Student Engagement (NSSE) distills survey results for first-year and senior undergraduates into 10 "engagement indicators." As assessed by NSSE, student engagement has been linked to important outcomes, including student achievement (as measured by GPA; Kuh et al., 2007; Pascarella &

Terenzini, 2005), persistence in college (Fredricks et al., 2004; Hu & McCormick, 2012; Kuh, 2003; Kuh et al., 2008), perceived and directly assessed learning gains (Hu & McCormick, 2012), and satisfaction (Filak & Sheldon, 2008). At the institutional level, NSSE indicators are associated with gains in outcomes such as critical thinking and problem solving, moral reasoning, inclination to inquire, and intercultural effectiveness (Pascarella et al., 2010). These findings further buttress the argument to rigorously investigate the engagement consequences of online learning.

Given NSSE's scale (500–700 four-year institutions annually), the diversity of institutions represented (NSSE, 2015b), and the broad applicability of most survey content to online and face-to-face delivery, the data uniquely offer the opportunity to examine engagement consequences of online delivery. While the validity of student surveys in general and NSSE specifically has been challenged (Porter, 2011), leaders of NSSE and its counterpart, the Community College Survey of Student Engagement, offered a point-by-point response to the critique (McCormick & McClenney, 2012), and 2- and 4-year institutions continue to participate in the projects in large numbers.

NSSE data have been used to examine average differences in engagement between OLs and FFLs net of individual and institutional characteristics (Chen et al., 2008; Chen et al., 2009; NSSE, 2013; Rabe-Kemp et al., 2009; Robinson & Hullinger, 2008). These studies found that compared to FFLs, OLs exhibit higher levels of academic challenge, more frequent use of effective learning strategies, greater perceived gains and satisfaction with their education, but lower levels of collaborative learning. Robinson and Hullinger (2008) found OLs scored higher than FFLs on NSSE measures of student-faculty interaction and enriching educational experiences. Chen et al. (2008) and Rabe-Hemp et al. (2009) found OLs spent more time preparing for class, participated more in class discussions,¹ and interacted more with professors but less with peers. A 2013 NSSE report found that on average, OLs reported higher levels of course challenge, made greater use of effective learning strategies, and rated their quality of interactions with others (peers, advisors, etc.) more highly than FFLs. However, OLs had lower average levels of collaborative learning and interactions with faculty compared to FFLs (in contrast to earlier research cited previously) and averaged lower scores on the supportive environment indicator. Chen et al. (2009) used multivariate techniques to find that instructional modality had a small but significant positive relationship with every student engagement indicator.²

While course-level studies of online learning cited earlier give insight into what individual teachers can do to improve student outcomes under different modalities, they offer little guidance to institutional decision makers with regard to the consequences of expanding online learning. The NSSE-based studies that examined engagement across all courses suggest that OLs are more engaged on indicators that should be unrelated to modality (e.g., course challenge, study strategies, and quality of interactions) while less engaged relative to FFLs on indicators that are facilitated by in-person interaction (e.g., collaborative learning and student-faculty interaction).

To date, the studies examining student engagement in online learning have not adequately accounted for systematic differences between the OL and FFL populations. As intuition

Table 1
Percentage Distribution of Undergraduates at 4-Year Institutions According to Course Modality, by Selected Student Characteristics, 2016

	Face-to-Face Courses Only	At Least One Online Course
Total	55.7	44.3
Age		
22 or younger	66.1	33.9
23–30	45.8	54.2
31 or older	33.2	66.8
Parents' highest level of education		
Less than a bachelor's degree	50.7	49.3
Bachelor's degree or higher	60.4	39.6
Single parent status		
Not a single parent	57.9	42.1
Single parent	36.1	63.9
Enrollment pattern for 2015–2016		
Exclusively full-time	61.6	38.4
Mixed full- and part-time	44.0	56.0
Exclusively part-time	51.9	48.1
Employment while enrolled		
Not employed	63.8	36.2
Worked 1–29 hours per week	61.1	38.9
Worked 30 or more hours per week	37.5	62.5

Source. U.S. Department of Education, National Center for Education Statistics (NCES), 2015–2016 National Postsecondary Student Aid Study (NPSAS:16; authors' calculations using NCES PowerStats online application).

suggests and national data confirm, OLs differ notably from FFLs with respect to background and enrollment characteristics (Table 1). Yet *t* tests do not account for covariates, and regression cannot account for bias when treatment and control covariates differ markedly (d'Agostino, 1998), as in the case of online learning. Consequently, there is a need to apply methods that better disentangle effects of instructional modality from large compositional differences in the two student populations.

Finally, previous studies have not examined dual-mode learners (DMLs)—those taking both online and face-to-face courses. The DML group has steadily grown over the past several years as brick-and-mortar institutions have expanded online offerings. As of 2016, more than a quarter of higher education students reported taking at least one online course (Allen & Seaman, 2016). Understanding the experience of DMLs is increasingly important as this trend is expected to continue.

Building on the extant literature, this study advances our understanding of the relationship between online learning and student engagement in higher education by (a) using methods that more effectively account for group differences and (b) extending the analysis to include the growing population of students who take both online and face-to-face courses.

Data

We used the 2015 NSSE data set with results from 541 U.S. institutions spanning the range of institutional diversity (NSSE, 2015b). We defined three groups based on survey questions about the number of current-term courses and the number identified as “entirely online”: those taking all of their courses face to

face (FFLs), those taking all of their courses online (OLs), and those taking both online and face-to-face courses (DMLs). We limited DMLs to students who took 40% to 80% of their classes online to afford a reasonable proportion of online classes. The dual-mode group does not refer to classes that blended face-to-face and online modalities (NSSE's survey design does not differentiate such courses). After removing cases with missing data, the sample included 122,347 FFLs, 11,344 OLs, and 7,081 DMLs. These groups included first-year and senior students because preliminary analyses found similar results across groups.

Congruent with previous literature, we found that relative to FFLs and DMLs, OLs were more likely to have enrolled part-time; enrolled in private for-profit institutions; transferred; majored in business, health professions, and “other” (i.e., unclassified) majors; lived away from campus; had higher grades; and had parents who hold less than a bachelor's degree (see Appendix Table C1 available on the journal website). Additionally, OLs were older and worked more than FFLs and DMLs, on average. We also found that DMLs differed from FFLs in many of the same ways as OLs but to a lesser degree. For example, the average age was 22.3 for FFLs and 37.1 for OLs, with DMLs in the middle at 27.3.

We used the following NSSE engagement indicators as outcomes because of their use in previous literature (Chen et al., 2008; Chen et al., 2009; Rabe-Hemp et al., 2009; Robinson & Hullinger, 2008) and their importance to student learning and success: collaborative learning, quality of interactions, supportive environment, student-faculty interaction, and learning strategies. We also included comparisons for two additional indicators not previously analyzed by modality, higher order

learning and reflective and integrative learning. (See Appendix A available on the journal website for items that make up each scale.) Collaborative learning taps the frequency with which students engage in learning activities with peers. Quality of interactions reflects student ratings of interactions with others in various roles (e.g., peers, faculty, etc.). Supportive environment represents students' perceptions of the institution's emphasis on programs and activities supporting student learning and success, including some campus-based activities that may inherently disadvantage OLs. Student-faculty interaction includes items assessing frequency of interactions with faculty (e.g., talking about course topics or career plans or discussing academic performance). Learning strategies captures the frequency with which students engaged in selected learning-enhancement practices. Higher order learning assesses the extent to which coursework emphasizes sophisticated cognitive tasks including application, analysis, synthesis, and evaluation. Reflective and integrative learning gauges activities like integration of learning across courses, perspective-taking, and reassessment of personal views. The engagement indicators have acceptable psychometric qualities, including Cronbach's alphas of .78 to .89 (NSSE, 2015a).

Methods

We account for the substantially different demographic characteristics of the groups using propensity score matching (PSM). PSM identifies cases that are essentially equivalent on observable variables but differ in treatment status (i.e., educational delivery modality). Differences that remain after these matches provide a more accurate estimate of the treatment effect than regression methods (Rosenbaum & Rubin, 1983). It thus increases the validity of a causal argument that the treatment of interest is responsible for outcome differences.

First, propensity scores are estimated using a logistic regression to predict a subject's propensity for treatment. We selected a logit model because of its broad familiarity and similar results achieved with logit or probit models (Caliendo & Kopeinig, 2008). The logistic regression produces a single, counterfactual-identifying data point, the propensity score, leveraging the model's covariates to balance on the distribution of these observed covariates, largely eliminating bias stemming from selection (Rosenbaum & Rubin, 1983).

Depending on the comparison in our study (OL vs. FFL, OL vs. DML, DML vs. FFL), the "treatment" is either online or dual modality. A key step in a PSM study is identifying all possible variables that could affect the relationship between the treatment and the outcome. Based on the extant literature (Carnoy et al., 2012; Schuetze & Slowey, 2002) as well as other variables feasibly related to the treatment and/or outcome that are available in the NSSE data set, we used the following: age, gender, race/ethnicity, parents' education, hours per week working for pay, enrollment status, type of living arrangement, care for dependents, institution control (public, private, private for-profit), major, transfer status, and grades at current institution. While we feel confident that this list captures much of what differentiates the three groups, one can never achieve complete certainty, so the risk of unobserved confounding factors may remain.

Next, we selected a caliper matching technique with radius matching within 0.03 (Caliendo & Kopeinig, 2008). This process limits matching to subjects in each group within a small propensity score bandwidth. The procedure then matches with every observation within the determined bandwidth. We selected a relatively small radius to decrease bias in estimates. While this approach is associated with increased variance, our large sample diminishes this impact. To demonstrate robustness, we also used nearest neighbor matching with three neighbors with replacement as well as kernel matching. As Appendix Tables B1 through B3 (available on the journal website) demonstrate, these different matching techniques produced very similar estimates.

The caliper matching method also effectively deals with the unequally distributed common support found in these comparisons (Stuart, 2010). Figure 1 shows the common support (i.e., the range of propensity scores for each treatment group), allowing for evaluation of the matching procedure. While the comparison of all OLs and FFLs shows all observations on the common support, the density of propensity scores above 0.5 among FFLs when compared to that of OLs is sparse. However, due to the large sample size for FFLs ($N = 122,347$), there are more than 20,000 FFLs with propensity scores above 0.5 from which to draw matches. The OL-DML comparison also shows all observations on the common support, with slightly more balance than the OL-FFL comparison. The comparison of the DMLs and FFLs also shows all observations on the common support and a more equal balance. Thus, we are confident that the matching provides meaningful comparisons at every level of propensity score for all comparisons.

Our results are based on the average effect of treatment on the treated (ATT), as opposed to the average treatment effect (ATE). To illustrate the difference, we slightly modify an excerpt from Morgan and Winship (2014) to fit our example:

The ATE is the expected what-if difference in [engagement] that would be observed if we could educate a randomly selected student in both [an all online and an all face-to-face sphere]. In contrast, the ATT is the expected what-if difference in [engagement] that would be observed if we could educate a randomly selected [all online] student in both [an all online and in an all face-to-face sphere]. (p. 55)

Both types of treatment effects tell an interesting story. Focusing on ATE can reveal the impact of different modalities on the random student. Focusing on ATT can indicate the impact of different modalities on the random all-online student. We find the latter to be more important because the students who elect to be all-online students are markedly different in most observable characteristics. ATE is valuable in its singular focus on the effect of the modality; however, by considering the treatment effect in terms of population characteristics that are much less associated with all online learning, it limits the practical implications for the current state of online learning in higher education. Additionally, the ATT addresses the challenge of unequal propensity score distributions because this estimate only requires a single observation with the comparable propensity score (Stuart, 2010). However, we recognize the value of the ATE estimates in understanding the effect of different modalities based on more

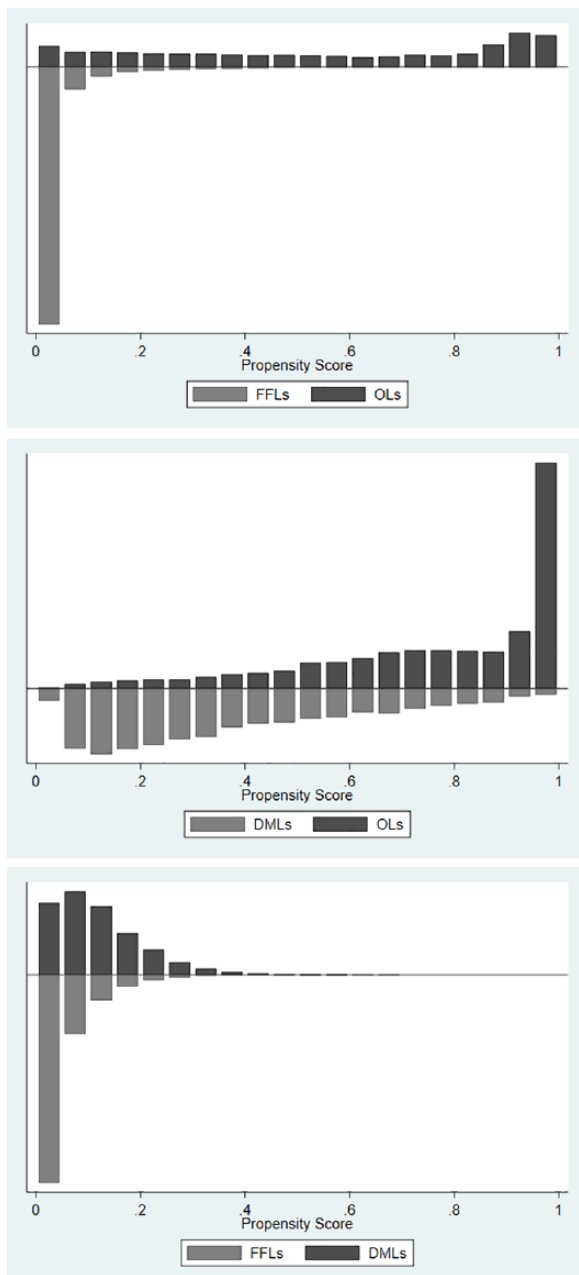


FIGURE 1. *Distribution of propensity scores by treatment status.*
 Note. FFLs = face-to-face learners; OLs = online learners;
 DMLs = dual-mode learners.

general student population characteristics and provide ATE estimates in Appendix B, available on the journal website.

One key assumption of PSM is that covariates included in the model entirely account for the selection process governing the treatment and control group assignments. Our modeling accounts for covariates identified by previous research as key characteristics differentiating and accounting for grouping. However, without random assignment, there is no certain way to ensure that confounding variables do not account for the findings. For example, Table 1 shows that single parenthood is correlated with learning modality, but it is unavailable in the NSSE data. We instead used time spent caring for dependents as an

imperfect proxy for single parenthood, but this represents a possible threat to our findings. To increase confidence in our findings, we conducted sensitivity analyses examining the robustness of our findings to possible unobserved confounds. Rosenbaum (2002) developed a method of sensitivity analysis that estimates different possible strengths (gamma factor) of an unobserved confounding binary variable that allow the analyst to determine at what strength the conclusions identified would change. These hypothetical gammas can be interpreted as the odds ratio of the coefficient of the unobserved confound (e.g., a gamma of 1.5 indicates that the unobserved confound increases the odds of being an OL compared to being an FFL 1.5 times holding other covariates constant). These values can be compared against odds ratio coefficients in the logistic regression used to estimate the model and expert judgment. Gamma values larger than those in the logistic regression suggest that the confound would need to be much more influential than covariates included in the model. On the other hand, gamma values similar to or smaller than those in the regression suggest the findings may be less robust to potential confounding variables if there is reason to believe such confounds exist.

Limitations

Despite the strengths of the NSSE data set and the PSM methodological approach, our study has limitations. First, NSSE is only administered at 4-year institutions. Findings may not generalize to the experiences of community college students. Second, NSSE is not designed to be a nationally representative sample. Nevertheless, the project has documented the extent to which its institutional and student samples resemble the national distribution of 4-year institutions and students (NSSE, 2015b). Third, NSSE is only administered to first-year and senior undergraduates. While we do not believe the inclusion of other undergraduates would meaningfully alter our conclusions, we cannot rule the possibility out. However, generalization to graduate students would be highly speculative. Fourth, NSSE's engagement measures are not course-specific. Rather, the survey calls on students to characterize their experience "during the current school year." Neither course design nor instructor qualifications are taken into account. Fifth, as noted previously, PSM depends on the inclusion of all influential covariates. To this end, we included those covariates suggested by prior literature from the NSSE data set and conducted sensitivity analyses. However, recent research indicated that course design and the professor's involvement with student learning (through assignment feedback and responsiveness to student communication) are important variables associated with the engagement and success of online students (Cho & Tobias, 2016; Jaggars & Xu, 2016). The NSSE data set includes an engagement indicator for effective teaching practices that is similar to the variables identified by Cho and Tobias (2016); however, they are aggregated across courses and would thus obscure impacts of individual courses and professors, a limitation that could be particularly problematic for DMLs. Finally, the treatment assignment is based on courses taken during spring semester, whereas NSSE's engagement indicators represent experiences over the academic year. Thus, it is possible that the

Table 2
Bias Reduction Statistics

Mode Comparison	Matching Status	Pseudo R^2	LR χ^2	p (χ^2)	Mean Bias (%)
OLs v. FFLs	Unmatched	.531	41,263.21	.0001	31.1
	Matched	.018	573.61	.0001	3.6
OLs v. DMLs	Unmatched	.347	8,506.70	.0001	17.9
	Matched	.027	855.55	.0001	4.1
DMLs v. FFLs	Unmatched	.139	7,629.90	.0001	14.1
	Matched	.005	101.34	.0001	1.4

Note. OLs = online learners; FFLs = face-to-face learners; DMLs = dual-mode learners; LR = likelihood ratio.

sample includes students with unobserved changes in treatment status between fall and spring semesters.

Findings

PSM diminished bias and increased balance of covariates across the comparisons. To assess matching quality, we report three measures of bias reduction in Table 2: pseudo R^2 (a measure of the extent to which regressors predict participation), a joint significance test of all regressors, and a measure of mean bias (“difference of sample means as a percentage of the square root of the average sample variances in both groups”; Caliendo & Kopeinig, 2008, p.15). An acceptable match would see pseudo R^2 approach 0, the mean bias be under 5%, and the joint significance test be nonsignificant (Kopeinig & Caliendo, 2008). As Table 2 shows, Pseudo R^2 reduced considerably for all comparisons to less than .03. Mean bias decreased to below 5% in each case. Both of these measures suggest matching significantly reduced the bias. The joint significance tests (noted χ^2 in Table 2), however, were all statistically significant, suggesting that across all variables, the groups differed. This likely reflects extremely large sample sizes given that joint significance tests are sensitive to sample size. Similarly, as shown in Appendix Tables C3, C5, and C7 available on the journal website, the few covariates with statistically significant differences showed only small differences after matching, suggesting the statistical significance is likely reflective of the large sample sizes. Overall, these measures suggest that the matching procedure effectively balanced the groups’ observed covariates.

A unique benefit of PSM is that results showing a change in the modality effect after matching suggest differences associated with the treatment (see Table 3). Matching includes only observations across modalities that are similar in terms of the variables in the logistic regression (i.e., the key demographic and enrollment variables differentiating OLs and FFLs). Any difference that remains after matching on these covariates likely indicates a disparity in terms of modality or unobserved causes not in the propensity model. An outcome measure that differs before matching but not after matching suggests the unmatched difference is related to covariates rather than modality. Changes in effect between unmatched and matched comparisons suggest a covariate-driven difference, whereas differences that persist after matching suggests a genuine modality-driven difference (more detailed results are given in Appendix B available on the journal website).

Collaborative Learning

Previous studies found that collaborative learning is higher among FFLs than OLs. PSM reduced the estimated difference in collaborative learning between OLs and FFLs and between OLs and DMLs (−13.90 unmatched, −9.22 PSM and −10.38 unmatched, −10.06 PSM, respectively), but differences after matching remain large (NSSE engagement indicators are scored on a 60-point scale.) Interestingly, a smaller but still statistically significant difference (−3.51 unmatched, −2.41 PSM) also favored FFLs over DMLs. Thus, every comparison favors the group that has more face-to-face contact even after matching. Despite technological advances and efforts to design effective ways to connect in online learning environments, students in online courses have fewer opportunities to benefit from peer collaboration.

Quality of Interactions

Previous literature suggested OLs rated quality of interactions higher than FFLs. Before matching, there is a significant and large difference favoring OLs compared to FFLs, but this decreases by more than half after matching (3.91 unmatched, 1.53 PSM). A similar change is seen in comparing OLs and DMLs (4.05 unmatched, 1.98 PSM), whereas there is no statistical difference between DMLs and FFLs (−0.14 unmatched, −0.30 PSM). This suggests that the online modality provides certain affordances in interacting with others that FFLs and DMLs do not perceive. One possible explanation may be that digital interactions provide greater efficiencies than analogous in-person interactions. Another potential explanation is that individualized interaction—even if technologically mediated—carries greater positive value for distance learning students.

Supportive Environment

Prior studies found FFLs experienced a more supportive environment than OLs. This makes sense given that some of the scale’s constituent items refer to activities that imply physical presence. PSM proves particularly valuable here as the OLs-FFLs difference disappears after matching (−4.84 unmatched, −0.17 PSM). A similar pattern is seen for OLs and DMLs (−3.04 unmatched, 0.22 PSM) and DMLs and FFLs (−1.81 unmatched, −0.26 PSM). Thus, among a comparable group of FFLs, DMLs, and OLs, the perception of supportive environment is essentially

Table 3
Findings Compared to Previous Literature and PSM Results

	Previous Literature: OLs vs. FFLs	OLs vs. FFLs		OLs vs. DMLs		DMLs vs. FFLs	
		Unmatched	Matched	Unmatched	Matched	Unmatched	Matched
Collaborative learning	–	–	–	–	–	–	–
Quality of interactions	+	+	+	+	+		
Supportive environment	–	–		–		–	
Student-faculty interaction	+/-	–	–	–	–	–	–
Learning strategies	+	+		+		+	
Higher order learning	N/A	+	+	+		+	
Reflective and integrative learning	N/A	+	+	+		+	

Note. – indicates the second group is larger than first group (e.g., FFLs higher than OLs); + is the opposite (e.g., OLs higher than FFLs). Empty cells indicate a nonsignificant difference. PSM = propensity score matching; OLs = online learners; FFLs = face-to-face learners; DMLs = dual-mode learners.

equal. This suggests that modality is less of a concern in providing an environment supportive of student success compared to individual differences in characteristics like age, residence, part-time status, and/or obligations to work or dependents.

Student-Faculty Interaction

The previous literature showed mixed results, with some studies finding OLs to have more frequent student-faculty interaction than FFLs and others finding the opposite. This study found that both FFLs and DMLs rated student-faculty interactions much higher than OLs even after matching (–10.22 unmatched, –4.61 PSM and –8.52 unmatched, –6.77 PSM, respectively). The difference between DMLs and FFLs is diminished after matching (–1.70 unmatched, –0.98 PSM). This finding is similar to the collaborative learning finding as every comparison favors the group with more opportunities for face-to-face interaction. This suggests face-to-face learning provides opportunities to engage with faculty that the online sphere does not provide. It is easy to imagine how sharing a physical space can facilitate interaction before or after class and even beyond the classroom. These findings should be of particular concern as a portion of the undergraduate population is significantly disadvantaged in this educationally beneficial aspect of college.

Learning Strategies

Previous studies indicated OLs engage in key learning strategies more frequently than FFLs. The initial difference substantially favors OLs over FFLs and DMLs, but matching reduces this to nonsignificance (5.13 unmatched, 0.19 PSM; 3.29 unmatched, –0.72 PSM). Matching also eliminates the difference between DMLs and FFLs (1.85 unmatched, 0.28 PSM). Thus, the difference in learning strategies reported in previous studies should be ascribed to underlying demographic and enrollment differences between the groups (e.g., age, residence, part-time status, and/or time spent working and caring for dependents) rather than the learning modality. This seems plausible because older students may bring greater commitment and life experience to their studies relative to their younger counterparts and as returning students, they may be especially attentive to maximizing their chances of success.

Higher Order Learning

We expand on previous literature by analyzing the relationship between modality and students' reports about their courses' emphasis on higher order learning. Both this construct and reflective and integrative learning (see below) assess processes associated with deep learning that should not differ by modality. Our findings suggest that even after matching, OLs had higher scores for higher order learning, although matching diminished the magnitude of the effect (3.57 unmatched, 1.54 matched). No statistically significant differences remained between OLs and DMLs after matching (3.23 unmatched, 0.68 matched), and DMLs and FFLs were essentially equal before and after matching (0.34 matched, –0.22 unmatched). This finding suggests that the online modality may lead students to engage in somewhat more challenging cognitive tasks than face-to-face learning.

Reflective and Integrative Learning

We also investigated how modality influences reflective and integrative learning. The findings for this construct resemble those for higher order learning. After matching, OLs had a significantly higher score on this measure compared to FFLs, but the magnitude of the effect was much reduced (3.04 unmatched, 0.96 matched). Initial unmatched differences do not hold when comparing matched OLs and DMLs (1.86 unmatched, –0.30 matched) or matched DMLs and FFLs (1.19 unmatched, 0.21 matched). While the difference between OLs and FFLs after matching is small, the fact that it remains after matching suggests there may be unique ways that online courses encourage OLs to draw connections between classes, connect classes with real-world issues, and reexamine prior beliefs.

Sensitivity Analysis

We conducted sensitivity analyses on each of the estimated non-zero differences between groups to see what strength of omitted confounding variables (gamma level) would be required to diminish the estimated effect to zero. Table 4 shows the gamma level at which the statistical significance of the estimated difference would reach the threshold of $p < .05$. As noted in the

Table 4
Sensitivity Analysis of Statistically Different Comparisons

Group Comparison	Outcome (Difference)	Gamma Level	Coefficients With Larger Odds Ratios (out of 53)
OLs vs. FFLs	Collaborative learning (-9.22)	3.8	5
OLs vs. FFLs	Quality of interactions (1.53)	1.4	17
OLs vs. FFLs	Student-faculty interaction (-4.61)	2.2	10
OLs vs. FFLs	Higher order learning (1.54)	1.3	20
OLs vs. FFLs	Reflective and integrative learning (0.96)	1.2	22
OLs vs. DMLs	Collaborative learning (-10.06)	4.0	2
OLs vs. DMLs	Quality of interactions (1.98)	1.6	7
OLs vs. DMLs	Student-faculty interaction (-6.77)	2.9	2
DMLs vs. FFLs	Collaborative learning (-2.41)	1.4	11
DMLs vs. FFLs	Student-faculty interaction (-0.98)	1.2	13

Note. OLs = online learners; FFLs = face-to-face learners; DMLs = dual-mode learners.

methods section, gamma levels are considered relative to the odds ratios from the logistic regressions used to estimate the propensity scores. We consider a finding to be robust to omitted covariates if many of the odds ratios in the logistic regression are smaller than the estimated gamma level. The largest odds ratios across the models are associated with variables noted previously that strongly differentiate the groups. As indicated in Table 4, all gamma levels are larger than more than half of the odds ratios associated with variables included in the models, suggesting that any hidden bias would have to be substantial to alter our findings.

Discussion

Consistent with previous work (NSSE, 2013), we found that OLs experienced substantially lower levels of collaborative learning than FFLs, albeit at a reduced magnitude after matching. Quality of interactions was another area in which our study supported prior findings favoring FFLs, but at a diminished magnitude. As online learning continues to expand, it will be important to implement approaches that promote opportunities for students to have meaningful interactions with peers and afford richer opportunities for collaborative learning. Whereas previous literature had mixed results with respect to student-faculty interaction, our results found this to be a vulnerability of online learning. The importance of collaboration and both student-student and student-instructor interaction in online learning is well documented (see the review of relevant literature in Jagers & Xu, 2016). As with quality of interactions more generally, the future development of online learning should identify ways to facilitate meaningful interactions with instructors. Jagers and Xu's (2016) mixed-methods study of 23 online courses is instructive in this regard, revealing both the characteristics of student-student and student-instructor interaction in courses highly rated for quality of interactions. Yet the examples revealed a heavy reliance on discussion boards as the mechanism for interaction, despite longstanding calls to move from "discussion" to "discourse" to facilitate student learning (Anderson, 2008). In this view, efforts to promote collaboration and interaction may need to go beyond the conventional affordances and implementations.

In contrast with our findings for collaboration and interaction, PSM results suggest that apparent differences in supportive environment and learning strategies reported in previous studies do not appear to reflect distinctive features of online learning. Rather, they reflect distinctly different OL population characteristics such as age, residence, enrollment status, major, responsibilities for dependents, and work commitments. These results suggest that online learning does not appear to be a barrier to institutional support for learners—the differences favoring FFLs have more to do with who OLs are than how their courses are delivered. The supportive environment finding bears deeper investigation to ensure colleges and universities meet unique needs of nontraditional students across all modalities in ways that take account of their life circumstances. A key focus of these efforts should be the already documented need to better support older, part-time, and commuting students who must balance their academic responsibilities with other obligations such as work and family. Potential accommodations include after-hours access to support services, flexible scheduling of exams and office hours, adequate parking facilities for commuting students, secure storage for nonresident students, and day-care services for students with dependents.

In a set of novel analyses, we investigated two measures of deep approaches to learning—higher order learning and reflective and integrative learning. In both cases, our results found *higher* scores for OLs relative to FFLs, though matching substantially reduced the magnitude of differences. These findings beg further investigation to identify the affordances of online instruction that facilitate richer cognitive engagement than traditional classroom instruction.


Another contribution of this study is examining the experience of students taking both online and face-to-face courses (DMLs). PSM had meaningful impact on comparisons involving DMLs. Comparing OLs with DMLs, the number of statistically significant differences in outcomes dropped from eight before matching to three after matching. For DML-FFL comparisons, significant differences dropped from seven to two. Although this study restricted the DML group to those with 40% to 80% of courses taken online, we found DMLs more closely resemble FFLs for most outcomes but in several cases


demonstrated engagement levels between FFLs and OLs. For example, DMLs experienced lower levels of collaborative learning than FFLs but to a much lesser extent than OLs. Similarly, DMLs experienced lower levels of student-faculty interaction compared with FFLs, but the magnitude of the difference was much lower than for OLs. Additionally, DMLs generally differed from OLs in the same way as FFLs but to a lesser extent (i.e., higher collaborative learning, student-faculty interaction; lower quality of interactions; similar supportive environment and learning strategies).³

In summary, some comparisons between FFLs and OLs were consistent with differences found in previous literature, but the magnitude of the differences appears to have been overstated by the failure to adequately account for the distinctiveness of the OL population. Previous literature suggested online learning was associated with both beneficial engagement outcomes (learning strategies, quality of interactions, time spent preparing for class) and drawbacks (collaborative learning, supportive environment) compared to face-to-face instruction (Chen et al., 2008; Chen et al., 2009; NSSE, 2006, 2013; Rabe-Hemp et al., 2009; Robinson & Hullinger, 2008). However, these findings were based on methodological approaches that did not adequately compensate for considerable demographic differences between OLs and FFLs. This study used PSM to compensate for substantial compositional differences in the FFL, OL, and DML populations, increasing confidence in estimates of how instructional modality affects student engagement. PSM achieved a satisfactory reduction in bias according to two of three tests, with the third affected by large sample size. Sensitivity analyses suggest our findings are not excessively vulnerable to the omission of confounding variables. Additionally, the literature has not examined the experience of the growing number of students taking face-to-face and online courses concurrently (DMLs). Other comparisons revealed differences previously attributed to instructional modality are in fact due to compositional differences between the student populations.

As online learning continues its expansion in higher education, this study points to two critical needs: overcoming obstacles to collaborative learning among students studying at a distance and overcoming limitations inhibiting meaningful interactions with faculty. Additionally, future research should seek to illuminate the ways that online delivery may facilitate deep approaches to learning to help improve the educational effectiveness of face-to-face instruction.

ORCID IDS

Justin Paulsen  <https://orcid.org/0000-0003-4200-853X>

Alexander C. McCormick  <https://orcid.org/0000-0003-2656-9636>

NOTES

Justin Paulsen was affiliated with Indiana University during the main work of this article but is now at HumRRO in Louisville, Kentucky.

¹The meaning of these findings is unclear because respondents' interpretation of questions about time spent preparing for class or participation in discussions likely differs for online learners (OLs) in asynchronous delivery formats compared to face-to-face learners (FFLs).

²The small impact may be due to the inclusion in the model of a control variable for learning technology use, which is related to modality of instruction.

³If we think of dual-mode learners (DMLs) as experiencing a lower "dosage" than OLs, the pattern of DML effects that are between those of FFLs and OLs offers reassurance relative to some of the National Survey of Student Engagement survey limitations identified earlier.

REFERENCES

- Allen, I. E., & Seaman, J. (2016). *Online report card: Tracking online education in the United States*. Babson Survey Research Group.
- Anderson, T. (2008). Teaching in an online learning context. In T. Anderson (Ed.), *The theory and practice of online learning* (2nd ed., pp. 343–365). Athabasca University Press.
- Astin, A. W. (1984). Student involvement: A developmental theory for higher education. *Journal of College Student Personnel*, 25(4), 297–308.
- Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, 22(1), 31–72.
- Carnoy, M., Jarillo Rabling, B., Castano-Munoz, J., Duart Montoliu, J. M., & Sancho-Vinuesa, T. (2012). Who attends and completes virtual universities: The case of the Open University of Catalonia (UOC). *Higher Education*, 63(1), 53–82.
- Chen, P. D., Guidry, K. R., & Lambert, A. D. (2009). *Engaging online learners: A quantitative study of postsecondary student engagement in the online learning environment* [Paper presentation]. Annual meeting of the American Educational Research Association, San Diego, CA, United States.
- Chen, P.-D., Gonyea, R., & Kuh, G. (2008). Learning at a distance: Engaged or not? *Innovate: Journal of Online Education*, 4(3), Article 3. <https://nsuworks.nova.edu/innovate/vol4/iss3/3>
- Cho, M. H., & Tobias, S. (2016). Should instructors require discussion in online courses? Effects of online discussion on community of inquiry, learner time, satisfaction, and achievement. *The International Review of Research in Open and Distributed Learning*, 17(2), 123–140.
- Coates, H. (2007). A model of online and general campus-based student engagement. *Assessment & Evaluation in Higher Education*, 32(2), 121–141.
- Cosgrove, S. B., & Olitsky, N. H. (2015). Knowledge retention, student learning, and blended course work: Evidence from principles of economics courses. *Southern Economic Journal*, 82(2), 556–579.
- d'Agostino, R. B. (1998). Tutorial in biostatistics: Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Statistics in Medicine*, 17(19), 2265–2281.
- Filak, V. F., & Sheldon, K. M. (2008). Teacher support, student motivation, student need satisfaction, and college teacher course evaluations: Testing a sequential path model. *Educational Psychology*, 28(6), 711–724.
- Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74(1), 59–109.
- Garratt-Reed, D., Roberts, L. D., & Heritage, B. (2016). Grades, student satisfaction and retention in online and face-to-face introductory psychology units: A test of equivalency theory. *Frontiers in Psychology*, 7, Article 673. <https://doi.org/10.3389/fpsyg.2016.00673>
- Garrison, D., Anderson, T., & Archer, W. (2010). The first decade of the community of inquiry framework: A retrospective. *Internet and Higher Education*, 13, p. 5–9.

- Hu, S., & McCormick, A. C. (2012). An engagement-based student typology and its relationship to college outcomes. *Research in Higher Education, 53*(7), 738–754.
- Jaggars, S., & Bailey, T. (2010). *Effectiveness of fully online courses for college students: Response to a Department of Education meta-analysis*. Community College Research Center, Columbia University.
- Jaggars, S. S., & Xu, D. (2016). How do online course design features influence student performance? *Computers & Education, 95*, 270–284.
- Kuh, G. D. (2003). What we're learning about student engagement from NSSE: Benchmarks for effective educational practices. *Change: The Magazine of Higher Learning, 35*, 24–32.
- Kuh, G. D., Cruce, T. M., Shoup, R., Kinzie, J., & Gonyea, R. M. (2008). Unmasking the effects of student engagement on first-year college grades and persistence. *Journal of Higher Education, 79*(5), 540–563.
- Kuh, G. D., Kinzie, J., Buckley, J. A., Bridges, B. K., & Hayek, J. C. (2007). *Piecing together the student success puzzle: Research, propositions, and recommendations*. Jossey-Bass.
- McCormick, A. C., Kinzie, J., & Gonyea, R. M. (2013). Student engagement: Bridging research and practice to improve the quality of undergraduate education. In M. Paulsen (Ed.), *Higher education: Handbook of theory and research* (Vol. 28, pp. 47–92). Springer.
- McCormick, A. C., & McClenney, K. (2012). Will these trees ever bear fruit? A response to the special issue on student engagement. *The Review of Higher Education, 35*(2), 307–333.
- Means, B., Toyama, Y., Murphy, R., Bakia, M., & Jones, K. (2009). *Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies*. U.S. Department of Education.
- Morgan, S., & Winship, C. (2014). *Counterfactuals and causal inference: Methods and principles for social research*. Cambridge: Cambridge University Press. doi:10.1017/CBO9781107587991
- National Survey of Student Engagement. (2006). *Engaged learning: Fostering success for all students – Annual Results 2006*. Indiana University Center for Postsecondary Research.
- National Survey of Student Engagement. (2013). *A fresh look at student engagement – Annual Results 2013*. Indiana University Center for Postsecondary Research.
- National Survey of Student Engagement. (2015a). *NSSE 2015 engagement indicators: Internal consistency statistics by class level*. <http://nsse.indiana.edu/html/reliability.cfm>
- National Survey of Student Engagement. (2015b). *NSSE 2015 overview*. http://nsse.indiana.edu/html/institutional_reporting.cfm
- Pace, C. R. (1980). Measuring the quality of student effort. *Current Issues in Higher Education, 2*, 10–16.
- Pace, C. R. (1982). *Achievement and the quality of student effort*. National Commission on Excellence in Education.
- Pascarella, E. T. (1985). College environmental influences on learning and cognitive development: A critical review and synthesis. In J. C. Smart (Ed.), *Higher education: Handbook of theory and research* (Vol. 1, pp. 1–61). Agathon.
- Pascarella, E. T., Seifert, T. A., & Blaich, C. (2010). How effective are the NSSE benchmarks in predicting important educational outcomes? *Change: The Magazine of Higher Learning, 42*(1), 16–22.
- Pascarella, E. T., & Terenzini, P. T. (2005). *How college affects students* (Vol. 2). Jossey-Bass.
- Perna, L. W., Ruby, A., Boruch, R. F., Wang, N., Scull, J., Ahmad, S., & Evans, C. (2014). Moving through MOOCs: Understanding the progression of users in massive open online courses. *Educational Researcher, 43*(9), 421–432.
- Porter, S. R. (2011). Do college student surveys have any validity? *The Review of Higher Education, 35*(1), 45–76.
- Rabe-Hemp, C., Woollen, S., & Humiston, G. S. (2009). A comparative analysis of student engagement, learning, and satisfaction in lecture hall and online learning settings. *Quarterly Review of Distance Education, 10*(2), 207–218.
- Reece, A. J., & Butler, M. B. (2017). Virtually the same: A comparison of STEM students' content knowledge, course performance, and motivation to learn in virtual and face-to-face introductory biology laboratories. *Journal of College Science Teaching, 46*(3), 83–89.
- Robinson, C. C., & Hullinger, H. (2008). New benchmarks in higher education: Student engagement in online learning. *Journal of Education for Business, 84*(2), 101–109.
- Rosenbaum, P. R. 2002. *Observational Studies* (2nd ed.). Springer.
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41–55.
- Schuetze, H., & Slowey, M. (2002). Participation and exclusion: A comparative analysis of non-traditional students and lifelong learners in higher education. *Higher Education, 44*, 309–327.
- Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science, 25*(1), 1–21.
- Tinto, V. (1975). Dropouts from higher education: A theoretical synthesis of the recent literature. *Review of Educational Research, 45*(1), 89–125.
- Tinto, V. (1986). Theories of student departure revisited. In J. C. Smart (Ed.), *Higher education: Handbook of theory and research* (Vol. 11, pp. 359–384). Agathon.
- Tseng, H., & Walsh, E. J., Jr. (2016). Blended versus traditional course delivery: Comparing students' motivation, learning outcomes, and preferences. *Quarterly Review of Distance Education, 17*(1), 43–52.
- U.S. Department of Education, National Center for Education Statistics. (2018). *List of current Digest tables, Table 311.15*. https://nces.ed.gov/programs/digest/current_tables.asp

AUTHORS

JUSTIN PAULSEN, PhD, is a research scientist at HumRRO, 700 Hurstborne Parkway, Suite 100, Louisville, Kentucky 40222; JPaulsen@humrro.org. His research focuses on using research methodology and psychometric theory to answer practical questions.

ALEXANDER C. MCCORMICK, PhD, is an associate professor of educational leadership and policy studies at the Indiana University Bloomington School of Education, 201 N. Rose Avenue, Bloomington, IN 47405; amcc@indiana.edu. His research focuses on student engagement, assessment, and evidence-informed improvement in higher education.

Manuscript received February 14, 2018
 Revisions received September 24, 2018,
 April 3, 2019, and August 15, 2019
 Accepted September 29, 2019