Student Outcomes in Online Courses: When Does Class Size Matter?

Rebecca A. Thomas

Ecampus, Oregon State University

Mary Ellen Dello Stritto

Ecampus, Oregon State University

Author Note

Rebecca Arlene Thomas, Ph.D., is currently the Postdoctoral Scholar in the Oregon State University Ecampus Research Unit (ECRU), and Mary Ellen Dello Stritto, Ph.D., is currently the Director of the ECRU. The ECRU conducts and supports research related to online teaching and learning in higher education, and promotes research literacy and collaboration within the field. Prior to joining the ECRU, Rebecca received a master's degree in Instructional Psychology & Technology from Brigham Young University and a Ph.D. in Psychology from the University of Texas San Antonio. Rebecca's research focuses on educational quality and stakeholder experiences, educational innovation, and interpersonal relationships.

Mary Ellen received a Ph.D. in Psychology from Claremont Graduate University, and specialized in quantitative methodologies, survey design, and statistical analysis.

Abstract

This quantitative study investigated the relationship between class size and student outcomes (final grades and DFW rates) in online higher education courses offered by a large, 4 year public institution in the United States. The following class size cut-off points were used: 8-15 vs. 16 or more students, 8-30 vs. 31 or more students, 8-40 vs. 41 or more students, and 8-50 vs. 51 or more students. Course level data included average final grades and DFW rates for 391 online undergraduate courses taught during the years 2017 and 2018. Significant results suggest that students earned higher grades in STEM (Science Technology Engineering Mathematics) and upper-division courses when online courses included 30 or fewer students. This suggests that it may be beneficial to limit certain kinds of courses to 30 students or fewer, as 30 students may be a tipping point where the benefits of smaller online classes wear off.

Keywords: online education, higher education, class size, student outcomes, student grades, student DFW rates, STEM, upper-division courses, lower-division courses

Student Outcomes in Online Courses: When Does Class Size Matter?

Enrollment in online courses and programs has continued to grow for over a decade (e.g. Seaman, Allen, & Seaman, 2018), and some leaders in higher education feel that online education may become even more common as instructors and students move forward after the COVID-19 pandemic (L. Templeton, personal communication, October 14, 2020). This growth in online higher education has several advantages, including the potential to provide individuals accessible means to higher education courses, certificates, and degrees, as online education generally allows students to work where and when they can. Additionally, students may find that taking courses in multiple modalities, such as face-to-face, blended, and online, fits better with their personal lives and goals.

Considering that online education is becoming increasingly common, investigation related to online learning efficacy is of interest to multiple stakeholder groups, including students, faculty and administrators. In education, class size has been discussed and examined for hundreds of years, especially in K-12 contexts (e.g. Angrist & Lavy, 1999; Blatchford, Bassett, & Brown, 2011). However, fewer studies have studied class size in online higher education courses. How many students should optimally be enrolled in each section of an online course? How does class size relate to student performance in online courses, such as final grades and DFW (course grades of D, F or Withdrawal) rates? This article addresses this topic by reviewing the current literature about class size in online higher education courses, summarizing the results of a study of the relationship between class size and student outcomes in a sample of over 300 undergraduate online higher education courses, and discussing implications for future research and practice.

Literature Review

Class size, or the number of students assigned to one instructor in one section of a course, has long been discussed and studied in K-12 education, and more recently, in higher education (e.g. Taft, Kesten, & El-Banna, 2019). In general, smaller class sizes have been seen as ideal across all levels of education, with university level courses traditionally following a pattern where the largest class sizes have occurred in introductory courses, with class sizes becoming progressively smaller as course content becomes more advanced and specialized (Monks & Schmidt, 2011). However, there is still ambiguity surrounding “ideal” class sizes in higher education, as it is an ongoing area of research (e.g. Ake-Little, von der Embse, & Dawson, 2020).

There are additional factors to consider when discussing class size for online courses in higher education. For example, online courses function differently than face-to-face courses, as discussions and other collaboration takes place in online environments, and online classes are not limited to the number of seats available in a physical space. Technology could also be used to help larger online courses “feel” smaller in ways that could not happen face-to-face (e.g. participants in large online classes could be assigned to one of multiple discussion boards). Additionally, since colleges and universities have experienced financial challenges, some have suggested that increasing class sizes in online courses may be a way to generate more efficient revenue (e.g. Bettinger, Doss, Loeb, Rogers, & Taylor, 2017; Taft, Perkowski, & Martin, 2011), and that universities should utilize economies of scale if class size could increase without adverse outcomes (Kokkelenberg, Dillon, & Christy, 2008). However, others have argued that many of the advantages found in smaller face-to-face classes remain in online classes (e.g. Taft et al., 2019). Therefore, institutions who aim to deliver quality online education to their students need clarity surrounding ideal class sizes for online courses and programs. The following sections of this review summarize research investigating how online class size may relate to student outcomes in online courses.

Class Size Definitions: What Makes An Online Class "Small" or "Large?"

Before summarizing research on how online class size relates to student outcomes, it is important to define what is meant by a "small" or "large" online course. However, this is somewhat difficult because there is a lack of consensus regarding class size definitions, in research, as well as practically (Lowenthal, Nyland, Jung, Dunlap, & Kepka, 2019). For example, in Lowenthal et al.'s (2019) survey of online faculty, roughly 40% considered a class of 30 or more students to be "high enrollment", but 10% did not consider a class to be "high enrollment" unless it included 100 or more students. Therefore, there may be widely varying perceptions of class size when using descriptors such as "small," "medium," and "large," and it is important to consider the specific class sizes when reviewing the literature, as well as discussing the topic of class size.

While different studies have defined "small" and "large" online classes differently, differences in outcomes between certain class sizes have been studied and discussed more than others. For example, college ranking systems have assigned more value to online classes with 30 or fewer students compared to online classes with more than 30 students (Monks & Schmidt, 2011). This cut-off point of 30 students per online class has also been seen in several studies. For example, in a study of "high enrollment" online courses, Lowenthal et al. (2019) studied the perceptions of faculty who taught online courses with 30 or more students, and Bettinger et al. (2017) compared student outcomes in classes with an average of 31 students to classes with 10% more students. Additionally, after a review of 58 studies related to online course enrollment, Taft et al. (2019) proposed the following online class size categories: ≤15 students were categorized as a small class, 16-23 students were categorized as a small/medium class, 24-30 students were categorized as a medium class, 31-39 students were categorized as a medium/large class, and 40 or more students were categorized as a large class. In this categorization, 30 students could be seen as a midpoint, as small and medium courses would include 30 or fewer students, while medium/large and large classes would include over 30 students.

Interestingly, the cut-off of roughly 30 students has also been seen in literature about K-12 education (e.g. Angrist & Lavy, 1999), as well as face-to-face higher education (e.g. Ake-Little et al., 2020). Some higher education research has suggested that instructors may adjust course objectives depending on class size, and may use more timesaving techniques, and assign fewer academically meaningful assignments such as writing prompts, when courses exceed 31 to 40 students (Ake-Little et al., 2020). Therefore, differences in courses smaller and larger than 30 students are important to assess, as this cut-off has been used in practice, as well as in the literature.

However, other literature has proposed other common cut-off points; namely, 15 students, 40 students, and 50 students. For example, in a paper about online higher education, Taft et al. (2019) categorized 15 and fewer students as “small” courses, and 40 or more students as “large” courses. Research on face-to-face higher education has suggested that “small” classes (15 or fewer students) may elicit a shift where students may perceive the environment as a “community” rather than a class, where students and instructors may be more likely to know each other by name, where students may be expected to participate in a variety of ways (e.g. discussion and writing), and where instructors may provide more detailed feedback (Cuseo, 2007). Additionally, the literature cited above by Ake-Little et al. (2020) proposed that the range between 31-40 students may be when instructors broaden course objectives and use more timesaving techniques, suggesting that considering the 40 student cut-off point may be as important as considering the 30 student cut-off point. Lastly, research on student preferences in online courses have indicated that most students prefer courses under 50 students (e.g. Roby, Ashe, Singh, & Clark, 2013), and some college ranking systems have penalized online programs with 50 students or more per class. Therefore, this study considered multiple cut-off points for online class size, including 15, 30, 40, and 50 students.

The Relationship Between Online Class Sizes and Student Outcomes

A small body of research has investigated how online class sizes relate to student learning and performance outcomes such as grades and DFW rates, while more of the research about online class size has focused on student and faculty perceptions and satisfaction (e.g. Lowenthal et al., 2019; Taft et al., 2019). However, the following categories of research can be used to inform hypotheses relating to class size and student outcomes in online education:

1. Research about how online class size relates to pedagogy and course design
2. Research about how online class size relates to instructor performance
3. Research about how online class size relates to student outcomes in online education

While the following review primarily focuses on literature about online higher education, sources from K-12 and general higher education are included to supplement these sections, as more class size research has been conducted in these contexts.

How Online Class Size Relates to Pedagogy and Course Design

In general, past research about class size suggests that certain course designs and pedagogies function most optimally if classes do not become too large. For example, research about class size in K-12 and face-to-face higher education has suggested that larger classes tend to be designed to include more timesaving techniques, such as automatic grading as opposed to personalized feedback (e.g. Ake-Little et al., 2020), and more reliance on the lecture method of instruction (e.g. Cuseo, 2007), while smaller classes tend to ask students to actively participate more in class (e.g. in discussions), as well as outside of class (e.g. through writing assignments; Cuseo, 2007). In general, this research suggests that class size can influence how courses are designed, as one instructor cannot provide the same level of attention to each student in larger classes, and it can also be more difficult to facilitate discussions and group work with larger numbers of students.

In research specific to online higher education, Taft et al. (2019) conducted a literature review of 58 articles published from 2012 to 2017 that were relevant to online class size. They found that the articles reviewed cited three main theories for determining online class size: objectivist/constructivist theory, Bloom's taxonomy, and the Community of Inquiry (CoI) framework. Classes with pedagogies that were more constructivist in nature, with learning objectives higher on Bloom's taxonomy, or that aimed to achieve more teaching presence (e.g. focusing course discussions and setting curriculum), cognitive presence (e.g. exchanging information and ideas) and social presence (e.g. portraying the self as real in a technology mediated environment) were theorized to function most optimally in "smaller" classes (see Garrison, Anderson, & Archer, 2000 for CoI framework definitions, and Taft et al., 2019 for specific articles that have discussed these three theories).

While larger courses (e.g. 40 or more students) tended to be more effective for learning objectives lower on Bloom's taxonomy, such as remembering and understanding, courses with 40 or fewer students were more optimal for higher levels of Bloom's taxonomy, such as application (Taft et al., 2019). Very small courses with 15 or fewer students were most ideal for the course objectives of synthesis and evaluation (e.g. Taft et al., 2011). Taft et al. (2019) summarized that smaller classes were ideal for nuanced learning, substantive online interaction, student development, mastery of more complex learning outcomes, and higher order thinking.

Research About How Online Class Size Relates to Instructor Performance

In addition to "smaller" class sizes lending toward certain pedagogies and course design, research has also suggested that online class size relates to instructors' abilities to perform ideally as educators. For example, online instructors in Lowenthal et al.'s (2019) study reported that classes with 30 or more students made grading and other responsibilities more time consuming, as well as limited the interaction they could have with students. Similarly, Russell and Curtis (2013) compared an online course with 125 students to a course with 25 students, and found that limited student-instructor interaction in the larger class, with the instructor reporting that the larger course's workload prohibited interaction. Some research has suggested that limited student-instructor interaction in online courses can negatively impact students' motivation to persist (Dykman & Davis, 2008). This research in online higher education is consistent with work on class size in K -12 and face-to-face higher education, which has found that in smaller classes, teachers tend to give more focus to each student (Blatchford et al., 2011), as well as more frequent interaction and feedback to each student, (Cuseo, 2007).

In addition to studying workload and student-instructor interaction, Sorenson (2015) conducted a study where they analyzed online teaching performance for 380 instructors at a large for-profit university. To do this, they asked faculty to walk through the online courses and conduct a peer review where they ranked the instructors in five categories: fostering critical thinking, providing instructive feedback, demonstrating high expectations, establishing relationships with students, and utilizing their expertise. While the authors found no significant differences in instructor ratings by class size, the study examined courses with class sizes ranging from 1 to 29 students, with the average class size being 15 students. Considering that other studies compared courses of 25 to 30 students with courses quite a bit larger, such as courses with over 100 students (e.g. Russell & Curtis, 2013), it is possible that teaching performance would have suffered more if class sizes had been increased more substantially. In general, the literature suggests that instructors may have a more difficult time providing quality interaction and feedback in courses larger than 30 students, which could potentially impact student learning outcomes.

Research About How Class Size Relates to Student Outcomes

To date, limited research has investigated how class size relates to student performance outcomes such as student grades and course withdrawals in online classes specifically. However, one study by Bettinger et al. (2017) studied over 4,000 sections of 111 online courses at DeVry University, and compared student outcomes in “regular classes,” which had an average of 31 students, to courses where 10% more students were added to the course, making the average “treatment classes” include 34 students. When the researchers compared the “regular classes” to the “treatment classes,” there were no significant differences in student outcomes such as final grades and future enrollment, but there were more students who withdrew from the larger classes themselves. However, it is important to consider that this study only added an average of 3 more students to each course, and that the class sizes for both groups stayed close to the common cut-off point of 30 students per class found elsewhere in the field and in the literature.

In research focusing on class size for face-to-face classes in higher education, multiple studies have found interactions between course content area, class size, and other factors on student outcomes. For example, studies have suggested that students tend to show lower engagement and achievement in larger STEM (Science Technology Engineering Mathematics) courses compared to other fields (e.g. Kara, Tonin, & Vlassopoulos, 2020; Scott, McNair, Lucas, & Land, 2017). Additionally, other studies have suggested that class size and content area may impact performance differently for students based on gender and ethnicity (e.g. Ake-Little et al., 2020), as well as ability level, with smaller courses benefitting lower ability students (DePaola, Ponzo, & Scoppa, 2013). Therefore, research suggests that it is important to consider multiple factors in the discussion about class size, such as course design and learning outcomes for the course, as well as field of study.

The Current Study

In general, the current research on class size in online higher education suggests that "smaller" courses tend to be more ideal for certain course designs and pedagogies, instructor performance, and student outcomes. However, research investigating student performance outcomes such as final grades and DFW rates in online higher education has been limited, with even less research investigating the interaction between class size and other factors such as course content area on student performance. The current study aims to address this gap by investigating the following research questions. For a visual representation of the research questions paired with the class size cut-off points analyzed see Table 1.

Table 1

Visual Representation of Study Research Questions and Class Size Comparisons

	Comparisons/Class Size Cut-Off Points			
Research Questions 1-4:	**8-15 vs. 16+ students**	**8-30 vs. 31+ students**	**8-40 vs. 41+ students**	**8-50 vs. 51+ students**
1. For all courses: Do student outcomes differ by class size?				
2. STEM vs. non STEM: Is there an interaction between class size and course content area (STEM vs. non-STEM) on student outcomes?				
3. Upper vs. lower-division: Is there an interaction between class size and course divison (upper vs. lower divison) on student outcomes?				
4. Is there an interaction between class size, division (upper, lower), and course content area (STEM, non-STEM) on student outcomes?				

Note. Student outcomes included the two dependent variables in this study: average course grade, and DFW rate.

Research hypotheses are listed below the four research questions.

Research Question 1: Overall, do student outcomes (average course grades and DFW rates) differ in online courses of different sizes?

Research Question 2: When considering class size in online courses, are there differences in student outcomes for upper and lower-division courses?

Research Question 3: When considering class size in online courses, are there differences in student outcomes for STEM (Science Technology Engineering and Mathematics) courses compared to other course content areas?

Research Question 4: Is there an interaction between class size, division (upper, lower), and course content area (STEM, non-STEM) on student outcomes?

Hypotheses: We predict that factors such as pedagogy and course design, and as well as instructor performance, may lead to lower grades and higher DFW rates in larger courses. We predict that there will be a stronger relationship between class size and student outcomes for upper division courses compared to lower division courses. We also predict that there will be a stronger relationship between class size and student outcomes for STEM courses compared to non-STEM courses. Lastly, we predict that student outcomes will differ the most by class size in STEM upper-division courses.

Method

This quantitative study utilized archival university data collected through our institution's division for online courses and programs. Our institution is a large, 4-year public institution in the northwestern part of the United States. This paper reports on the results of data analysis using data from a larger study about both online and face-to-face courses. We selected online undergraduate courses taught during the years of 2017 and 2018 for this analysis. In cases where the course was taught more than one term, we selected the most recent term that the course was taught for the analysis. Courses with fewer than 8 students enrolled were excluded from the dataset (e.g. excluding independent study and other very small courses). The resulting dataset included 391 online undergraduate courses taught in diverse fields at the institution (see Tables 1 and 2 for descriptive statistics about the courses in the dataset). Class size ranged from 8 students per class to 226 students per class.

The data for each course were provided at the course level. For example, our dataset included enrollment numbers for each course (i.e. 32 students were enrolled in this course), the numbers of students who received a final grade of A, B, C, D, and F in each course, and the number of students who withdrew from each course. This data allowed us to compute average grade point averages (GPAs) for each course, as well as calculate the proportion of students in each course who received certain grades, or who withdrew. The dataset also included course level descriptive statistics, such as the number of students of color in each course. However, since the data were provided at the course level, not the individual student level, the data did not allow us to analyze whether student outcome variables (grades, DFW rates) correlated with any of the descriptive variables at the individual level. Therefore, all analyses were conceptualized at the course level.

While the full dataset was used to address Research Question 1, the data needed to be separated into upper and lower-division courses for Research Question 2, and separated into STEM and non-STEM courses for Research Question 3. Course numbers were used to categorize courses into “lower-division” and “upper-division” courses, with 100-200 level courses being defined as lower-division, and 300-400 level courses being defined as upper-division. The institutional college where each course was offered was used to categorize courses into “STEM” and “non-STEM.” See Table 2 for categorizations.

Table 2

Information About the Online Courses Analyzed

University College	STEM Classification	Number of Courses in Dataset	Number of Lower-Division Courses	Number of Upper-Division Courses
Agricultural Sciences	STEM	77	4	73
Business	Non-STEM	23	1	22
Education	Non-STEM	4	0	4
Engineering	STEM	15	3	12
Earth, Ocean, and Atmospheric Sciences	STEM	20	5	15
Forestry	STEM	13	1	12
Liberal Arts	Non-STEM	185	35	150
Public Health and Human Services	STEM	27	6	21
Science	STEM	24	3	21
Academic Preparation Courses	Non-STEM	3	2	1

Note. Lower-division courses included courses at the 100-200 level, while upper-division courses included courses at the 300-400 level.

Analyses were conducted using IBM SPSS Statistics 26, with each course represented as a case. To analyze the data, a series of one-way ANOVAs were used to test differences in GPA and DFW rates between courses of different sizes, as one-way ANOVAs are robust to assumptions of normality (Schmider, Ziegler, Danay, Beyer, & Bühner, 2010). Based on the literature, all analyses were conducted with the following cut-off points for class size: 15, 30, 40, and 50 students per class. In other words, courses with 8-15 students were compared to courses with 16 or more students, courses with 8-30 students were compared to courses with 31 or more students, courses with 8-40 students were compared to courses with 41 or more students, and courses with 8-50 students were compared to courses with 51 or more students.

To address Research Question 1, analyses were conducted using the full sample of courses. To address Research Question 2, analyses were run separately for lower division courses (courses at the 100 and 200 levels) and upper division courses (courses at the 300 and 400 levels). To address Research Question 3, analyses were run separately for STEM and non-STEM courses. To address Research Question 4, analyses were run separately for upper-division STEM courses and lower-division STEM courses. While we would have been interested in analyzing data for lower division STEM and non-STEM courses separately, these analyses were excluded due to low numbers of lower division STEM courses ($n = 22$) and lower division non-STEM courses ($n = 38$) in the dataset.

Results

Table 3 shows descriptive statistics for all course groupings, and Table 4 shows the numbers of courses for each of the groupings. One-way ANOVAs were conducted for all cut-off points for class size (15, 30, 40, and 50 students), with average GPA, as well as DFW rates (percentage of students who received a D or F grade, or who withdrew from each course) as the dependent variables. See Table 5 for a visual of all of the analyses run for GPA differences, and Table 6 for a visual of all of the analyses run for differences in DFW rates. In general, there were few statistically significant results (see Tables 4 and 5). Therefore, the following summary emphasizes significant results. As Table 3 shows, average class sizes and GPAs were similar between groups and the DFW rates ranged from 14% to 17% in the course groupings. We had similar numbers of STEM and non-STEM courses, but had a larger group of upper-division courses compared to lower-division courses.

Table 3

Descriptive Statistics About the Online Courses Analyzed

Course Groupings	**Number of Courses**	**Average Class Size *M (SD)***	**Average GPA *(M, SD)***	**Average DFW Rate *(M, SD)***
All Courses in Dataset	391	32.9 (27.0)	3.16 (0.4)	14% (11%)
STEM Courses	176	41.7 (36.0)	3.16 (0.5)	14% (11%)
Non-STEM Courses	215	25.7 (12.5)	3.16 (0.4)	14% (11%)
Lower-Division Courses	60	36.82 (34.3)	3.10 (0.4)	17% (11%)
Upper-Division Courses	331	32.19 (25.5)	3.17 (0.5)	14% (11%)
STEM Upper-Division Courses	154	39.24 (33.5)	3.17 (0.5)	14% (12%)
Non-STEM Upper-Division Courses	177	26.06 (12.9)	3.18 (0.5)	14% (10%)

Note. Lower-division courses included courses at the 100-200 level, while upper-division courses included courses at the 300-400 level.

Table 4

Descriptive Statistics: Number of Courses Per Group

Number of Courses (N) Per Group									
Courses Included	***N***	**8-15**	**16+**	**8-30**	**31+**	**8-40**	**41+**	**8-50**	**51+**
Full dataset	391	75	316	230	161	307	84	334	57
STEM	176	29	147	74	102	117	59	129	47
Non-STEM	215	46	169	156	59	190	25	205	10
Lower-Division	60	9	51	34	26	44	16	48	12
Upper-Division	331	66	265	196	135	263	68	286	45
STEM Upper-Division	154	28	126	67	87	108	46	118	36
Non-STEM Upper-Division	177	38	139	129	48	155	22	168	9

Results: Online Class Size and GPA

Full results for the relationship between class size and GPA are included in Table 5. Average GPA data were missing for 10 courses in the sample, resulting in a sample size of 381 courses for the following analyses. There were no significant differences in course GPA (students' final grades) when using the cut-off points of 15 students, 40 students, and 50 students. However, there were some statistically significant results when comparing courses with 30 or fewer students to courses with greater than 30 students. For the full sample, significantly higher course GPAs occurred in courses with 30 students or fewer ($N = 222$, $M = 3.21$, $SD = 0.45$) compared to courses with more than 30 students ($N = 159$, $M = 3.10$, $SD = 0.42$), $F(1, 379) = 5.736$, $p = 0.02$).

Table 5

Differences in Average GPAs by Class Size

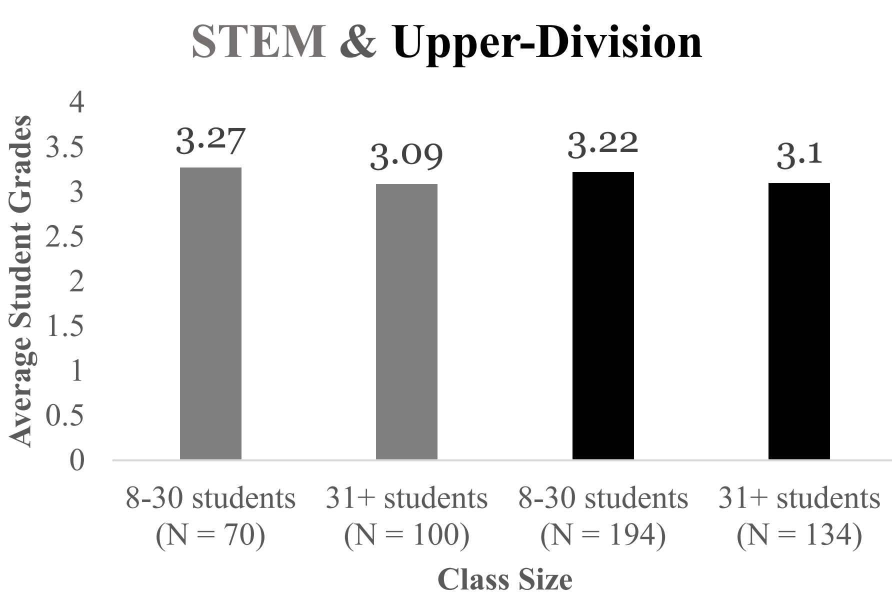
		Class Sizes (in number of students)			
Courses Included	N	**8-15 vs. 16+**	**8-30 vs. 31+**	**8-40 vs. 41+**	**8-50 vs. 51+**
Full dataset	381	$F(1, 379) = 3.23, p = 0.07$	***Higher GPA in smaller courses $F(1, 379) = 5.74, p = 0.02$**	$F(1, 379) = 0.45, p = 0.51$	$F(1, 379) = 0.04, p = 0.85$
STEM	170	$F(1, 168) = 1.60, p = .21$	***Higher GPA in smaller courses $F(1, 168) = 6.672, p = 0.011$**	$F(1, 168) = .00, p = 0.98$	$F(1, 168) = 0.01, p = 0.92$
Non-STEM	211	$F(1, 209) = 1.67, p = 0.20$	$F(1, 209) = 1.08, p = 0.30$	$F(1, 209) = 1.37, p = 0.24$	$F(1, 209) = 0.48, p = 0.49$
Lower-Division	53	$F(1, 51) = 0.06, p = 0.81$	$F(1, 51) = 0.46, p = 0.50$	$F(1, 51) = 0.03, p = 0.86$	$F(1, 51) = 0.01, p = 0.95$
Upper-Division	328	$F(1, 326) = 2.91, p = 0.09$	***Higher GPA in smaller courses $F(1, 326) = 5.05, p = 0.03$**	$F(1, 326) = 0.50, p = 0.48$	$F(1, 326) = 0.01, p = 0.92$
STEM Upper-Division	152	$F(1, 150) = 1.47, p = 0.23$	***Higher GPA in smaller courses $F(1, 150) = 5.86, p = 0.02$**	$F(1, 150) = 0.11, p = 0.74$	$F(1, 150) = 0.23, p = 0.63$
Non-STEM Upper-Division	176	$F(1, 174) = 1.42, p = 0.24$	$F(1, 174) = 0.64, p = 0.43$	$F(1, 174) = 2.46, p = 0.12$	$F(1, 174) = 0.97, p = 0.33$

Note. *Indicates p < .05. Bolded results indicate statistical significance.

When STEM and non-STEM courses were analyzed separately, significantly higher GPAs occurred in courses with 30 students or fewer ($N = 70$, $M = 3.27$, $SD = 0.43$) compared to courses with more than 30 students for STEM courses ($N = 100$, $M = 3.09$, $SD = 0.44$), $F(1, 168) = 6.672$, $p = 0.01$ (see Figure 1), but not for non-STEM courses. Similarly, when lower and upper-division courses were analyzed separately, significantly higher GPAs occurred in courses with 30 students or fewer ($N = 194$, $M = 3.22$, $SD = 0.46$) compared to courses with more than 30 students for upper-division courses ($N = 134$, $M = 3.10$, $SD = 0.42$), $F(1, 326) = 5.054$, $p = 0.03$ (see Figure 1), but not for lower-division courses. Lastly, when upper-division courses were analyzed separately by STEM and non-STEM courses, significantly higher GPAs occurred in courses with 30 students or fewer ($N = 66$, $M = 3.27$, $SD = 0.44$) compared to courses with more than 30 students for upper-division STEM courses ($N = 86$, $M = 3.09$, $SD = 0.45$), $F(1, 150) = 5.857$, $p = 0.02$, but not for upper division non-STEM courses. Overall, these results suggest that the class size cut-off point of 30 students may be the point where we see differences in student performance in upper-division STEM classes.

Figure 1

Comparing Class Sizes of 8-30 Students With 31+ Students: STEM and Upper-Division Courses



Note. Average grade point averages are graphed for STEM courses (gray bars on graph), and Upper-Division courses (black bars on graph). Some courses may be represented in two bars on the graph, as some courses may be included in the STEM group, as well as the upper-division group.

Results: Online Class Size and DFW Rates

Full results for the relationship between class size and DFW rates are included in Table 6. There were no significant differences in DFW rates when using the cut-off points of 15 students, 30 students, and 50 students. Similarly, there were no significant differences in DFW rates when using the cut-off point of 40 students, and conducting the analysis with the full sample, STEM courses only, and lower-division and upper-division courses. However, there were statistically significantly higher DFW rates in non-STEM courses with 40 or fewer students (N = 190, M = 15%, SD = 11%) compared to non-STEM courses with greater than 40 students (N = 25, M = 11%, SD = 10%), $F(1, 213) = 3.784$, p = .05.

Table 6

Differences in Average DFW Rates by Class Size

Courses Included	*N*	Cut off: 15 students	Cut off: 30 students	Cut off: 40 students	Cut off: 50 students
Full dataset	391	$F(1, 389) = 0.34$, $p = 0.56$	$F(1, 389) = 0.41$, $p = 0.52$	$F(1, 389) = 1.59$, $p = 0.21$	$F(1, 389) = 0.80$, $p = 0.37$
STEM	176	$F(1, 174) =1.02$, $p = 0.31$	$F(1, 174) = 0.67$, $p = 0.42$	$F(1, 174) = 0.00$, $p = 0.99$	$F(1, 174) = 0.00$, $p = 0.96$
Non-STEM	215	$F(1, 213) = 0.02$, $p = 0.88$	$F(1, 213) = 2.47$, $p = 0.12$	***Higher DFW rates in smaller courses** **$F(1, 213) = 3.78$, p = .05**	$F(1, 213) = 2.36$, $p = 0.13$
Lower-Division	60	$F(1, 58) = 0.01$, $p = 0.94$	$F(1, 58) = 0.13$, $p = 0.72$	$F(1, 58) = 0.06$, $p = 0.80$	$F(1, 58) = 0.98$, $p = 0.33$
Upper-Division	331	$F(1, 329) = 0.57$, $p = 0.45$	$F(1, 329) = 0.34$, $p = 0.56$	$F(1, 329) = 1.95$, $p = 0.16$	$F(1, 329) = 0.41$, $p = 0.52$
STEM Upper-Division	154	$F(1, 152) = 1.13$, $p = 0.29$	$F(1, 152) = 0.24$, $p = 0.63$	$F(1, 152) = 0.25$, $p = 0.62$	$F(1, 152) = 0.04$, $p = 0.83$
Non-STEM Upper-Division	177	$F(1, 175) = 0.00$, $p = 0.98$	$F(1, 175) = 2.57$, $p = 0.11$	$F(1, 175) = 3.36$, $p = 0.07$	$F(1, 175) = 1.31$, $p = 0.25$

Note. *Indicates p < .05. Bolded results indicate statistical significance.

Discussion

Understanding the relationship between class size and student outcomes in online higher education courses can inform beneficial decisions for online students. The current study investigated the relationship between class size and student grades and DFW rates in online undergraduate courses. The study looked at four cut-off points for class size found previously in the literature (15, 30, 40, and 50 students), and investigated the relationship between class size and student outcomes overall, based on course content area (STEM vs. non-STEM), and based on level (upper vs. lower-division).

Final Grades and Class Size

We hypothesized that final grades would be higher in smaller courses, and that there would be greater differences in student outcomes based on class size in STEM courses, as well as upper division courses. These hypotheses were supported, but only when comparing courses with 30 or fewer students to courses with more than 30 students. Students earned higher final grades in classes with fewer than 30 students overall, with greater differences in grades for STEM courses, as well as upper division courses. When comparing final grades using the cut-off points of 15 students, 40 students, or 50 students, there were no significant differences.

This suggests that aspects of “smaller” courses, such as pedagogy and course design, as well as instructor performance capabilities, may help some students achieve higher grades in STEM and upper division courses. It is important to note that STEM and upper division courses are more likely to teach learning outcomes higher on Bloom’s taxonomy, and therefore might benefit from more constructivist teaching methods and instructor attention and feedback compared to students in lower division courses (e.g. Taft et al., 2019). However, our findings suggest that students did not earn higher grades in courses with 8-15 students compared to courses with more than 15 students. This is interesting because some research has suggested that “small” classes (15 or fewer students) may feel like more of a community, and may involve more student participation (through discussion, writing, etc.), as well as more instructor feedback (Cuseo, 2007). Additionally, since previous research has found classes larger than 30 students to be more time-consuming for the instructor (Lowenthal et al., 2019), one could assume that classes with 15 or fewer students would be less time-consuming than classes with 30 students. It could be that the pedagogical benefits of courses with 15 or fewer students do not transfer to students' final grades, or that instructors may more efficiently use their time in courses with 16-30 students. However, if STEM and upper division courses exceed 30 students, lack of instructor time may lead to sacrifices in valuable pedagogy.

While the differences in GPA weren't large for the 30 student cut-off, with 0.1-0.2 GPA increases, these increases may be practically significant for certain students. Future work in this area could investigate student outcomes other than final grades, such as performance for specific projects or labs, as it is possible that pedagogical or instructor factors may be more beneficial for students doing specific projects.

DFW Rates and Class Size

We hypothesized that DFW rates would be higher in larger courses, and predicted that there would be greater differences in student outcomes based on class size in STEM courses, as well as upper division courses. However, our results did not support these hypotheses. In general, we found very few significant differences in DFW rates by class size, with one analysis finding significantly higher DFW rates in non-STEM courses with 40 or fewer students compared to greater than 40 students. Considering we hypothesized that this relationship would be in the other direction, with higher DFW rates in larger courses, this result may warrant further consideration.

While past research has found that students were more likely to withdraw from larger courses (Bettinger et al., 2017), our results did not support this overall. In addition to withdrawals, this result suggests that differences in grades by class size may be more likely to be differences in A, B, and C grades, as opposed to differences in D and F grades. In other words, the pedagogy and instructor time allowed in smaller classes may enhance students' performance, but other factors may be more influential in students' decisions to withdraw, or failure. Future research can further analyze why certain class sizes (such as courses with 30 students or fewer) may relate to specific grades, or if they relate to various aspects of persistence.

Could 30 Students Be a Tipping Point Between "Small" and "Large" Online Classes?

The cut-off point of 30 students has been found frequently in the literature, as well as practically, in college ranking systems for online courses and programs (e.g. Monks & Schmidt, 2011; Taft et al., 2019). Therefore, it is interesting to note that our significant findings regarding final grades corresponded to this cut-off point. This suggests that around 30 students may be the tipping point where the benefits of "small" online courses may wear off, with instructors changing course aspects such as their pedagogy, learning outcomes, or communication once courses become larger than 30 students. A potential factor in this relationship may be that the types of course assignments that occur in courses may shift as class size increases. For example, in a course with 15-20 students, instructors my find that grading weekly papers manageable, but this may become more difficult in courses with over 30 students. In courses with higher numbers of students they may be more likely to use self-grading assessments (e.g. quizzes), and thus there may be fewer opportunities for individualized feedback. Further, in an online course the level of interaction an instructor has with individual students on a discussion board may be reduced as the number of students in the course increases. Instructors may also find courses with 30 or fewer students to be easier to manage when it comes to collaborative group work, as instructors would have fewer discussions to moderate.

Implications

Overall, these results imply that students may benefit if specific types of online courses, such as STEM courses or upper division courses, are kept under 30 students per class. However, when it comes to managing class sizes in other types of courses, such as introductory liberal arts courses, students may achieve similar outcomes in larger classes. Therefore, we recommend that stakeholders such as faculty and administrators evaluate whether particular courses align with the needs of smaller class sizes (for a guide on how to do this, see Taft et al., 2019). This alignment on a course-by-course basis may allow some institutions to keep certain courses larger, while leveraging the benefits of smaller class sizes for courses where students could benefit the most.

For online courses where higher enrollments could be appropriate for learning outcomes, certain strategies may save instructors time without sacrificing pedagogy. For example, Trammell and LaForge (2017) suggested practices such as enhancing instructor social presence through video and using rubrics to streamline grading so that students can receive immediate feedback. Future research could further investigate how technology could streamline processes for larger online courses.

Limitations

Several limitations of this study's results are worth noting. First, previous research in general higher education has suggested that student level factors, such as gender or ethnicity, may interact with class size and student success (e.g. Ake-Little et al., 2020). Therefore, factors other than those studied here should be considered when determining class size. Outside of the realm of student success, there may be other reasons to keep class sizes at a certain level, such as instructor preferences or burn-out (e.g. Lowenthal et al., 2019). Therefore, stakeholders should consider multiple factors, including student success variables, when determining online class sizes in higher education.

Limitations regarding the dependent variables, course grades, and DFW rates, should also be noted. While student grades can be an indication of student learning, grades can also be an indication of other factors such as an instructors' grading policies or student life factors (i.e. a student going through a personal life event may have difficulty focusing). It is possible for two students who earned the same grade to have different levels of course content understanding. Therefore, other research could further investigate how class size relates to more specific learning outcomes.

Additionally, this dataset was obtained from one 4-year public institution in the United States, and therefore, results may not generalize to all institutions that provide online higher education. While we aimed to investigate diverse course content areas, the content areas studied here may also differ from other institutions. While we think it was useful to use the cut-off points we did (15, 30, 40, and 50 students), some of the sample sizes for the group comparisons were uneven and small (see Table 4), which limited the interpretability of the tests for the higher and lower cut-off points (under and over 15 students, and under and over 50 students). Lastly, this study investigated broad differences in student outcomes based on very general course content areas (i.e. STEM and non-STEM) and educational levels (i.e. upper and lower division). If courses were analyzed more specifically, such as comparing different course types within the school of Liberal Arts, additional results may have surfaced. Therefore, future research should explore these effects in more specific course types, as well as include other variables.

Future Research

Considering the size of online classes is only the beginning when it comes to ensuring quality online education. While class size may impact how instructors teach, other factors, such as classroom dynamics and teacher-student communication (Blatchford et al., 2011) and instructor bias can also contribute to student outcomes in online courses. For example, research about class size in face-to-face courses suggests that underrepresented students are less likely to experience the benefits of smaller classes, possibly due to a lack of mentoring or support system (Ake-Little et al., 2020). Future work could also consider specific fields, such as specific college majors, as there may be differences in performance for students in different STEM fields. It is also possible that institution type, such as community college vs. 4 year institution, or public vs. private institution, may impact how class size is perceived as well as how individual students experience their education. Lastly, students cannot optimally perform in the classroom without the tools (e.g. technology, internet access) necessary to succeed, suggesting that further work is needed in areas related to educational equity.

While this study measured student grades and DFW rates, other outcomes may be important to consider, such as student understanding of specific content, student participation in high impact practices (i.e. research labs), or future employment. Grades are not the "be-all-end-all" of ideal outcomes for college students, and class size may impact other variables, or interact with variables such as instructor bias to produce outcomes. Future work in this area should further consider institutional factors, instructor factors, and student factors in the context of class sizes in online higher education. For example, work could investigate how class size relates to decisions regarding assessment strategies. Work using multiple methodologies, such as quantitative and qualitative research, could contribute to this area.

Conclusion

Investigation of class size is important in the study of online learning efficacy and student success. In the current study, results supported the idea that there is no "one-size-fits-all" approach to online class size (Taft et al., 2019), as we found that students earned higher grades in courses with 30 students or fewer in STEM courses, as well as in upper division courses, but that class size made less difference in other courses. While the current study did not test why grades were higher in classes with 8-30 students compared to 31+ students, theory suggests that smaller classes are ideal for learning objectives higher on Bloom's Taxonomy, allow instructors to give more time and feedback per student, and lend toward more student participation in assignments such as discussions and writing. While it may be beneficial for certain kinds of online classes to be limited to 30 students or fewer, it is important to remember that the discussion of online learning efficacy and student success is multifaceted, with multiple stakeholders and perspectives to consider. Therefore, we recommend that readers take these results and use them in combination with other data to make informed decisions.

References

Ake-Little, E., von der Embse, N., & Dawson, D. (2020). Does class size matter in the university setting? *Educational Researcher, 49*(8), 595-605. doi: 10.3102/0013189X20933836

Angrist, J., & Lavy, V. (1999). Using Maimonides' rule to estimate the effect of class size on scholastic achievement. *Quarterly Journal of Economics, 114*(2), 533–575. https://doi.org/10.1162/003355399556061

Bettinger, E., Doss, C., Loeb, S., Rogers, A., & Taylor, E. (2017). The effects of class size in online college courses: Experimental evidence. *Economics of Education Review, 58*, 68-85.

Blatchford, P., Bassett, P., & Brown, P. (2011). Examining the effect of class size on classroom engagement and teacher–pupil interaction: Differences in relation to pupil prior attainment and primary vs. secondary schools. *Learning and Instruction, 21*(6), 715–730. http://doi.org/10.1016/j.learninstruc.2011.04.001

Cuseo, J. (2007). The empirical case against large class size: Adverse effects on the teaching, learning, and retention of first-year students. *Journal of Faculty Development, 21*(1), 5–21.

De Paola, M., Ponzo, M., & Scoppa, V. (2013). Class size effects on student achievement: Heterogeneity across abilities and fields. *Education Economics*, *21*(2), 135-153.

Dykman, C. A., & Davis, C. K. (2008). Online education forum, part three. A quality online educational experience. *Journal of Information Systems Education, 19*(3), 281-289.

Garrison, D. R., Anderson, T., & Archer, W. (1999). Critical inquiry in a text-based environment: Computer conferencing in higher education. *The Internet and Higher Education*, *2*(2-3), 87-105.

Kara, E., Tonin, M., & Vlassopoulos, M. (2020). Class size effects in higher education: Differences across STEM and non-STEM fields. IZA Discussion Papers, No. 12996, Institute of Labor Economics (IZA), Bonn.

Kokkelenberg, E., Dillon, M., & Christy, S. (2008). The effects of class size on student grades at a public university. *Economics of Education Review, 27*(2), 221–233.http://doi.org/10.1016/j. econedurev.2006.09.011

Lowenthal, P. R., Nyland, R., Jung, E., Dunlap, J. C., & Kepka, J. (2019). Does class size matter? An exploration into faculty perceptions of teaching high-enrollment online courses. *American Journal of Distance Education, 33*(3), 152-168.

Monks, J., & Schmidt, R. M. (2011). The impact of class size on outcomes in higher education. *The BE Journal of Economic Analysis & Policy*, *11*(1), 1-17.

Roby, T., Ashe, S., Singh, N., & Clark, C. (2013). Shaping the online experience: How administrators can influence student and instructor perceptions through policy and practice. *The Internet and Higher Education*, *17*, 29-37.

Russell, V., & Curtis, W. (2013). Comparing a large- and small-scale online language course: An examination of teacher and learner perceptions. *Internet and Higher Education, 16*, 1–13. doi:10.1016/j.iheduc.2012.07.002

Schmider, E., Ziegler, M., Danay, E., Beyer, L., & Bühner, M. (2010). Is it really robust?. *Methodology*: *European Journal of Research Methods for the Behavioral and Social Sciences, 6*(4), 147-151.

Scott, A. N., McNair, D. E., Lucas, J. C., & Land, K. M. (2017). From gatekeeper to gateway: Improving student success in an introductory biology course. *Journal of College Science Teaching*, *46*(4), 93-99.

Seaman, J. E., Allen, I. E., & Seaman, J. (2018). Grade increase: Tracking distance education in the United States. *Babson Survey Research Group*.

Sorenson, C. (2015). An examination of the relationship between online class size and instructor performance. *Journal of Educators Online, 12*(1), 140-159.

Taft, S. H., Kesten, K., & El-Banna, M. M. (2019). One size does not fit all: Toward an evidence based framework for determining online course enrollment sizes in higher education. *Online Learning, 23*(3), 188-233. doi:10.24059/olj.v23i3.1534

Taft, S. H., Perkowski, T., & Martin, L. S. (2011). A Framework for Evaluating Class Size in Online Education. *The Quarterly Review of Distance Education, 12*(3), 181-197.

Trammell, B. A., & LaForge, C. (2017). Common challenges for instructors in large online courses: Strategies to mitigate student and instructor frustration. *Journal of Educators Online, 14*(1), n1.