Suggested citation:
Lin, C.-H., Bae, J., & Zhang, Y. (2019). Online self-paced high-school class size and student achievement. *Educational Technology Research and Development*, 67, 317-336. doi: 10.1007/s11423-018-9614-x

Introduction

The number of enrollments in K-12 online courses in the United States tripled from 1.5 million in 2009-2010 (Watson, Murin, Vashaw, Gemin, & Rapp, 2010) to 4.5 million in 2014-2015 (Watson, Pape, Murin, Gemin, & Vashaw, 2015). Schools are continuing to expand their online course offerings, both to overcome school-level challenges and to meet student needs. For instance, they have been deployed as alternative courses to resolve scheduling conflicts; to offset shortages of highly qualified teachers, especially in Advanced Placement; and to provide a wider range of electives and other accelerated options for college-bound students (Watson et al., 2015).

The provision of online courses, however, creates another challenge. On the one hand, offering them helps to improve educational access, expand curricular choices, and increase high-quality learning opportunities (Barbour & Reeves, 2009; Berge & Clark, 2005; Cavanaugh & Blomeyer, 2007). On the other, such goals appear to be associated with extreme class sizes. Miron and Gulosino (2016) found that 356 students were enrolled in one virtual-school class, and that the average virtual-school class size was 35 – far above the U.S. averages for face-to-face class sizes, of 26.2 for primary schools, 25.5 for middle schools, and 24.2 for high schools (Coopersmith, 2009). At the other end of the scale, Watson et al. (2015, p. 56) found some virtual-school classes so small that their costs were “difficult to justify”; and of the online classes examined in the current study, 47% had 10 students or fewer (see Figure 1).

A meaningful answer to the question of whether the costs of small online classes are justified requires careful evaluation of the relationship between class size and students' learning outcomes. Accordingly, the purpose of this study is to examine this under-researched relationship between class size and online learners' performance, using data from a statewide virtual school. The main research questions are as follows:

1. What is the optimal class size (across all subjects) for self-paced courses at the high-school level?
2. What are the optimal class sizes in each subject for self-paced courses at the high-school level?

Effects of Class Size

Face-to-face Settings

A crucial indicator of classroom context, class size has been widely studied in face-to-face K-12 settings (Konstantopoulos & Sun, 2014). In general, researchers have favored smaller classes over larger ones when discussing class size's impact on teaching effectiveness, teacher-student interaction, and student achievement (Blatchford, Russell, Bassett, Brown, & Martin, 2007; Burruss, Billings, Brownrigg, Skiba, & Connors, 2009; Education Next, 2007; Ehrenberg, Brewer, Gamoran, & Willms, 2001; Hattie, 2005; Kokkelenberg, Dillon, & Christy, 2008). For example, a synthesis of more than 500 meta-analyses of class size suggested that small classes in face-to-face K-12 settings were usually perceived as desirable by both teachers and students, and as beneficial for students' learning (Hattie, 2005).

The literature has also consistently reported a relation between small class sizes and students' improvements in learning (for a review, see Kokkelenberg et al., 2008).

Specifically, students in small classes have been found to experience higher rates of teacher-student interaction than students in large classes do (Brühwiler & Blatchford, 2011; Zyngier, 2014). Students in smaller classrooms naturally gain more intense individual attention from teachers (Blatchford, Bassett, & Brown, 2011; Blatchford et al., 2007; Ehrenberg et al., 2001), which in turn improves their chances of learning (Konstantopoulos & Sun, 2014), of engaging in active learning (Blatchford et al., 2011), and of achieving high grades (Zyngier, 2014). In addition to fostering more active teacher-student interaction, small class sizes have been found to correlate with decreases in students' misbehavior and increases in their positive learning behaviors in class (Babcock & Betts, 2009; Bascia, 2010; Finn, Pannozzo, & Achilles, 2003).

From the perspective of teaching, reducing classroom sizes has been found to result in positive changes in the effectiveness of teaching styles and strategies, e.g., more individualization of teaching with the aim of increasing class engagement (Brühwiler & Blatchford, 2011), better interaction patterns, use of humor, and classroom organization/rule-setting (Harfitt, 2013), and increased teacher-parent interaction (Bascia, 2010). Perhaps unsurprisingly, K-12 teachers have consistently been found to strongly prefer small class sizes. For example, 81% of one group of teachers reported that they would prefer a reduction of class size over an increase in salary (Education Next, 2007). Conversely, large class sizes have been identified as a major driver of teacher attrition (Isenberg, 2010).

Despite the widely reported positive pedagogical effects of small classes in K-12 settings, researchers should be cautious about overemphasizing this positive impact, as the relationship between class size and student learning is not linear (Borland, Howse &

Trawick, 2005). Evaluations of optimal class size should take account of various factors (Hattie, 2005) including, but not limited to, ethnicity (Krueger & Whitmore, 2001) and grade level (Konstantopoulous & Sun, 2014). Additionally, it should be borne in mind that factors other than pedagogy often affect administrative decisions regarding class size. For instance, if instructors can teach larger classes without students' learning outcomes being adversely affected, it may be tempting to reduce overall educational costs through economies of scale (Hattie, 2008).

Online Settings

No prior studies of the relationship between class size and learning behaviors or outcomes in online K-12 settings appear to have been published, though the dearth of and need for such studies have been pointed out by some researchers (Miron & Gulosino, 2016; Zhang, Liu, & Lin, 2018). The following review of the literature on class size in online learning therefore contains only studies that were conducted in post-secondary settings.

Looking for the Optimal Class Size in Online Learning

Like research on class size in face-to-face environments, scholarship on online class size has suggested that it is not a stand-alone issue. Rather, it interacts with other components of online learning, and decision-making regarding online class sizes should therefore be context-dependent (Tomei, 2006; Zhang, Liu, & Lin, 2018). The existing scholarly discourse on online class size reflects two conflicting assumptions, both drawn from studies of higher education: namely, that small class sizes are better than large ones, or vice versa.

On the one hand, instructors have tended to argue – based on some combination of teacher- and student-level considerations – that the online environment requires smaller classes if it is to achieve desirable learning outcomes (Aragon, 2003; Arbaugh & Benbunan-Fich, 2005; Arzt, 2011; Orellana, 2006; Qiu, Hewitt, & Brett, 2012; Sorensen, 2014; Taft, Perkowski, & Martin, 2011; Tomei, 2006). At the teacher level, small online classes have been seen as keeping working loads at a reasonable level, and thus enabling a sufficient quantity and quality of feedback and student-teacher interaction, as well as adequate time for grading (Sorensen, 2015; Tomei, 2006). At the student level, meanwhile, online instructors have argued that large classes impede active student-student interactions as well as student-teacher ones (Arzt, 2011; Orellana, 2006; Taft et al., 2011).

Studies aimed at identifying the optimal online class size in post-secondary settings have recommended sizes in the range of 12 to 30 students. Tomei (2006), for example, estimated the optimal class size for a graduate-level course based on specific local conditions, including faculty teaching load (i.e., 85% of time to be spent on teaching, 10% on service, and 5% on research, with a three-course assignment per semester), and concluded that the ideal number of students would be 17 in a face-to-face format, but 12 in an online format. Orellana (2006) surveyed 131 teachers of online undergraduate or graduate courses, and found that while the respondents' actual average class size was 22.8, they perceived the optimal class size as 18.9 if their goal was to boost the existing level of interaction, or 15.9 if it was to maximize interaction. More recently, Qiu et al. (2012) asserted that 13 to 15 students was the optimal size for online classes, based on their finding of a significant positive association between class size and information overload. Approaching the matter from a pedagogical perspective, however, Arbaugh and Benbunan-Fitch (2005) concluded that the ideal online class size was considerably larger: between 25 and 30. And based on an examination of the social presence required for online courses, coupled with personal teaching experience, Argon (2003) also argued that 25 students would be an optimal online class size in higher education.

In contrast to instructors, administrators often favor the large classes that are made possible by the online environment's lack of physical-space limitations, as a means of lowering teaching costs (Sorensen, 2014; Tomei, 2006), though many institutions simply set online class sizes at the same level as their face-to-face counterparts (Mupinga & Maughan, 2008). One of the few studies of class size in K-12 online settings, Miron and Gulosino's (2016) report on more than 400 virtual schools, found an overall average class size of 35, but with a large variation in within-school average class size, from a minimum of 1.3 to a maximum of 356.

Online Class Size and Learning Outcomes

Research on the relationship between online class sizes and learning outcomes is scant. However, Qiu et al.'s (2012) study of graduate-level online courses found that large class sizes caused information overload among students, which decreased their engagement with learning and the quality of their note-taking and note-reading. Based on those findings, the authors concluded that a class size of 15 would be optimal.

Online Class Size and Subject Effects

Although no research on the relationship between online class size and subject matter in either secondary or post-secondary settings has yet been published, it is possible that the effect of class size varies across subjects in K-12 online learning environments (Zhang et al., 2018). According to Cavanaugh, Gillan, Kromrey, Hess, and Blomeyer (2004), math and science seem to be more difficult to learn in virtual schools than other subjects are. Other researchers have also suggested that world languages can be especially challenging for students to learn online (Lin & Warschauer, 2015; Lin, Zheng, & Zhang, 2017); and this has been borne out empirically in K-12 virtual-school contexts, where world-language students have been found to face more challenges and perform less successfully than their face-to-face counterparts (Cavanaugh, 2001; Oliver, Kellogg, & Patel, 2012).

In sum, there are three distinct gaps in the literature on class size in online settings. First, because the above-noted optimal ranges of class size were all derived from post-secondary online settings, it is unclear whether such numbers are also applicable to K-12 online settings – especially given that online learning requires a high level of self-regulation (Lin, Zhang, & Zheng, 2017; Lin, Zheng, et al., 2017) and students at the secondary level are still developing such skills (Elliot, Dweck, & Yeager, 2017). Second, none of the studies reviewed above looked directly at the relationship between online class size and learning outcomes. As previously mentioned, it is problematic (but commonplace) for schools to simply pre-set a size for their online classes without considering contextual factors, including especially students' learning achievement. And third, all prior research on online class size has utilized small sample sizes and focused on a narrow range of subjects in higher education, without considering possible differences across subjects. Therefore, estimating the optimal online class sizes for secondary students is still uncharted territory, and doubly so if budgetary concerns and between-subject differences are also taken into account.

Methods

Research Site

This study utilized a dataset covering all students enrolled during the 2013-14 school year at an accredited state-wide virtual school in the Midwestern U.S. Though always officially referred to as a school, it in fact comprises a supplementary program of a la carte online courses for students who are all enrolled elsewhere, either in physical or cyber schools, mostly though not exclusively within the same state.

During the year in question, the virtual school's courses used digital texts in which learning was self-paced, and communications between and among teachers and students were asynchronous. Online instructors in this school were not required to devise curricula, but only to supplement fully designed online courses that were provided to them. These courses, which were designed based on research on effective online teaching and learning strategies, were reviewed and certified by a third-party quality assurance program (Quality Matters) as well as by the virtual school for their compliance with its own quality standards. All such courses followed the state's and/or national curriculum standards and were taught by state-certified teachers with endorsements for the relevant content area and grade level. Representative instructional practices included guiding and supporting students' learning through communication with them, their parents/guardians, and their mentors at their own respective schools; providing progress-monitoring, informative feedback, and personalized supports throughout the term; and facilitating specific student tasks and activities, especially class discussions. Instructional materials, tools and assignments included streaming audio and video, computer animations, email, chat rooms, digital portfolios, individual and team projects, and discussion forums, and shared the general aim of offering problem-based learning opportunities that would enhance interaction and collaboration. However, the school's administrators reported that due to the nature of self-paced online learning, students' interactions with their peers were limited.

At the time of writing, none of the course providers had granted the research team access to any of their courses; though problematic, this is commonplace in K-12 online learning (Barbour, 2017). Thus, instructional approaches could not be evaluated.

Data

Initially, the data included 21,253 enrollment records from 12,445 students, but students who withdrew from or dropped a course were excluded, as they did not receive any grade. This yielded a sample size of 20,540 records relating to 12,032 students and 233 courses in six subjects, taught by 155 instructors. The enrollment data for each individual student included the name(s) of the course(s) taken, the semester(s) in which they were enrolled, their grades, and their instructors' unique identifying numbers. The dataset also included information about each student's local school, including its name and location; the name of the student's local-school mentor; and the student's gender and self-reported reason for taking online courses.

Measures

For purposes of this study, *course* refers to the subject a student is enrolled in (e.g., Algebra, Geometry, beginning Chinese), and *class size* as the number of students in one course section. An online instructor might teach a course in multiple sections (e.g., Algebra section I and section II). A given student could take one or more courses from the school in one, two or all three semesters of the year (i.e., Fall, Spring and Summer).

Gender. Among the 12,032 students who completed at least one course during the year, 5,324 (or 44.2%) were males and 6,708 (or 55.8%) were females.

Grade level. All courses offered in the virtual school were at high-school levels, and the great majority of its student population was of high-school age. Some middle-school students were enrolled subject to approval from their own schools, but their exact numbers cannot be known. This was because the virtual school asked its students for their demographic information, but did not require them to provide it; and their response rate to the school's question about their current grade level was below 30%. However, a survey with a 29% response rate that two of the researchers conducted in the same virtual school in spring 2014 found that the student body consisted of 9% middle schoolers, 15% 9th graders, 31% 10th graders, 26% 11th graders, and 20% 12th graders (Lin, Zheng, et al., 2017).

Reasons for enrollment. The dataset included students' reasons for enrollment in online classes, based on a multiple-choice question with five possible answers: 1) the course being unavailable at local schools, 2) credit recovery, 3) the learning preferences of the student, 4) scheduling conflicts, and 5) other. By far the most common reason given was the local unavailability of similar courses (46.7%; see Table 1). A further 15.2% responded that their enrollment was due to their personal learning preferences, while 11.3% cited scheduling conflicts. Only 8.1% of students stated that their enrollment was for credit recovery, and nearly one in five students (18.8%) did not provide any clear reason for why they enrolled, except to the extent that it did not fit into any of the other four categories.

Learning outcomes. Learning outcomes were the course grades reported by the virtual school to the students' own schools at the end of each semester. All courses with the same name shared the same assessment regime across their different sessions. In other words, the design of the assessment was at the course level, not at the instructor- or class level. As noted above, all the courses were certified by Quality Matters; and one aspect of such certification involves standards of assessment and measurement. Specifically, each course's assessment strategy must be capable of evaluating the effectiveness of student learning based on its stated learning outcomes.

In all subjects, each course grade consisted of a mixture of the scores from auto-graded and instructor-graded assignments, transformed into a percentage format. To a certain extent, this dual grading system ensured the consistency of the assessment system.

Class size. The size of a class (i.e., section, in the case of courses with two or more sections) was calculated as the sum of the students who had completed it, regardless of whether they had passed or failed. However, as noted above, students who dropped a class were not included when calculating its size.

The virtual school's decisions regarding online class size were based on estimated numbers of enrollments, and its administrative staff sought to ensure that the sections of the same course were of roughly equal sizes. In some instances, classes ended up smaller than the administrators had anticipated, but the school did not cancel any classes for this reason alone.

The sizes of the 1,380 class sections in which at least one of our student participants was enrolled ranged from one student to 60 (see Figure 1). There were five or fewer students in 29% of the online classes in our sample; six to 10 students in 18% of the classes; 11 to 20 students in another 20%; and 21 or more in 33%. The average class size was 14.88, with a standard deviation of 12.09.

Subjects. The virtual school's administrators had divided the subjects taken by the students in the dataset into the following categories: 1) English (making up 7.54% of all enrollments); 2) foreign languages (21.51%); 3) science (11.73%); 4) math (14.51%); 5) social science (19.03%); and 6) other subjects (25.68%; e.g., art, business, and physical education).

Data Analysis

Stata 14 software was used to conduct all the quantitative analyses for this study. Hierarchical linear modeling (HLM) with maximal likelihood estimates (Singer & Willett, 2003) was employed to examine all of the research questions. HLM is a type of regression model used to account for correlated errors in nested data structures. In this case, students were nested in different class sections, which violates an assumption of multiple regression: that individuals should be independent. As compared to multiple regression, HLM provides a more accurate form of estimation that accounts more rigorously for sources of statistical error.

A series of five two-level models were fit to answer the research questions regarding class size. Level 1 consisted of students' enrollment records, and level 2 of each class section. In all models, the students' final grades were entered as the dependent variable. The majority (54%) of the students in the dataset took only one course during the school year in question, and in cases where a student had multiple enrollments, only his/her first record was used, to ensure that no students' learning outcomes were biased upward due solely to their greater familiarity with the virtual-school environment. However, all the enrollments of students who had multiple enrollments were still included in calculations of class size. In addition, given that HLM requires a considerate numbers of students in level 1 of each model (McNeish & Stapleton, 2016), classes with fewer than 10 students were removed from HLM analysis. After these removal criteria were applied, the final sample size was 10,648 students.

Model 1 was the unconditional model, essential for determining overall section-related random effects. Model 2 added class size as a level-2 variable, which enabled examination of the linear relationship between class size and learning outcomes. However, since this relationship is not necessarily linear, Model 3 added a class-size quadratic term and examined whether the relationship between class size and learning outcomes was a parabolic curve. Likewise, Model 4 added a class-size cubic term to examine whether the relationship between class size and learning outcomes was nonlinear and nonparabolic.

To compare which model fit the data best, the researchers used the two types of information criterion that are most frequently utilized in the HLM literature: Akaike's information criterion (AIC; Akaike, 1998) and the Bayesian information criterion (BIC; Schwarz, 1978). Both are used comparatively, to evaluate which of two or more models has the best combination of fit and complexity, with the model whose index value is closest to zero being the best-fitting.

The use of pre-determined shapes to describe the relationship between class size and final grades, as in Models 2, 3 and 4, may not result in accurate descriptions of this relationship. To overcome this issue, Model 5 used fractional polynomial (FP) analysis with multilevel modeling. FP regression provides a flexible parametric method for modeling non-linear relationships using the smallest possible number of parameters, and allows the use of logarithms, non-integer powers, and repeated powers (Royston & Altman, 1997). It was performed in Stata using the *fp* command.

After identifying the model that was best able to describe the relationship between class size and student learning outcomes in general, we used that model to examine the effects of class size on student learning outcomes for each subject.

Results

Optimal Class Size across All Subjects

The unconditional model, or Model 1, estimated the overall mean attainment across classes as 71.59 points out of 100 (see Table 2). The between-class (level-2) variance in attainment was estimated as 185.71, and the within-class/between-student (level-1) variance in attainment as 617.77. The total variance was the sum of between-class and within-class/between-student variance, i.e., 803.48. Intraclass correlation (i.e., between-class variance divided by total variance) was 0.231, indicating that 23.1% of the observed variance in attainment was attributable to differences between classes. It should be noted at this point that our dataset did not include entry scores, so the class effect was not value-added.

Model 2, which was used to examine a linear class-size effect, found that each one-person increase in class size resulted in a 0.32-point increase in average final grade, $p < .001$, and the effect size was 0.024. This linear relationship is illustrated in Figure 2.

In Model 3, the quadratic model, both class size (B = 1.10, $p < 0.001$) and class size squared (B = -0.01, $p < .001$) were significantly correlated with average final grades, indicating that the parabola shape fit the relationship between class sizes and learning outcomes. The effect size for class size linear term increased to was 0.084, and the effect size for class size quadratic term was -0.0009. Both indicators of overall model fit were higher for Model 2 (AIC = 99749.40, BIC = 99778.50) than for Model 3 (AIC = 99737.86, BIC = 99774.23), indicating that the latter fit the data better than the former. This suggests that the non-linear relationship between class size and final grade was a reverse-U shape, with peak academic performance occurring at class sizes of around 45 students (see Figure 3) when all academic subjects were considered collectively.

In Model 4, class size was not significantly related to average final grades (B = 0.69, $p = .41$), and neither were the class-size quadratic and cubic terms, suggesting that this model did not fit the relationship between class size and learning outcomes. Model 4's AIC (99739.60) was lower than that of Model 2 (99749.40), but its BIC (99783.24) was higher than Model 2's BIC (99778.50), confirming that Model 4 did not fit the relationships in the data better than Model 3 did.

For Model 5, before performing FP, it was necessary to determine the best powers for each parameter. This required fitting 44 models, as shown in Table 3. FP comparison indicated that the second model, with power of two and two, was significantly different at the 0.05 level both from the other FP models and from the linear model. Model 5 therefore utilized the powers reported (i.e., two and two), and the results of HLM are also presented in Table 4. The first class-size quadratic term had a coefficient of 0.02, and the second, of -0.0003. When a power repeats in an FP analysis, it is multiplied by another ln(x). The equation for class size is as follows:

$$\text{grades} = 60.505 + 0.082 * \text{class size}^2 - 0.019\ \text{class size}^2 * \ln(\text{class size}) \qquad \text{equation 1}$$

The resulting curve of Model 5 is shown in Figure 4, and suggests that increasing a typical class' size had a positive effect on its students' learning outcomes until such size reached around 45 people, after which it had a negative effect. The overall shapes of Figures 3 and 4 were very similar, each being a reverse-U with a peak around 45; but after the peak point, Figure 4 had a sharper drop than Figure 3 did. Model 5 had a smaller AIC and BIC than either Model 2 or Model 3, suggesting that it fit the data better than they did.

Optimal Class Size for Each Subject

Because the FP-based model best described the relationship between class size and learning outcomes for all six academic subject areas taken together, FP analysis with multilevel modeling was next performed for each subject separately. To allow each subject different powers for different class sizes, the powers for each of these FP analyses were determined by the model with the lowest deviance. The results are presented in Table 5 and Figure 5.

English. The model results for English classes revealed a trend in which final grades increased as class size increased, but this relationship was not significant (B_1 = -387.456, $p = 0.166$; $B_2 = 0.000$, $p = 0.288$). In other words, English class size did not have an impact on students' final grades.

Foreign languages. In foreign languages, student achievement was highest in classes with 15 students, and above that level, the relationship between class sizes and final grades was consistently negative: i.e., a person's final grade tended to decrease as class size increased ($B_1 = -581.240$, $p = 0.525$ and $B_2 = 556.506$, $p = 0.57$). However, the relationship between foreign-language class sizes and final grades was not significant, as shown in Table 5.

Science. The relationship of science-class size to final grades was similar to that across all subjects, but the position of its peak was different ($B_1 = 0.003$, $p = 0.05$; $B_2 = -0.001$, $p = 0.06$). Specifically, students' final grades increased as class size increased until the latter reached 35; then, their final grades decreased as class size increased further. This relationship was statistically significant.

Math. The relationship between class sizes in math and students' performance was also parabolic: final grades increased as class size increased up to a maximum of 38 students, but decreased if class size rose beyond that point ($B_1 = 0.004$, $p = .004$; $B_2 = -0.001$, $p = 0.006$). This relationship was statistically significant.

Social science. The relationship between class size and students' final grades in social-science subjects was very similar to that in math classes ($B_1 = 522.413$, $p < 0.001$; $B_2 = -353.297$, $p < 0.001$), albeit with the turning point falling around a size of 42 students, as opposed to 38 in math classes. This relationship was significant.

Other subjects. The relationship between class sizes and students' final grades in classes that did not fit into any of the above five categories (e.g., art) was very similar to the previously described situations in social-science and math classes ($B_1 = 2090.210$, $p =$

0.011; $B_2 = 1323.329$, $p = 0.005$), albeit with the turning point falling around 35 students. This relationship was also significant.

Discussion

This study represents an important contribution to the body of research on class size in K-12 learning, and extends it to an online K-12 learning context. Specifically, findings from the present study address common false expectations about online-class size in K-12 settings. In contrast to face-to-face settings, online learning is unconstrained by physical space, and this sometimes results in extremely large K-12 classes (e.g., more than 300 students in one section: see Miron & Gulosino, 2016). Meanwhile, as part of efforts to provide niche courses that are not available in certain geographic areas or not taught in physical schools at all, online K-12 learning sometimes features extremely small classes, with just one or two students in some sections (see Watson et al., 2015; also see Figure 1). The results of the present research suggest that extreme online-class sizes, in either direction, are likely to have negative impacts on students' learning outcomes.

On the one hand, this implies that class-size reduction might be helpful, if the original size is above a certain threshold, which aligns with findings from the majority of peer-reviewed papers that support class-size reduction (Zyngier, 2014). However, the present findings are considerably more robust than those of studies using observational data in traditional K-12 settings, due to an unusual characteristic of our study context: the fact that the online instructors were not responsible for the curriculum, making it extremely unlikely that they taught classes differently because of differences in student numbers. As Ehrenberg et al. (2001) noted, small class sizes can affect teaching in two ways. First, instruction may be improved; and second, certain teaching practices may become more effective. In the current study, given that the curriculum was fixed, the effects of class-size differences that were identified should have stemmed mainly from such differences themselves.

On the other hand, in line with the findings of Rice (1999) and Dustmann, Rajah, and van Soest (2003), the current study's results suggest that reducing class size, when it is already at or below a certain threshold, may actually have a negative impact on students' learning outcomes. One possible explanation is that smallness may impede learner-instructor interactions, which prior research has shown to have a positive effect on learning, after controlling for students' individual differences in motivation and learning strategies (Lin, Zheng, et al., 2017).

It is also worth noting that the optimal class sizes for self-paced courses reported in this study (i.e., 38 in math, 42 in social science, and 35 in other non-language subjects) are much higher than the average sizes of face-to-face high-school classes (i.e., 24.2: see Coopersmith, 2009). This calls into question prior studies' recommendations that online classes always be smaller than face-to-face ones (e.g., Qiu, Hewitt, & Brett, 2012; Sorensen, 2014). This discrepancy could relate to special characteristics of the present study's virtual-school research setting: for instance, the previously discussed lack of responsibility for curriculum development on the part of its teachers, which could have reduced their preparation time and thus enabled them to effectively teach larger classes than might normally be possible.

This study raises several policy issues. Its more striking findings include the sharp differences among academic subject areas in terms of the impact of class size on students' final grades. This confirms that decision-makers should avoid one-size-fits-all approaches to setting class sizes, and instead remain sensitive to different subject areas' divergent requirements, as Rice (1999) contended. However, the present findings should not be used to generate prescriptive guidelines for optimal class sizes in different subjects; nor are they in any way intended to suggest that class size be used simplistically as a measure of school quality. Rather, they indicate that when an online class is of an extreme size – whether large or small – it may be detrimental to students' learning, at least relative to similar classes of a more normal size.

This study's results also imply that policymakers should be more cognizant of the balance that must be struck between maximizing educational access and having small classes. Expanding access and providing curricular choices are among the primary purposes of offering online courses to K-12 students (Barbour & Reeves, 2009; Berge & Clark, 2005; Cavanaugh & Blomeyer, 2007). However, these purposes should not be achieved at the expense of students' learning outcomes. As illustrated by the current study, being members of a small class may in certain circumstances actually be harmful to students' learning. When offering niche courses that are unavailable in students' own schools, online-learning providers should consider these potential disadvantages of small classes, alongside better-known issues such as cost.

Conclusions

This study has provided empirical evidence of both general and subject-specific class-size effects in self-paced online high-school courses. Its examination of the relationship between class size and students' learning outcomes in multiple subjects was a major departure from the prior literature, which has focused primarily on math (e.g., Li

& Konstantopoulos, 2016). The relationships between class sizes and students' final grades can be depicted as a reverse-U shape for math, social science, and other subjects (e.g., arts). In these areas, as class size increased, final grades increased until the peak in this curve was reached; and once the class size rose to any level beyond the peak, it had a negative impact on students' final grades. In English, foreign languages, and science, however, the impact of class size on learning outcomes was non-significant.

These unique findings contribute to the study of online learning in multiple ways. To the best of the researchers' knowledge, this is the first study that examines class sizes' relation to academic outcomes in an online K-12 setting. Amid ever-increasing numbers of K-12 students taking courses online, and the prevalence of very small and very large classes in virtual schools, this study provides clarification of key issues for policymakers and school administrators tasked with setting and modifying class sizes. In addition, its results refine and extend findings derived from face-to-face settings. For the most part, the results of HLM analysis in the current study indicated a significant, non-linear association between self-paced online classes' sizes and their students' final grades. The identification of the parabolic nature of this relationship may help to disentangle the mixed results obtained from past research in face-to-face settings.

Several limitations of the current research need to be noted. First, it used observational data, making it very difficult to isolate confounding or omitted variable bias. The assignment of students and teachers to particular classes, for example, could have been influenced by students' characteristics – such as motivation and previously identified abilities – as well as by interventions on the part of instructors, principals, and parents. These omitted variables relating to student placement are difficult to measure, and accordingly, causal inferences should be drawn from this study only with caution. In future research, this limitation can be addressed in one of two ways: through a true experimental design in which online students are randomly assigned to classes of different sizes, or through modes of statistical analysis that can appropriately account for omitted variables (Ehrenberg et al., 2001). Instrumental variables, used in combination with regression discontinuity, have the potential to address the unobserved-variable issue (see Konstantopoulos & Shen, 2016 for example), but a strong instrument that is unrelated to unobserved variables had yet to be developed at the time of writing.

Second, the class-size effects we reported in this study with regard to self-paced courses were relatively small, in keeping with the effect sizes reported in many meta-analyses (see Hattie, 2005). Among many factors that have been shown to influence students' learning outcomes (e.g., feedback, instruction, and their own study skills), class size has consistently been ranked at the lower end of the spectrum (Hattie, 2005, 2008). The inherent limitations of this study's data prevented examination of other factors that might have explained some of the observed differences between classes. Although reason-for-enrollment data was collected, a parsimonious model that fit this and the other data well without violating statistical assumptions could not be identified within the time available. And third, this study only examined online learning in a self-paced context, meaning that its findings may not be generalizable to other online delivery modes such as cohort-based online learning.

Future studies should aim to further disentangle the effects of class size by considering additional factors. First, the difficulty levels of courses within a subject (e.g., Algebra versus Advanced Placement Calculus) should be included as model covariates, to help determine whether the widely theorized positive effects of small class sizes are present, but masked by the examination of hard and easy classes jointly. Second, course-design factors that affect the amounts of time teachers spend preparing for classes and providing feedback should be measured, as this could provide a clearer understanding of the mechanism(s) through which class size affects students' learning outcomes (e.g., Rice, 1999). Third, credit-recovery students may have lower levels of self-regulation skills than other categories of students (Watson & Gemin, 2008), which in turn could affect their learning outcomes. Therefore, including the percentage of credit-recovery students in each class might shed further light on the small effect sizes reported in this study. Lastly, in addition to final grades, we recommend that future research on the effects of class size use other types of outcome variables, such as student satisfaction and the amount of student-teacher interaction. Broader assessments of this kind will be especially important when research in this area is extended to include other types of online-course settings.

References

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In *Selected Papers of Hirotugu Akaike* (pp. 199–213). Springer, New York, NY.

Aragon, S. R. (2003). Creating social presence in online environments. *New Directions for Adult and Continuing Education*, *2003*(100), 57–68. https://doi.org/10.1002/ace.119

Arbaugh, J. B., & Benbunan-Fich, R. (2005). Contextual factors that influence ALN effectiveness. *Learning Together Online: Research on Asynchronous Learning Networks*, *1*, 123–144.

Arzt, J. (2011). Online Courses and Optimal Class Size: A Complex Formula. *Online Submission*.

Babcock, P., & Betts, J. R. (2009). Reduced-class distinctions: Effort, ability, and the education production function. *J. Urban Econ.*, *65*(3), 314–322.

Barbour, M. K. (2017). K-12 online learning and school choice : Growth and expansion in the absence of evidence. In R. A. Fox & N. K. Buchanan (Eds.), *The Wiley handbook of school choice* (pp. 421–440). Hoboken, NJ: Wiley Blackwell.

Barbour, M. K., & Reeves, T. C. (2009). The reality of virtual schools: A review of the literature. *Computers & Education*, *52*, 402–416. https://doi.org/10.1016/j.compedu.2008.09.009

Bascia, N. (2010). Reducing class size: What do we know. *Ontario Institute for Studies in Education*.

Berge, Z. L., & Clark, T. (2005). *Virtual schools: Planning for success*. New York: Teachers College Press.

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.

Blatchford, P., Russell, A., Bassett, P., Brown, P., & Martin, C. (2007). The effect of class size on the teaching of pupils aged 7 – 11 years. *School Effectiveness and School Improvement*, *18*(2), 147–172. https://doi.org/10.1080/09243450601058675

Brühwiler, C., & Blatchford, P. (2011). Effects of class size and adaptive teaching competency on classroom processes and academic outcome. *Learning and Instruction*, *21*(1), 95–108.

Burruss, N. M., Billings, D. M., Brownrigg, V., Skiba, D. J., & Connors, H. R. (2009). Class Size as Related to the Use of Technology, Educational Practices, and Outcomes in Web-Based Nursing Courses. *Journal of Professional Nursing*, *25*(1), 33–41. https://doi.org/10.1016/j.profnurs.2008.06.002

Cavanaugh, C. (2001). The effectiveness of interactive distance education technologies in K-12 learning: A meta-analysis. *International Journal of Educational Telecommunications*, *7*(1), 73–88.

Cavanaugh, C., & Blomeyer, R. L. (2007). *What works in K-12 online learning*. Eugene, OR: International Society for Technology in Education.

Cavanaugh, C., Gillan, K. J., Kromrey, J., Hess, M., & Blomeyer, R. (2004). The effects of distance education on K-12 student outcomes: A meta-analysis. *Learning Point Associates/North Central Regional Educational Laboratory (NCREL)*.

Coopersmith, J. (2009). *Characteristics of public, private, and Bureau of Indian Education elementary and secondary school teachers in the United States: Results from the 2007–08 Schools and Staffing Survey.* Washington, D.C: National Center for Education Statistics.

Education Next. (2007). *Program on Education Policy and Governance (PEPG) 2007 Survey*. Hong Kong. Retrieved from http://educationnext.org/files/EN-PEPG_Complete_Polling_Results.pdf

Ehrenberg, R. G., Brewer, D. J., Gamoran, A., & Willms, J. D. (2001). Class size and student achievement. *Psychological Science in the Public Interest*, *2*(1), 1–30.

Elliot, A. J., Dweck, C. S., & Yeager, D. S. (2017). *Handbook of competence and motivation: Theory and application* (2nd ed.). Guilford Publications.

Finn, J. D., Pannozzo, G. M., & Achilles, C. M. (2003). The "Why's" of Class Size: Student Behavior in Small Classes. *Rev. Educ. Res.*, *73*, 321–368.

Harfitt, G. J. (2013). Why 'small'can be better: An exploration of the relationships between class size and pedagogical practices. *Research Papers in Education*, *28*(3), 330–345.

Hattie, J. (2005). The paradox of reducing class size and improving learning outcomes. *International Journal of Educational Research*, *43*(6), 387–425. https://doi.org/10.1016/j.ijer.2006.07.002

Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.

Isenberg, E. P. (2010). *The effect of class size on teacher attrition: Evidence from class size reduction policies in New York State*. US Census Bureu Center for Economic Studies.

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

Konstantopoulos, S., & Shen, T. (2016). Class size effects on mathematics achievement in Cyprus: evidence from TIMSS. *Educational Research and Evaluation*, *22*(1–2), 86–109. https://doi.org/10.1080/13803611.2016.1193030

Konstantopoulos, S., & Sun, M. (2014). Are teacher effects larger in small classes? *School Effectiveness and School Improvement*, *25*(3), 312–328.

Li, W., & Konstantopoulos, S. (2016). Class Size Effects on Fourth-Grade Mathematics Achievement: Evidence From TIMSS 2011. *Journal of Research on Educational Effectiveness*, *0*(0), 1–28. https://doi.org/10.1080/19345747.2015.1105893

Lin, C.-H., & Warschauer, M. (2015). Online foreign language education: What are the proficiency outcomes? *The Modern Language Journal*, *99*(2), 394–397. https://doi.org/10.1111/modl.12234_1

Lin, C.-H., Zhang, Y., & Zheng, B. (2017). The roles of learning strategies and motivation in online language learning: A structural equation modeling analysis. *Computers & Education*, *113*, 75–85. https://doi.org/10.1016/j.compedu.2017.05.014

Lin, C.-H., Zheng, B., & Zhang, Y. (2017). Interactions and learning outcomes in online language courses. *British Journal of Educational Technology*, *48*, 730–748. https://doi.org/10.1111/bjet.12457

McNeish, D. M., & Stapleton, L. M. (2016). The effect of small sample size on two-level model estimates: A review and illustration. *Educ. Psychol. Rev.*, *28*(2), 295–314.

Miron, G., & Gulosino, C. (2016). *Virtual schools report 2016: Directory and performance review*. Retrieved from Boulder, CO.

Mupinga, D. M., & Maughan, G. R. (2008). Web-based instruction and community college faculty workload. *College Teaching*, *56*(1), 17–21.

Oliver, K., Kellogg, S., & Patel, R. (2012). An investigation into reported differences between online foreign language instruction and other subject areas in a virtual school. *CALICO Journal*, *29*(2), 269–296.

Orellana, A. (2006). Class size and interaction in online courses. *Quarterly Review of Distance Education*, *7*(3), 229–248.

Qiu, M., Hewitt, J., & Brett, C. (2012). Online class size, note reading, note writing and collaborative discourse. *International Journal of Computer-Supported Collaborative Learning*, *7*(3), 423–442. https://doi.org/10.1007/s11412-012-9151-2

Rice, J. K. (1999). The Impact of Class Size on Instructional Strategies and the Use of Time in High School Mathematics and Science Courses. *Educational Evaluation and Policy Analysis*, *21*(2), 215–229. https://doi.org/10.3102/01623737021002215

Royston, P., & Altman, D. G. (1997). Approximating statistical functions by using fractional polynomial regression. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *46*(3), 411–422. https://doi.org/10.1111/1467-9884.00093

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford university press.

Sorensen, C. (2014). Classrooms Without Walls: A Comparison of Instructor Performance in online Courses Differing in Class Size. *J. Online Learn. Teach.*, *10*(4), 569.

Taft, S. H., Perkowski, T., & Martin. (2011). A framework for evaluating class size in online education. *The Quarterly Review of Distance Education*, *12*(3), 181.

Tomei, L. (2006). The impact of online teaching on faculty load: Computing the ideal class size for online courses. *Journal of Technology and Teacher Education*, *14*(3), 531–541.

Watson, J., & Gemin, B. (2008). *Promising practices in online learning: Using online learning for at-risk students and credit recovery*. Vienna, VA: International Association for K-12 Online Learning.

Watson, J., Murin, A., Vashaw, L., Gemin, B., & Rapp, C. (2010). Keeping pace with K-12 online and blended learning: An annual review of policy and practice. Retrieved from http://www.kpk12.com/wp-content/uploads/KeepingPaceK12_2010.pdf

Watson, J., Pape, L., Murin, A., Gemin, B., & Vashaw, L. (2015). *Keeping pace with K-12 digital learning: An annual review of policy and practice*. Retrieved from http://www.kpk12.com/wp-content/uploads/Evergreen_KeepingPace_2015.pdf

Zhang, Y., Liu, H., & Lin, C.-H. (2018). Research on class size in online K-12 learning. In K. Kennedy & R. Ferdig (Eds.), *Handbook of Research on K-12 Online and Blended Learning* (2nd ed., pp. 273–283). Pittsburgh, PA.

Zyngier, D. (2014). Class size and academic results, with a focus on children from culturally, linguistically and economically disenfranchised communities. *Evidence Base*, *1*, 1–23.

Table 1. Reasons for Enrollment

Enrollment reasons	Frequency	Percentage
Course unavailable locally	9,491	46.69
Credit recovery	1,655	8.06
Learning preferences	3,116	15.17
Scheduling conflict	2,310	11.25
Other reasons	3,868	18.83
Total	20,540	100

Table 2. Hierarchical Linear Modeling for Class Size and Final Grades

	Final grades			
	Model 1	Model 2	Model 3	Model 4
Class size		0.32***	1.10***	0.69
		(5.73)	(5.07)	(0.83)
Class size quadratic term			-0.01***	0.00
			(-3.70)	(0.07)
Class size cubic term				-0.00
				(-0.51)
_cons	71.59***	63.73***	53.87***	57.12***
	(123.19)	(42.78)	(17.67)	(8.11)
Variance (between class)	185.71***	175.58***	170.90***	170.84***
	(74.61)	(73.17)	(72.29)	(72.29)
Variance (within class)	617.77***	617.68***	617.74***	617.4***
	(452.36)	(452.48)	(452.47)	(452.47)
Observations	10,648	10,648	10,648	10,648
Intraclass correlation	0.231			
AIC	99779.56	99749.40	99737.86	99739.60
BIC	99801.38	99778.50	99774.23	99783.24

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3. Fractional Polynomial Comparisons for Class Size

Class size	Degree of freedom	Deviance	*p*	Powers
omitted	0	99773.56	0.000	
linear	1	99741.40	0.003	1
m=1	2	99733.02	0.045	.5
m=2	4	99727.23	--	2 2

Table 4. Fractional Polynomial Analysis for Class Size

	Model 5
Class size term 1	0.082***
	(0.015)
Class size term 2	-0.019***
	(0.003)
_cons	60.505***
	(1.809)
Random effect	
Variance (between-class)	170.725***
	(12.14)
Variance (within-class)	617.737***
	(8.77)

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5. Fractional Polynomial Analysis of Class Size for Each Subject

	Final grades					
	English	Foreign language	Science	Math	Social science	Others
Class size 1 power	-2	-2	3	3	-1	-2
class size 1	-387.456	-581.240	0.003	0.004**	522.413***	2090.210*
	(-1.39)	(-0.64)	(1.93)	(2.86)	(3.65)	(2.53)
Class size 2 power	3	-2	3	3	-1	-2
class size 2	0.000	319.655	-0.001	-0.001**	-363.297***	-1323.329**
	(1.06)	(0.57)	(-1.86)	(-2.77)	(-4.15)	(-2.80)
Constant	54.748***	72.434***	67.385	59.076***	101.441***	79.470***
	(13.64)	(30.14)	(28.30)	(23.83)	(18.06)	(50.16)
Variance between classes	2.818***	2.545***	2.459***	2.716***	2.078***	2.268***
	(25.88)	(35.64)	(23.63)	(32.42)	(20.57)	(30.16)
Variance within class	3.452***	3.205***	3.242***	3.269***	3.162***	3.196***
	(129.40)	(213.36)	(160.15)	(173.03)	(205.90)	(248.91)
Observations	786	2,383	1,338	1,553	2,281	3,257

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

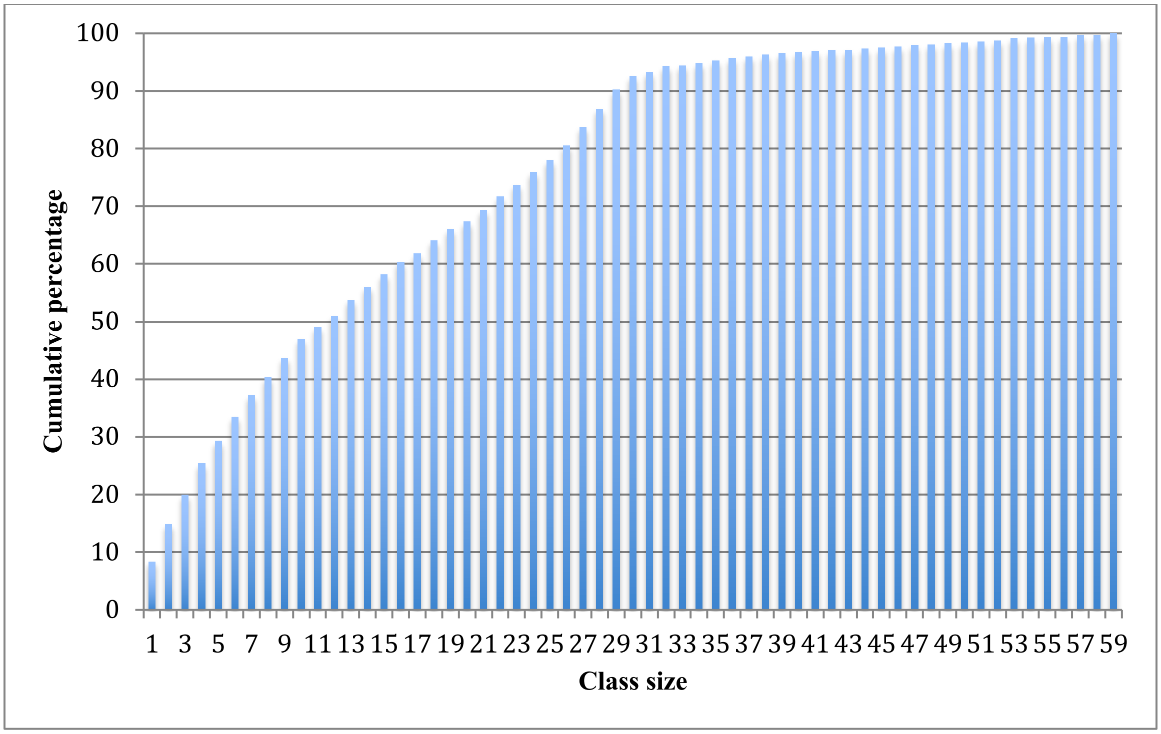

Figure 1. Cumulative Percentage of Classes by Size

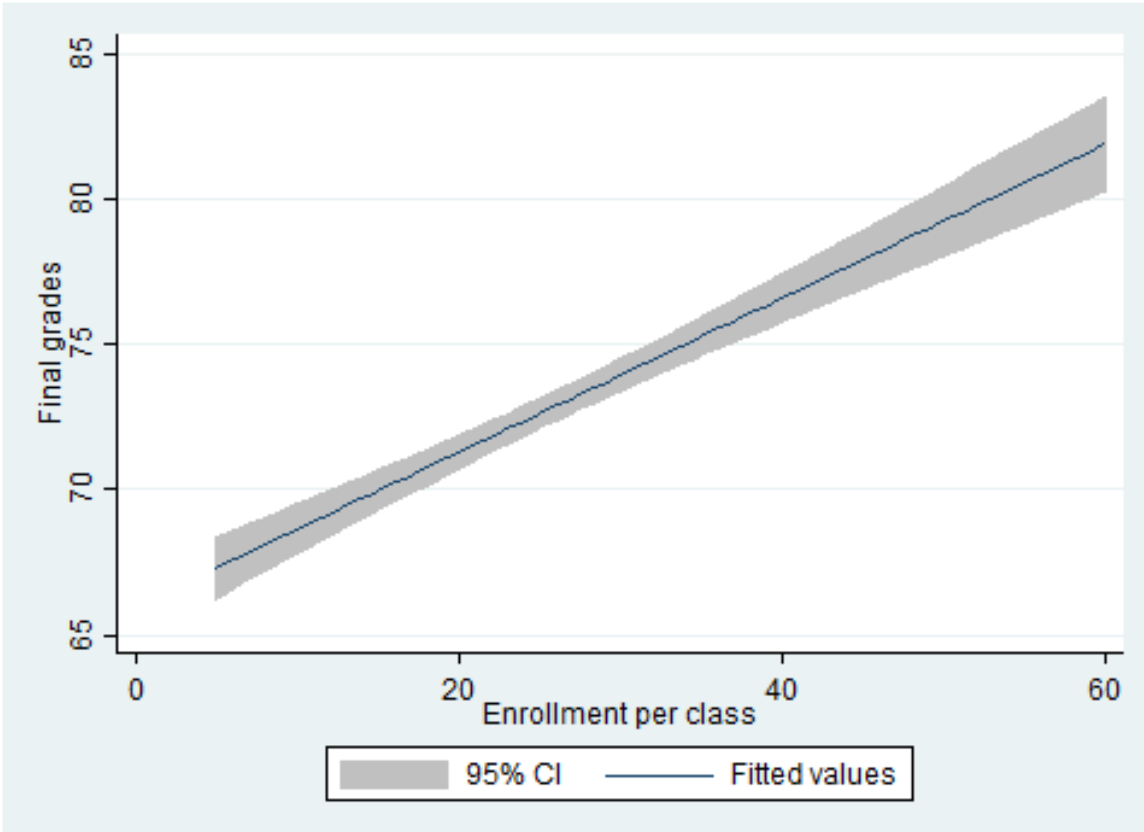

Figure 2. Linear Regression Fitting Line for the Effect of Class Size on Final Grades

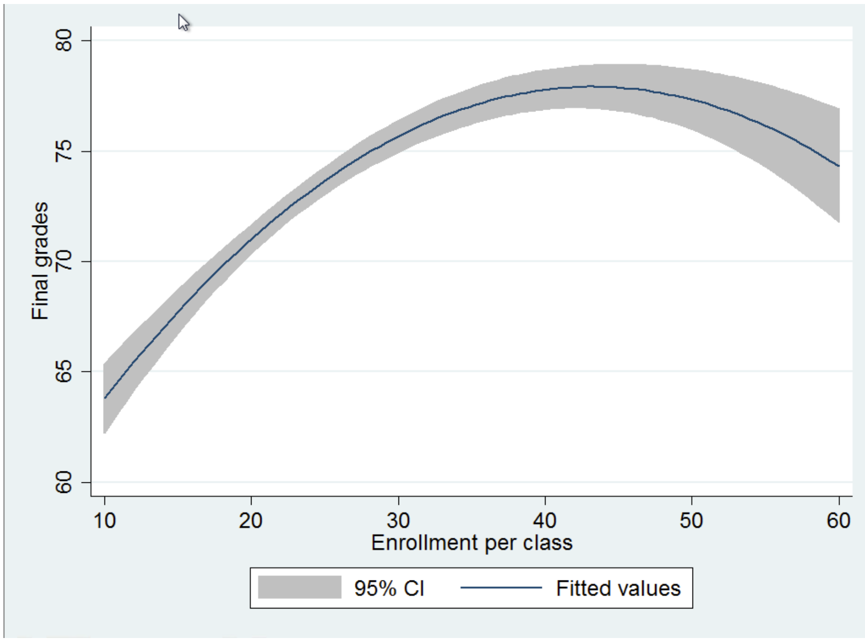

Figure 3. Parabolic Regression Fitting Line for the Effect of Class Size on Final Grades

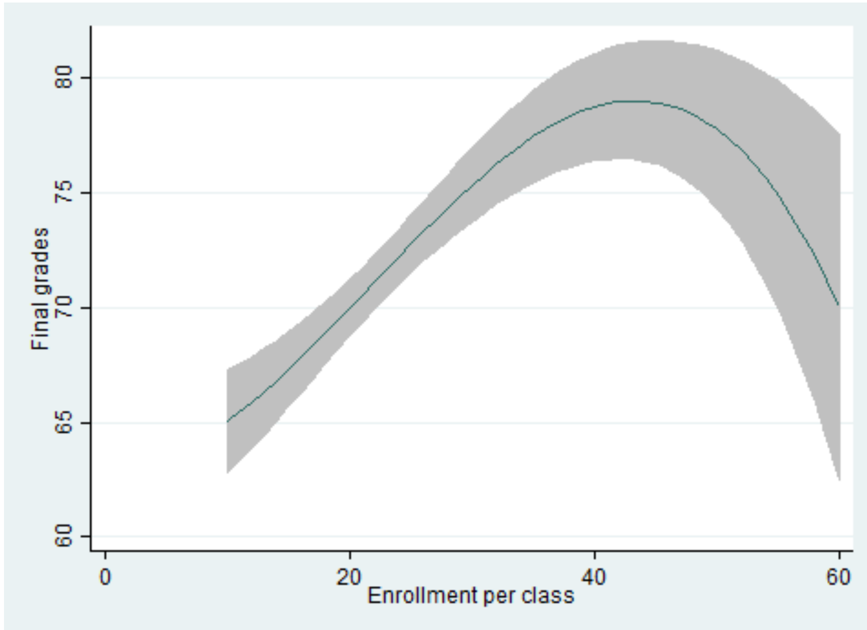

Figure 4. Fractional Polynomial Analysis of the Effect of Class Size on Final Grades

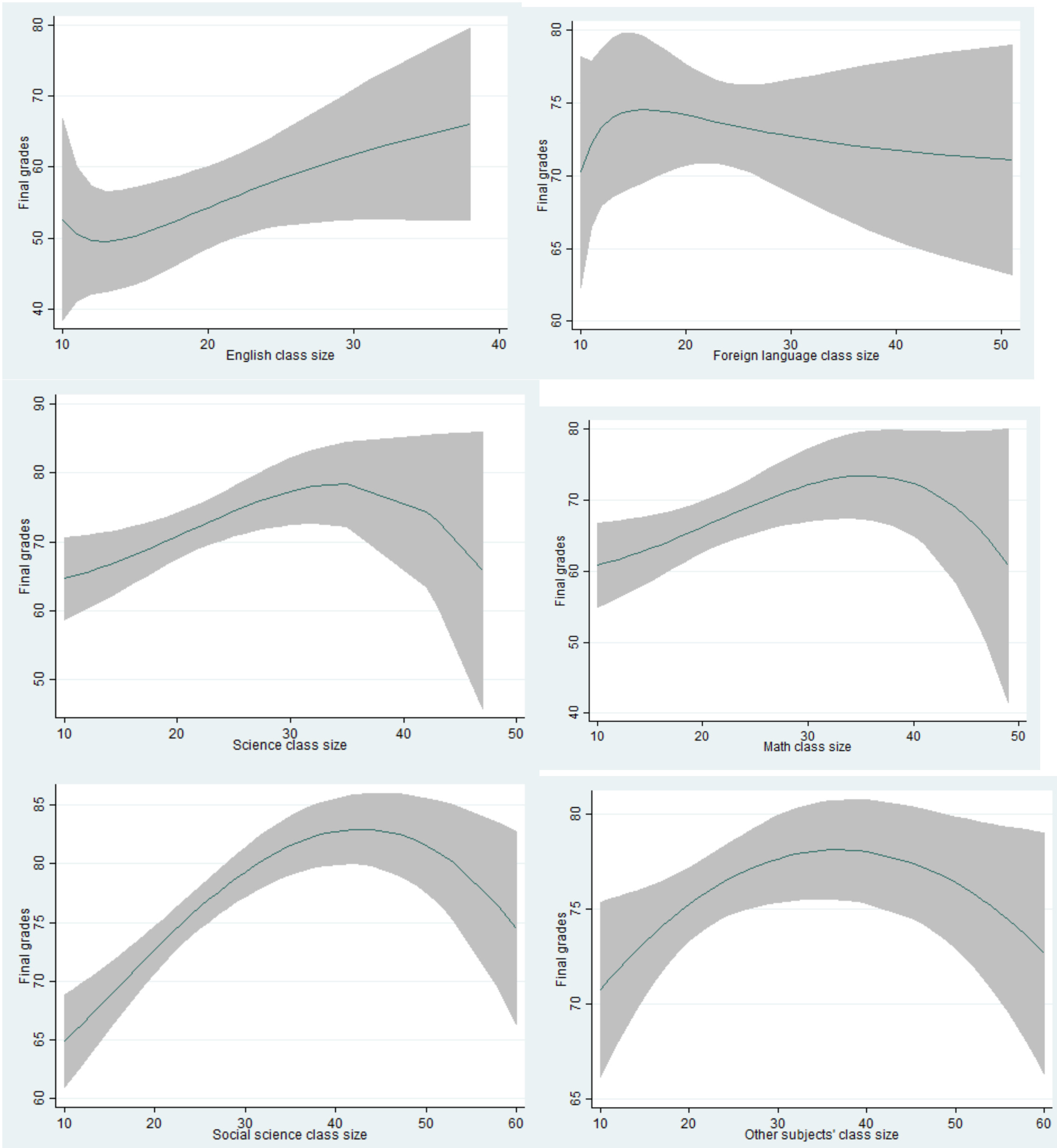

Figure 5. Fractional Polynomial Analysis of the Relationships between Class Sizes and Final Grades, by Subject