ELSEVIER

International Journal of
Educational Research 43 (2005) 387–425

International Journal of
Educational
Research

www.elsevier.com/locate/ijedures

The paradox of reducing class size and improving learning outcomes

John Hattie*

Faculty of Education, The University of Auckland, Private Bag 92019, Auckland, New Zealand

Abstract

This paper addresses four questions: What are the effects of reducing class size? How important are these effects? How can we explain these effects? and How can we improve the outcomes when class sizes are reduced? A major aim is to provide directions for resolving the paradox as to ''Why reducing class size has not led to major improvements in student learning?'' and the conclusion is that class size reductions can lead to worthwhile increases provided certain conditions are met.

Keywords: Class size; Reduction; Learning outcomes; Meta-analysis

1. Introduction

It is not difficult to find claims for both sides of the argument about whether or not reducing class sizes leads to enhancements in learning outcomes. One side argues that reducing class size leads to more individualized instruction, higher-quality instruction, greater scope for innovation and student-centered teaching, increased teacher morale, fewer disruptions, less student misbehavior, and greater ease in engaging students in academic activities. On the other side, there is a voluminous literature that does not support the claim that learning outcomes are enhanced when class sizes are reduced. A major aim of this paper, therefore, is to resolve these competing claims and to offer an explanation as to how learning outcomes can be enhanced in smaller classes, and to suggest why indeed this is not currently the case.[1]

*Tel.: +64 9 373 7599x82496; fax: +64 9 308 2355.
E-mail address: j.hattie@auckland.ac.nz.

[1]The concern is class size and not teacher–student ratio, as they are quite distinct. Class size is the number of students for whom a teacher is primarily responsible during a typical lesson (Lewit & Baker, 1997, p. 113).

0883-0355/$ - see front matter
doi:10.1016/j.ijer.2006.07.002

Section 1 of this paper will outline findings from the many studies on the influence of class size, and conclude that there is little evidence for major effects on student learning. There are probably few other domains of educational endeavor where there are such clear systematic findings of a close to zero relationship—the variance in studies is surprisingly small, and similar findings have been consistently found over the past 100 years. The earliest empirical studies were published at the turn of this century (Rice, cited by Cram, 1968), although there are earlier claims in the Talmud that the maximum size of bible classes should be 25 students. As early as 1974, Jamison, Suppes, and Wells reviewed this literature and found that 35 studies favored smaller classes, 32 were inconclusive, and 18 favored larger classes. They concluded that reducing class sizes may only "slightly" improve achievement (Jamison, Suppes, & Wells, 1974; see also Blake, 1954; Fleming, 1959; Sitkei, 1968). Hanushek (1998, p. 36) aptly sums up this literature by saying "Most discussions of reducing class size begin with an assertion that student performance will increase if only class sizes can be reduced, a proposition shown to be generally erroneous". It seems hard to reconcile the obvious—reducing class size enhances the time per student—with the evident—there is limited evidence to support the positive claims.

The claims defended in this paper are that

(a) the results are systematically small (and not that they are centered on zero);
(b) that there is much difficulty in reconciling the small effects with the rhetoric about the positive and, for many, obvious profound effects;
(c) the effects of those supporting lower class sizes are more related to teacher and student work-related conditions, and the claims of those not supporting lower class sizes are more related to the small effects on student learning;
(d) the effects of reducing class size *may* be higher on teacher and student work-related conditions which then *may* or *may not* translate into effects on student learning;
(e) we should trust past evidence (which is what literature reviews and meta-analyses are based on) as indicative of what has been;
(f) if we cannot stop the tide of parent and teacher lobbying for smaller class-sizes then it is imperative that there is a public demonstration of consequential effects when compared to other interventions that are more likely to enhance student learning.

The major arguments in this review are that a synthesis of meta-analyses and other studies of class size demonstrate a typical effect-size of about 0.1–0.2, which relative to other educational interventions could be considered "small" or even "tiny", especially in relation to many other possible interventions—and certainly not worth the billions of dollars spent reducing the number of children per classroom. The more important question, therefore, should not be "What are the reasons for this enhanced effect-size?", but "Why are the effect-sizes from reducing class size so small?" It is suggested that the reason for these small effect-sizes relate to teachers of smaller classes adopting the same teaching methods as they use in larger classes and thus not optimizing the opportunities presented by having fewer students. Most importantly, the concept of excellence in teaching in a class of 30–80 is not appropriate for discussing excellence in teaching in a class of 20–30, and similarly excellence in teaching a class of 20–30 is not appropriate for a class of 15–20.

2. The evidence

> the basic cause of failure in reading, as well as in other subjects, is the large class" (Gates, 1937, cited in Douglass & Parkhurst, 1940, p. 217).
>
> Teachers advocated for a reduction from 42 to a more reasonable 30–35 (National Education Association, 1939, cited in Douglass & Parkhurst, 1940)

There have been many studies asking the most simple and critical question—"What is the effect of reducing class sizes?" The following section outlines some of the more important studies that have addressed this question.

2.1. Glass and Smith's meta-analyses

One of the more powerful advances in research over the past 50 years has been the advent of meta-analysis: the systematic analyses of prior studies by estimating the size of the effect on the outcome and then assessing the influences of various moderators on these effects. The very first meta-analysis undertaken was related to class size. Glass and Smith (1978) synthesized 77 studies, leading to a total of 725 effect-sizes. The average effect-size when class sizes were reduced from 25 to 15 was 0.09, but, more importantly, there was a non-linear effect. Reducing class sizes from 40 or more to 20 students led to close-to-zero increment in achievement, but when class sizes dropped to 15 students or lower, there were larger effects on achievement. Smith and Glass (1980) later synthesized 59 studies covering 371 effects relating to class size and non-achievement-based outcomes such as self-concept, interpersonal regard, engagement, quality of instruction, teacher attitude, and school climate. The average effect-size for these non-achievement outcomes was 0.24. Smith and Glass (1980) concluded that achievement; attitude, teacher morale, and student satisfaction gains were "appreciable" in smaller classes (i.e., 10–15 students)—with negligible gains from reducing class sizes as high as 40–20 students. This effect was greater in secondary than in primary schools, but the same across all subjects and across various ability levels.

Not surprising, their meta-analysis was subject to much debate and criticism. For example, Hedges and Stock (1983) reanalyzed Glass and Smith's (1978) set of studies using slightly more rigorous statistical methods, but found no differences to the earlier conclusions. A more telling criticism of the Glass and Smith meta-analysis was that the studies were of short duration, included many one-on-one tutoring studies, and were in some cases non-school related (e.g., tennis coaching). Slavin (1989) found only eight studies that met his inclusion criteria of lasting at least 1 year, involving a substantial reduction in class size, and involving random assignment or matching of students across larger and smaller classes. He concluded that substantial reductions in class size have small positive effects on students (effect-size from 25 to 15 = 0.13) and the effect was not cumulative and disappeared within a few years.

2.2. Indiana's Prime Time study

McGiverin, Gilman, and Tillitski (1989) conducted a meta-analysis of 10 studies of Indiana's Prime Time project, a longitudinal study which aimed to reduce class size to 14 in 24 Year 1–3 classes. They reported that Year 3 students, who had been in smaller classes for 2 years had significantly higher achievement test scores than did students in larger classes, with an overall average effect-size of 0.34 (see also Chase, Mueller, & Walden, 1986; Malloy & Gilman, 1989; Mueller, Chase, & Walden, 1988). It is difficult to credit this effect to class size, however, as the study had few controls, it is not clear that small classes were kept small for the entire day, and, while the average class size for the 'smaller' classes was set at 18, actual 'small' class sizes ranged from 18 to 31, and classes of 24 were considered small if there was a teacher aide to assist the teacher.

2.3. Project STAR

Project STAR (Student–Teacher Achievement Ratios) is among the most talked about, most impressive, and most powerful study on class size. Project STAR was motivated by Prime Time and began in Tennessee in 1985 (for a history see Ritter & Boruch, 1999). The project involved a random assignment of about 6,500 students in 329 classrooms in 79 schools entering kindergarten, in 29 schools to a regular class (22–26 students), or small class (13–17 students), or a regular class with a full-time teacher aide (the latter were no different from regular classes with a teacher, indicating that any effects are related to class size and not teacher–student ratio). It was intended that the students so assigned stay in classes of the same size for the next 3 years, when the students then moved into regular-sized Year 5 classes. Teachers were also assigned at random to the class groups, and no special instructions were given to the teachers of the different sized classes (Achilles, 1999, 2002; Achilles & Finn, 2000; Achilles, Finn, & Pate-Bain, 2002; Finn & Achilles, 1990; Finn, Folger, & Cox, 1991; Word et al., 1990).

Across all effect-sizes, the smaller class advantage in Year 1 was approximately 0.15–0.18; for Year 2, 0.22–0.27; and for Years 3 and 4, 0.19–0.26. These overall effects (0.15–0.27) are very similar to the average reported in the Smith and Glass meta-analysis. Effect-sizes also increased monotonically as students spent additional years in a small class. Again, in each grade, the benefits of additional years in a small class were greater. In Grade 2, for example, the achievement advantage for students who had participated in small classes for 1, 2, and 3 years was 0.12, 0.24, and 0.36 SD, respectively, in reading and 0.16, 0.24, and 0.32 s, respectively, in mathematics. After the students returned to regular-sized classes, the effects began to decrease (Year 4 = 0.12, Year 5 = 0.22, Year 6 = 0.17, and Year 8 = 0.13). A different story emerges when the gains over time are considered. Fig. 1 presents the effect-size gains from Kindergarten to Year 1, Year 1–Year 2, and Year 2–Year 3. It can be seen that the greatest gains in Reading are made by the students in the regular size classes from Year 1 onwards, and similarly the gains are greatest in regular classes from Year 1 onwards. Fig. 1 also demonstrates that any gains may have been a one-off start up gain for Kindergarten students (Tomlinson, 1990).

The effect-sizes were greater for minority (close to double) compared to white students for all achievement areas. Zero effects were found for motivation and self-concept. Finn and Achilles (1990) reported that the difference between minorities and whites in mastery rates on the Year 1 reading test was "reduced from 14.3% in regular classes to 4.1% in small classes" (p. 568), almost double for minority compared to white students (Fig. 2). On the other hand, Jepsen and Rivkin (2002) investigated California's $1 billion innovation to reduce K-3 class sizes from 30 to 20 and the effects on teacher quality and student achievement. They found that the rapid expansion of the teaching force needed to teach the additional classes appears to have led to deterioration in average teacher quality in schools serving a predominantly black student body. Teachers of smaller classes,

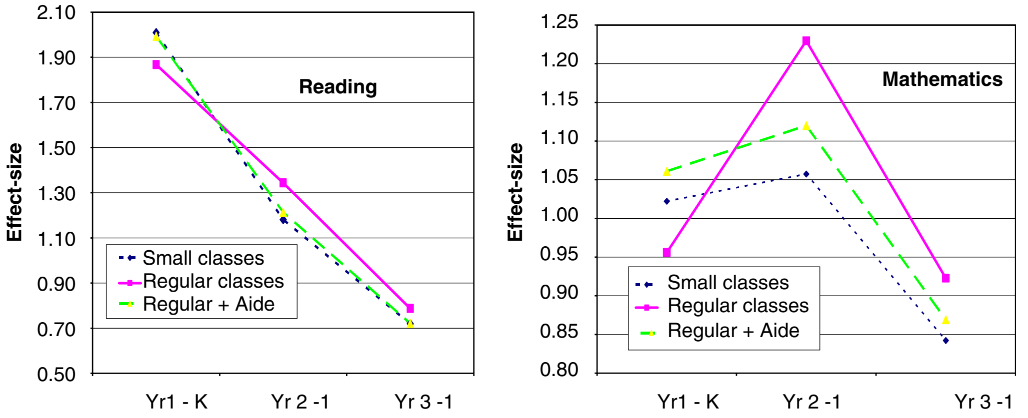

Fig. 1. Effect-sizes gains across the years of Project STAR for reading and mathematics.

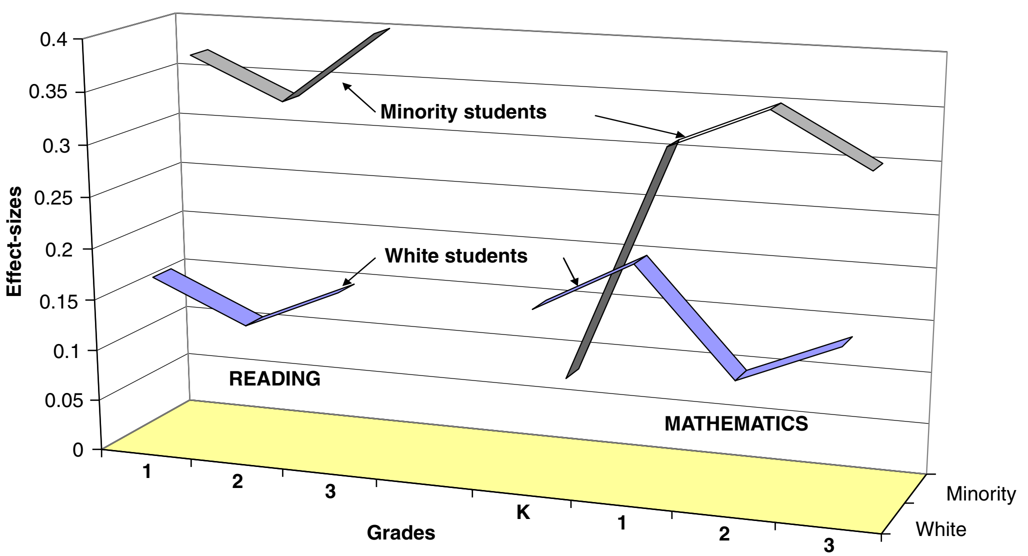

Fig. 2. Effect-sizes of smaller vs. larger class sizes for white and minority students.

particularly in non-white and free lunch schools, tended to lack full certification, have no postgraduate education, and were more likely to be in their first or second year of teaching. This deterioration can partially or, in some cases, fully offset the benefits of smaller classes, demonstrating the importance of considering all implications of any policy change.

At Year 5, all of the students in the study returned to regular-sized classes. The Lasting Benefits Study followed many of these students, some through to Year 11 (Finn et al., 1991) and beyond (Finn, Gerber, & Boyd-Zaharias, 2005). The effect-sizes in favor of those who had begun in smaller classes were primarily in the 0.10–0.15 range, indicating that there were positive effects of this early age intervention, even when the small-class intervention was disbanded. The effects for student engagement in learning (initiative taking, lack of disruption, attentiveness) were greater in the smaller classes (effect-size = 0.13) a year after the students returned to normal classes (Finn & Achilles, 1999). Wenglinsky (1997), in an analysis of production functions based on Project STAR, also reported positive effects for small classes at Year 5 but not at Year 9. Other claims include that students who attended small classes displayed improved learning behavior, increased engagement in school, and decreased disruptive or withdrawn behavior compared to their counterparts in regular-sized classes (Finn, 1998; Finn, Fulton, Zaharias, & Nye, 1989; Nye, Hedges, & Konstantopoulos, 2000).

There is also evidence that students from smaller classes graduated on schedule at a higher rate (76%) than those from regular classes (64%), completed school with an honors diploma more often (45%) than students from regular classes (29%), and dropped out of school less often (15%) than regular classes (24%) (Boyd-Zaharias & Pate-Bain, 2000; Pate-Bain, Boyd-Zaharias, Cain, Word, & Binkley, 1997). These claims that students who had been in the smaller classes appeared to have maintained academic achievement advantages is somewhat misleading as the differences are 'appearances' and were not statistically significant differences.

Project STAR has been widely summarized and lauded as among the best research designs, and the power of their positive findings have been used to justify many other ventures into reducing class sizes. It is not without its critics, however. Nye, Hedges, and Konstantopoulos (1999) controlled for the effects of those who dropped out of the small classes and found slightly higher effects, whereas Hanushek (1997a, 1997b, 1999, 2003) remained highly critical of the large attrition rate in the Project STAR data. He noted that slightly less than half of the original students in the experiment remained in the study until the end of the third grade (Year 4), and the assignment while random, was not blind—teachers knew of the assignment (obviously). Harker (2004) also noted that the STAR experiment had all three "treatments" running in each of the participating schools, which could have led to compromises through the normal social and collegial interaction processes of the school communities. Hence, the results could have been related to more resources going to the smaller classes, and other "more direct motivation and incentives of teachers and principals that could bias the results of the different treatment groups" (Hanushek, 1997a, p. 153). He concluded that "it is only slightly better than an even bet from the STAR data that the small class achievement will exceed that of the regular and the regular-with-aide classes in any of the sampled schools" (p. 159).

2.4. *The Wisconsin Student Achievement Guarantee in Education (SAGE) study*

The Wisconsin Student Achievement Guarantee in Education (SAGE) program was designed as a 5-year project commencing in the 1996–1997 school year. Schools with a high proportion of students living in poverty were required to implement four interventions: reduced class size (from 25 to 12–15), opening from early in the morning until late in the evening, developing rigorous curricula, and creating a system of staff development and professional accountability. After the first 2-year evaluation, Molnar et al. (1999) concluded that the effect-size from the first-year SAGE students for class size reductions was about 0.2. From the tables in their article, I estimated that the effect-sizes were 0.16 for Reading, 0.20 for Language, and 0.25 for Math, but this effect needs to be compared to the effect-size change (across the three subjects) from 1 year to the next of 1.20 for the smaller classes and 1.00 for the comparisons classes: a small advantage (0.20) relative to the yearly increase (1.20). The effect-size was much higher for African-American students. They noted no differences between class sizes of 15 with one teacher and class sizes of 30 with two teachers, concluding that this "suggests that the benefits of reducing class size may be achievable without the attendant capital costs of building additional classrooms" (Molnar et al., 1999, p. 177).

2.5. The Connecticut study

Hoxby (1998, 2000) used data from Connecticut students who entered the sixth grade in 1991 and follow-up data for the next 11 years. She found that reducing class size from 30 to 15 had no effects on student achievement: "the effect of reducing class size is rather precisely estimated to be close to zero" (Hoxby, 2000, p. 1273). She found no evidence that class size reductions are more efficacious in early grades, or in schools that contain higher numbers of African-American students.

2.6. Blatchford et al. and the United Kingdom studies

Blatchford et al. (2002) followed a cohort of students over a 3-year period (students aged 4–7). The study was based on 220 schools, with 368 classes and 9330 children in eight Local Education Authorities. The class sizes typically varied from 10 to 35 in reading, and 15 to 33 in mathematics. They found decreasing class size was related to increasing test scores, but little apparent change between class sizes of about 18 and 25. They found little evidence to claim that more and/or positive peer interactions accompany the benefits in achievement until class sizes are reduced to around 15 students. For Reading, there is little apparent change between class sizes between 25 down to 18 (Fig. 3), whereas for Mathematics the effects become more pronounced between 25 and smaller (Fig. 4). For both, the large effects are for the lower achievers and those from the lower socioeconomic backgrounds.

Later, Blatchford, Bassett, Goldstein, and Martin (2003) argued that approximately 25 may be an important number of children, below which relationships with classroom processes, such as the number and size of within-class groups, become most evident. Their claim is that smaller classes were particularly important during the first year of schooling, particularly for those who are most in need academically; that is, those with the lowest school entry scores who thus have the most ground to make up.

Also in the UK, Goldstein, Yang, Omar, Tumer, and Thompson (2000) found nine studies that met their criteria for inclusion in their meta-analysis: randomized or matched design, achievement outcomes, longitudinal, and large classes no larger than 40 and small classes no smaller than 15. They used multilevel modeling and found that the effect on reading scores was a decrease of 0.02 SD per additional student, or 0.20 when reducing from 25 to 15. Hargreaves, Galton, and Pell (1997, 1998) observed 14 'expert teachers' who they expected would adapt their teaching strategies to cope with new situations. They made these observations in three settings: in their own class, in another teacher's class (which had a different number of students) and, for a subset of the teachers, in their own class with only half the students. They found that there was no statistically significant difference between small and larger classes in individual attention, teachers still used 'group' as the audience (more whole class group in smaller, more groups within larger classes), a greater use of 'sustained interactions' in smaller classes, and the teachers were more likely in smaller classes to engage in more enquiring questioning, ask more

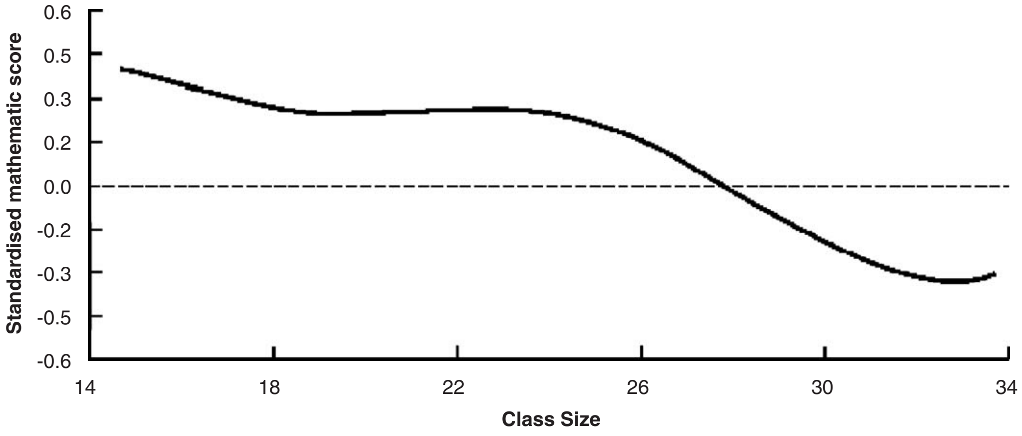

Fig. 3. Post-reception literacy by class size (Blatchford et al., 2002).

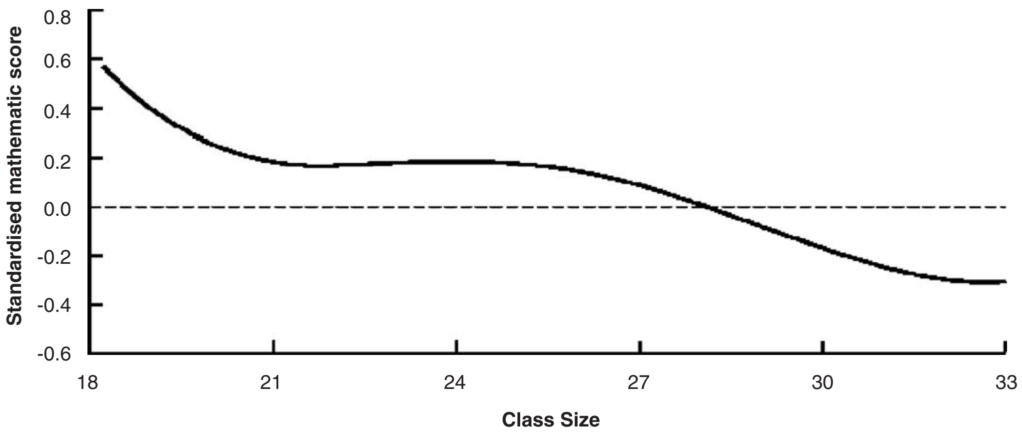

Fig. 4. Post-reception mathematics by class size (Blatchford et al., 2002).

task-related questions, make more task statements, and are generally more involved with the task when interacting with students. They argued that teachers needed to be specifically trained to work with smaller classes.

Dustmann, Rajah, and van Soest (2003) estimated the effects of reducing class size on the decision to remain in school beyond the age of 16 for pupils in England and Wales. Using a longitudinal study (started in 1958), they found a significant negative effect of class size on the probability to stay at school at age 16. The "significant" effect-size was only -0.04!

2.7. Other recent studies

Similarly in Australia, it has been difficult to find studies identifying differential effects on achievement relating to class size. Johnson, Jensen, Feeny, and Methakullawat (2004) used one of the larger data sets based in Victoria generated from 1232 primary schools, 264 secondary schools, and 44 primary and secondary schools. They concluded that "we have been unable to find any evidence that class size is an important determinant of academic performance in primary or secondary schools" (p. 33). Nowhere did they find systematic effects of class size, with most effects centering around zero, and certainly "rarely important" (p. 27).

There have been various econometric studies that also have found small effects from reducing class size, with effects clustering around 0.00–0.10 (Boozer & Rouse, 1995; Hanushek, Rivkin, & Taylor, 1996; Krueger, 1999). Even when sophisticated data methods have been used to estimate value-added effects, the effects are close to zero. Based on 26 studies in 1052 US schools (the NELS data), Ludwig and Bossi (1999) found that reducing class size (from 25 to 15) improved test scores in reading but have a "very modest negative effect" on math scores (p. 391). Akerhielm (1995) used the National Education Longitudinal Study (NELS), based on a US nationally representative sample of over 24,000 eight graders. She found that lower ability students are disproportionately found in smaller classes, and the overall effect-size of 0.18 for Science and 0.13 for History, and concluded that "the incremental benefits may not surpass the incremental costs of decreasing class size" (p. 239).

2.8. Hanushek's vote counting summaries

Hanushek (1986b, 1997a, b, 1999; Hanushek et al., 1996) has long maintained that there is little evidence to support the benefits of smaller classes on student learning. He completed a vote count of the number of statistically significant and non-significant effects based on 90 publications (leading to 377 estimates; he uses the term "insignificant" which I assume means not-significant or close to zero). I have grouped the "statistically insignificant" and "not different from zero" together, and it is clear that the most dominant effect is that there are no differences from reducing class sizes, and just as many statistically significant positive as negative findings (Fig. 5).

Hanushek's conclusion is that about 13% of the studies were statistically significant and positive in favor of smaller classes, 15% were negative, and 72% were not different from zero. Hence, there is "little reason to believe that smaller classes systematically lead to improvements in student achievement" (Hanushek, 1999, p. 148). When he added studies conducted within a single state in the United States, Hanushek concluded that "more studies actually suggest that small classes are harmful… [and that overall, there is] no consistent or clear indication that overall class size reductions will lead to improved student performance" (p. 149). It is perhaps not surprising that the box-scores of positive and negative are so close, when it is noted that the effect-sizes in the above reviewed studies are also close to zero. The box-score method, however, is not preferable to ascertaining the effect-sizes, as it loses too much information (e.g., an effect-size of 0.99 is positive, and an effect-size of –0.01 is negative, but the box-score concludes that there are as many positive as negative conclusions!).

Krueger (2003) critiqued Hanushek's method of selecting studies, extracting and counting the estimates, claiming that it was irrational and leads to a misleading result. He claimed that Hanushek's choice of studies was biased to those with small effects, although this criticism seems contrary to well-cited arguments that it is more likely that only "statistically significant" results are published. Instead, Krueger weighted the studies in a

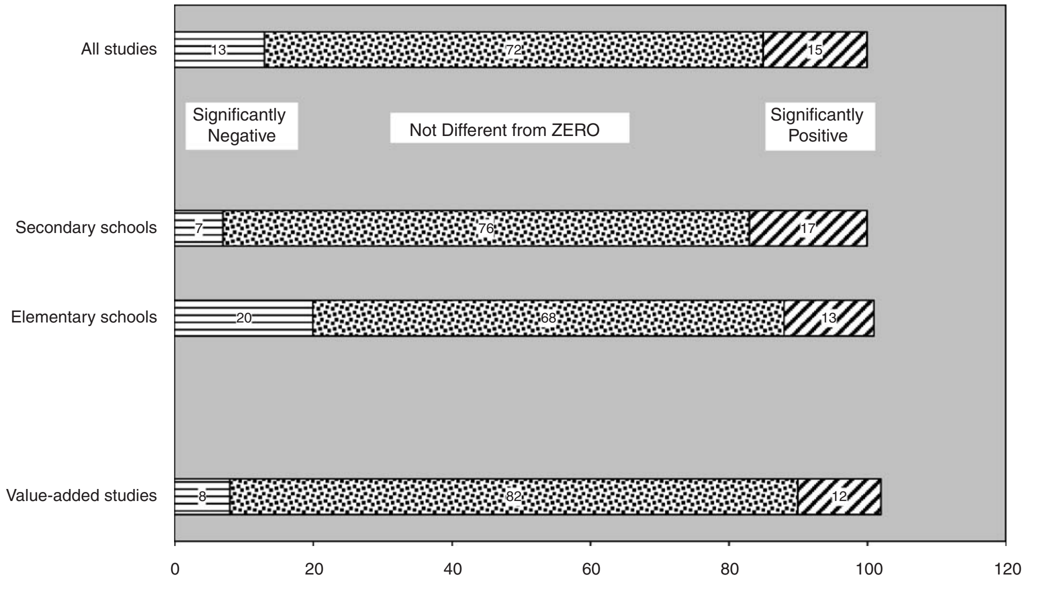

Fig. 5. Summary of Hanushek's vote counting.

number of ways (by quality of design, by "impact" of the journal quality, by number of estimates), and this led to twice as many positive and half as many negative. But again, another lost opportunity—the magnitude of these positive and negative findings is not reported.

2.9. Conclusions from these studies

Table 1 summarizes some of these studies. Across these meta-analyses, summaries of major initiatives, and newer studies, the average effect-size is 0.13. Thus, the typical effect of reducing class sizes from 25 to 15 is about 0.10–0.20. Perhaps as interesting as the typical value, is that there is not a lot of variance in these estimates; the mean is a reasonable summary of the effects of reducing class size.

These studies represent a variety of designs including meta-analysis, longitudinal studies, cross-cohort studies, are from many countries (USA, UK, Israel, Bolivia), from across all grades, and use some of the most sophisticated statistical methods available. There is remarkable consistency across the effect-sizes from these many diverse studies.

2.10. Non-western studies

There is an important codicil that is worth mentioning before moving on to address the worthwhileness of this typical 0.10–0.20 effect-size. Most of this research is based on classes in Western countries. It is noted that the Western literature is primarily interested in reducing class sizes from 25–30 to 15–18, whereas in many non-Western countries the typical class size well exceeds this 25–30 class size. Indeed, in Hong Kong, the typical class size is more like 30–40 (Biggs, 1998).

Table 1
Synthesis of meta-analyses of reducing class size from 25 to 15

Authors (year)	No. of studies	No. of effects	No. of classes	No. of students	Effect-size	Outcome
Glass and Smith (1978)	77	725	14,358	520,899	0.09	Achievement
Smith and Glass (1980)	59	371			0.24	Non-achievement outcomes
Finn (1998)	1	1	79	6500	0.22	Achievement
	1	1	79	6500	0.12	Achievement (Grades 4–6)
	1	1	79	6500	0.02	Self-concept, Motivation
McGiverin et al. (1989)	10	24			0.34	Achievement
Molnar et al. (1999)	1	1	411	9790	0.21	Achievement
Hoxby (2000)	1	1	14,593	306,453	0.03	Achievement
Blatchford et al. (2005)	1	1	368	9330	0.23	Achievement
Goldstein et al. (2000)	9	36	1178*	29,440	0.20	Achievement
Dustmann et al. (2003)	1	1	224	3811	−0.04	Achievement
Akerhielm (1995)	1	1	1052*	24,000	0.15	Achievement
Rice (1999)	1	1	8760	24,599	−0.04	Achievement
Johnson et al. (2004)	1	1	168*	3700	0.00	Achievement
Angrist and Lavy (1999)	1	1	1327	46,455*	0.15	Achievement
Urquiola (2000)	1	1	608	10,018	0.20	Achievement
Average	164	1165	40,728*	948,540*	0.13	

* = estimated.

Fuller (1987) reviewed nine studies in developing countries (Botswana, Thailand, India, Chile, Iran, Egypt, Kenya, Malaysia, Puerto Rico, Tanzania, Bolivia, and Argentina). The studies all used primary students, the mean class size was 44, and Fuller found no difference in learning outcomes relating to class size. He further reported that in five studies, students working within larger classes performed at higher levels than in smaller classes, and Fuller only identified one study that found smaller classes at primary level significantly related to higher achievement (in Columbia).

Angrist and Lavy (1999) used a large Israeli database to assess the effects of varying class sizes, where the class varies in size up to 40. They used 1327 class means in 625 schools and scores in Verbal and Math achievement. They estimate effect-sizes across classes of 0.18 s for fifth grade and about 0.13 for fourth-grade students. In a similar study, Urquiola (2000) studied 10,018 third-grade students in Bolivia where the 608 class sizes varied from very small up to 40. The effect-sizes were estimated between 0.18 and 0.23. In a similar study, Urquiola (2000) studied 10,018 third-grade students in Bolivia where the 608 class sizes varied from very small up to 40. The effect-sizes were estimated to be between 0.18 and 0.23. Lindahl (2001) used 556 Swedish fifth graders, and concluded that the estimates for class-size are not significantly different from zero.

A recent Hong Kong study outlined the results from 28 "potential exemplars" of reducing small classes. Certainly teachers and principals "felt" that reduced classes led to more individual attention and higher motivation, but there was a "gap" between what teachers felt and what actually took place. From observations of the classrooms, the interactions were mainly characterized by close-ended questions or questions with model answers with little follow-up discussion or exploration of ideas, teaching was textbook-bound, there was not much support/attention given to individual students, the main purpose was class exercises and cross-checking of answers. The conclusion was that "there was no significant difference in the nature and quality of teacher–student interactions between small classes and regular/large classes" (Legislative Council Panel on Education, 2004, p. 5).

Wößmann and West (2002) investigated the effect of class size on student achievement across many countries using the Third International Mathematics and Science Study (TIMSS) data. TIMSS is based on representative samples of students in the two adjacent grades with the highest share of 13-year-old students from about 40 countries. In 18 of these countries, they identified class size effects by relating differences in the relative performance of students in two adjacent grades within individual schools to that part of the between-grade difference in class size in the school that reflects between-grade differences in average class size. Thus, they effectively excluded both between-school and within-school sources of student sorting.

They were unable to detect a statistically significant effect of class size on student achievement for most school systems. I have plotted the typical class sizes with the math (Fig. 6) and science performance (Fig. 7) and it can seen that for most countries the relationship is zero (i.e., the class sizes are constant whereas performance is appreciably different). Indeed, many of the countries where class sizes were greatest were among the highest performers. Within school systems, they found four cases in which there were statistically significant beneficial effects from being in smaller classes (France and Iceland in mathematics, Greece and Spain in science); and for the rest they rejected any sizable class size effect with reasonable confidence. There were no effects related to the level of development or the country's overall level of resources, and this led to their conclusion, that the

> evidence on class size effects presented in this paper suggests that capable teachers are able to promote student learning equally well regardless of class size (at least within the range of variation that occurs naturally between grades). In other words, they are

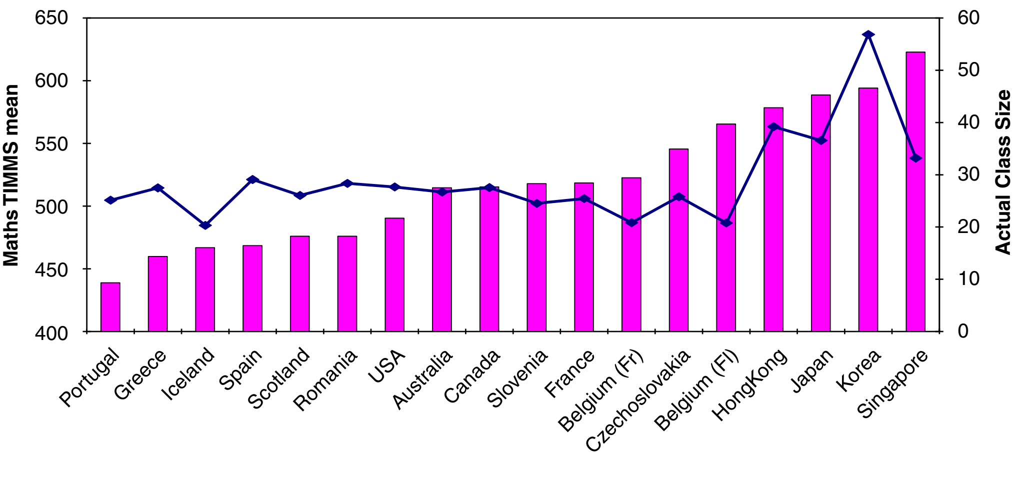

Fig. 6. Average class size and mathematics performance (from TIMSS).

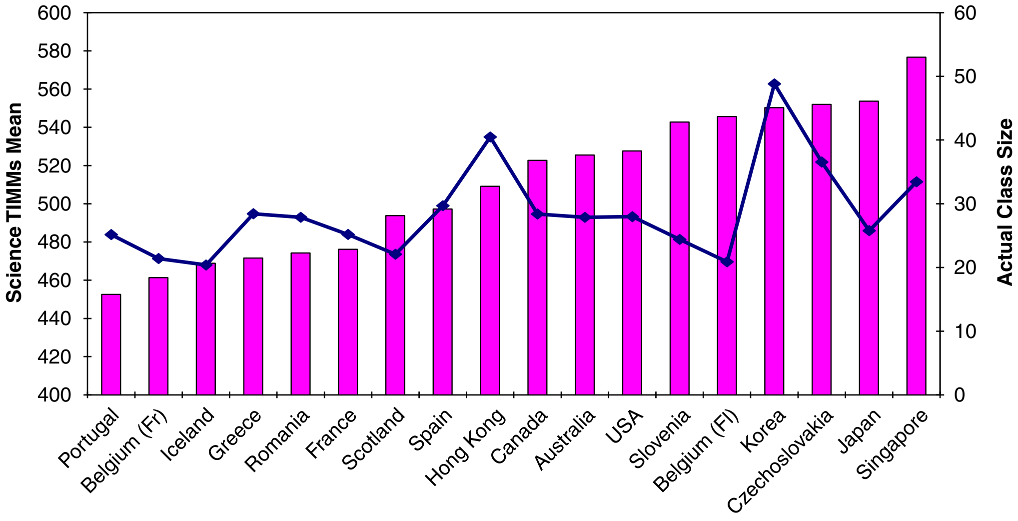

Fig. 7. Average class size and science performance (from TIMSS).

capable enough to teach well in large classes. Less capable teachers, however, while perhaps doing reasonably well when faced with smaller classes, do not seem to be up to the job of teaching large classes. (Wößmann & West, 2002, p. 30)

3. How important are these effects?

3.1. Cost–benefit analyses

One way to address the importance of this 0.10–0.20 effect-size is to undertake a cost–benefit study. For example, Buckingham (2003) estimated that the effects of reducing the overall average ratio of New Zealand primary and secondary students by one student (to 18.4 and 14.5, respectively) was around NZ$113 million per year. While noting that this ratio is not the same as reducing class size, this cost only provides for one less student per class on average, it would be an ongoing commitment, is not a one-off investment as it would account only for extra staffing costs. Other additional costs include classrooms, classroom resources and ongoing professional development (see also Greenwald, Hedges, & Laine, 1996).

In a study comparing the relative magnitude of achievement effects resulting from the introduction of textbooks, establishment of radio instruction, and lowering mean class size, Jamison (1982), see also Heyneman, Jamison, and Montenegro (1983) estimated that to obtain the achievement benefit gained from raising the availability of textbooks at a constant increment of cost, schools must lower average class size from 40 to 10 students per teacher. Fuller concluded "in most situations, lowering class size with the intent of raising achievement is not an efficient strategy" (p. 276). Similarly, Levin (1988) compared the cost-effectiveness of four reforms for raising student achievement at the elementary level in reading and mathematics: a longer school day, computer-assisted instruction (CAI), cross-age tutoring, and reduced class size. Peer tutoring showed the largest effects: nearly a year of student gain in mathematics and about one-half a year in reading. The longer school day and reducing class size by five students showed the smallest effects. CAI was associated with gains in the middle of the range of results.

Brewer, Krop, Gill, and Reichardt (1999) estimated the costs of reducing class sizes to 18 students in Years 1–3 in the United States would require hiring an additional 100,000 teachers at a cost of $US5–6 billion per year, and an additional 55% more classrooms. To reduce again from 18 to 15 students would cost a further $US5–6 billion per year. Instead, they estimate that this investment could, instead, be used to raise teachers' salaries by $20,000 per year (see also Blatchford et al., 2002).

3.2. Converting effect-sizes to other metrics

A second way to address the importance is to use conversions of effect-sizes to more understandable metrics. For example, when the 0.13 effect-size is translated into the percentage of students who would gain from smaller classes, then "59% of those in smaller classes would outperform the typical student in the larger class." This may sound an appreciable amount but indeed it is only 9% as the comparability is to the typical student placed at the 50th percentile. Others have made conversions to "months gained" and the usual message is that these months gained (e.g., 3 months) are appreciable when compared too many other interventions, but as we shall see it is not. For example, Grissmer (1999) and Molnar (1999) argued that the ongoing benefits from having spent the first 4 years of schooling in small Tennessee classes was 6–9 months in Grade 4 to over 1 year by Grade 8.

3.3. Comparing to many other interventions

A third way, and the argument advanced in this paper, is that such typical effects of reducing class size (from 25 to 15) need to be compared not to the zero point (of no changes in outcomes) but to the effects of many other educational interventions—as only then can more defensible interpretations be made. I have been synthesizing over 500 meta-analyses and the average of all these effects can be used as the more appropriate comparison point. Table 2 presents a sample of 50 effects from across this distribution and the place of class size is clearly among the smaller effect-sizes, while Fig. 8 presents a bar graph of the 4000 or so effect-sizes from these 500 meta-analyses based on about 300,000 research studies relating to student outcomes.

The immediate message is that there are very few influences that are detrimental; indeed almost everything enhances achievement (i.e., is greater than 0). Thus to make claims that smaller classes are more effective than larger classes is hardly an exciting finding as there are so few influences that do not have a positive effect on learning outcomes. The more appropriate comparison point is to compare the typical effect-size of reducing class size to the average of all possible interventions on student learning. It can be seen in Fig. 8 that the effect of 0.13 is rather small and certainly well below the average of all influences (0.40).

Thus, the first and correct conclusion is that the mean effect-size of 0.13 indicates that on average it is worthwhile in terms of achievement to reduce class sizes. Those who claim there is systematic evidence that reducing class size enhances learning outcomes are correct. But this message alone is misleading. The comparison point is not the zero point, but is the average of all other influences on student achievement. The effect of reducing class size is among the least effective of influences on student outcomes.

Table 2
A sample of 46 influences on student achievement

	Influence	No. of studies	Effect-size
1	Feedback	13,209	0.81
2	Direct instruction	1925	0.81
3	Prior achievement	619	0.80
4	Lack of disruptive students	1511	0.79
5	Quality of teaching	808	0.67
6	Phonological awareness	429	0.66
7	Early intervention	30,275	0.64
8	Peer assessment	308	0.63
9	Challenging goals	959	0.59
10	Self-assessment	521	0.56
11	Mastery learning	1933	0.55
12	Interactive video	1008	0.54
13	Peer influences	366	0.50
14	Bilingual programs	1457	0.49
15	Study skills	3224	0.49
16	Socio-economic status	1899	0.48
17	Professional development	18,644	0.48
18	Tutoring	2101	0.47
19	Advance organizers	2106	0.46
20	Hypermedia instruction	317	0.46
21	Parent involvement	2597	0.43
22	Home environment	25,685	0.42
23	Self-concept	4925	0.40
24	Individual instruction	4747	0.39
25	Time on task	1680	0.37
26	Homework	558	0.35
27	Computer-assisted teaching	16,415	0.32
28	Acceleration	345	0.32
29	Testing frequency	2346	0.32
30	Calculators	238	0.20
31	Learning hierarchies	168	0.19
32	Desegregation	1590	0.19
33	Mainstreaming	1635	0.19
34	Finances	1634	0.18
35	Behavior objectives	157	0.18
36	Teacher questioning	476	0.17
37	Programmed instruction	801	0.14
38	Ability grouping	5078	0.14
39	Teacher expectations	912	0.14
40	**Class size**	**2559**	**0.13**
41	Diet	255	0.12
42	Problem-based learning	41	0.06
43	Whole language programs	13	0.06
44	Open vs. traditional classes	3426	0.04
45	Summer vacation	269	−0.07
46	Retention	3626	−0.20

Note: Class size is highlighted.

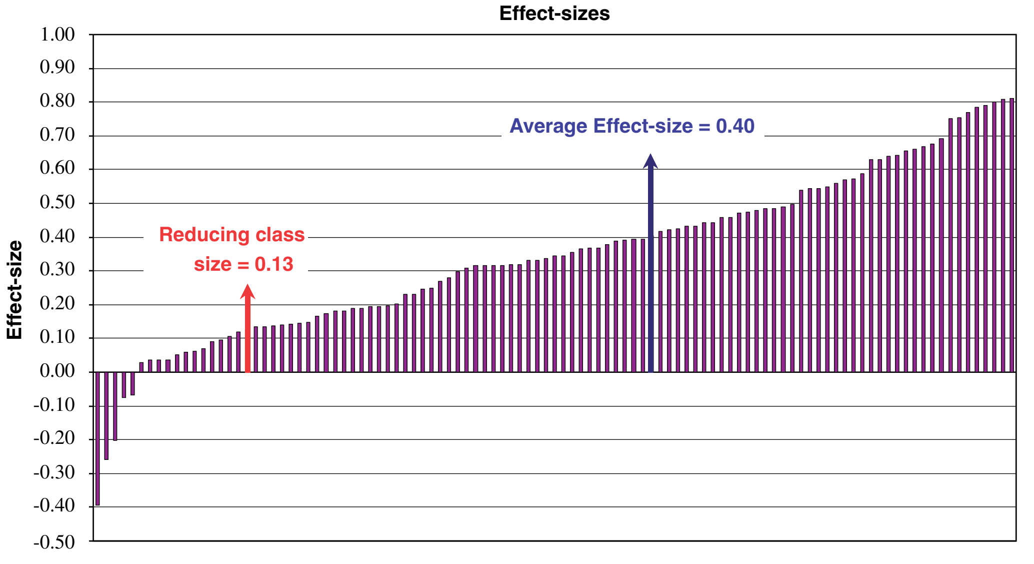

Fig. 8. Effect-sizes from meta-analyses of influences on learning.

It is notable that most structural effects are low in this continuum. Among the lower effect-sizes along with reduced class size, are ability grouping, retention, open classrooms, finances, and mainstreaming (see Table 2). The common denominators of the effects at the top of the continuum relate to specific teaching issues controlled by the professional judgment of teachers—such as direct instruction, reciprocal teaching, quality of teaching, and presence of challenging goals. These particular teacher-controlled methods all aim to optimize feedback, challenge, increase engagement on task and they will become important as we move to the next question as to "Why the difference between smaller and larger classes is so small?"

4. Why is this difference so small?

It seems ironic that the list of reasons as to why smaller classes are more effective is very long, but so little research is undertaken asking why the differences are so small? Further, the expected differences seem more related to quantity than quality (e.g., more on-task behavior, less student–student interactions), and there is little, if any, evidence that the fundamental nature of teaching differs when there are smaller classes (even when the same teacher is teaching small and large classes within the same day). Some of the arguments tend towards the reductive—that is, the effects must be obvious simply because there are fewer students. In the same manner, there must be less time on non-teaching administrative functions simply because there are fewer students. There is more time spent on instruction and less on discipline simply because in smaller classes there are 15 less students to be "naughty" than in a class of 30. But even if these "quantity" claims are correct, they do not explain why the effects are so small. There are three major interrelated reasons why this effect may be small:

- it is difficult to find studies whereby the nature of classroom experiences are differentially related to class size;

- teachers tend to use the same teaching methods regardless of class size; and
- there may be greater attention to peer effects in smaller classrooms.

4.1. Variation in teachers

Using the Project STAR data, I separated the students into 166 regular classrooms with teacher aides, 174 regular (about 25 student) classes, and 199 smaller (12–18) classes. Fig. 9 presents the means of these classes for the Reading and Fig. 10 shows the means for Mathematics. It is clear that there is an enormous amount of overlap in the performance of the students in these three types of classes. It is not the case that there are 'clear'' advantages of smaller classes.

4.2. Differences in the nature of classroom experiences

Table 3 summarizes some of the arguments that have been supported and not supported in the research on reducing class size. The list looks supportive of the claim that there are more desirable behaviors in the smaller classes. The major message, however, is that the evidence of even these differences are small. For example, in Project STAR, Evertson and Folger (1989) investigated the impact of class size on teaching practices using observations. There were no statistically significant differences between small and regular classes for the

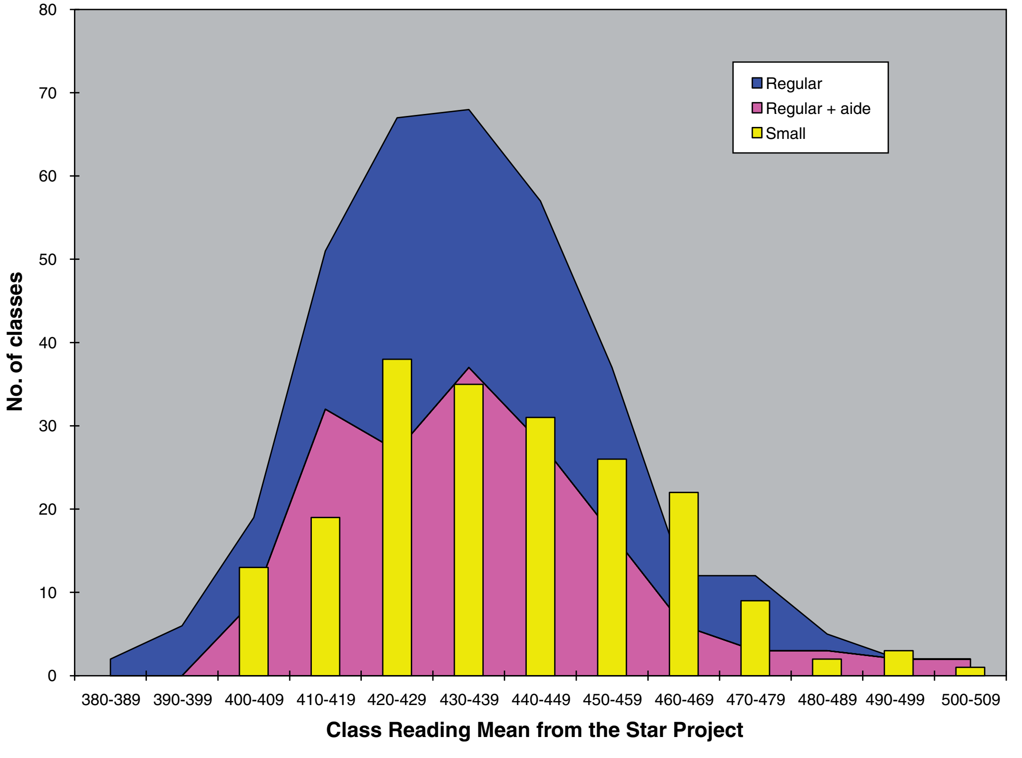

Fig. 9. Means of STAR classes on reading.

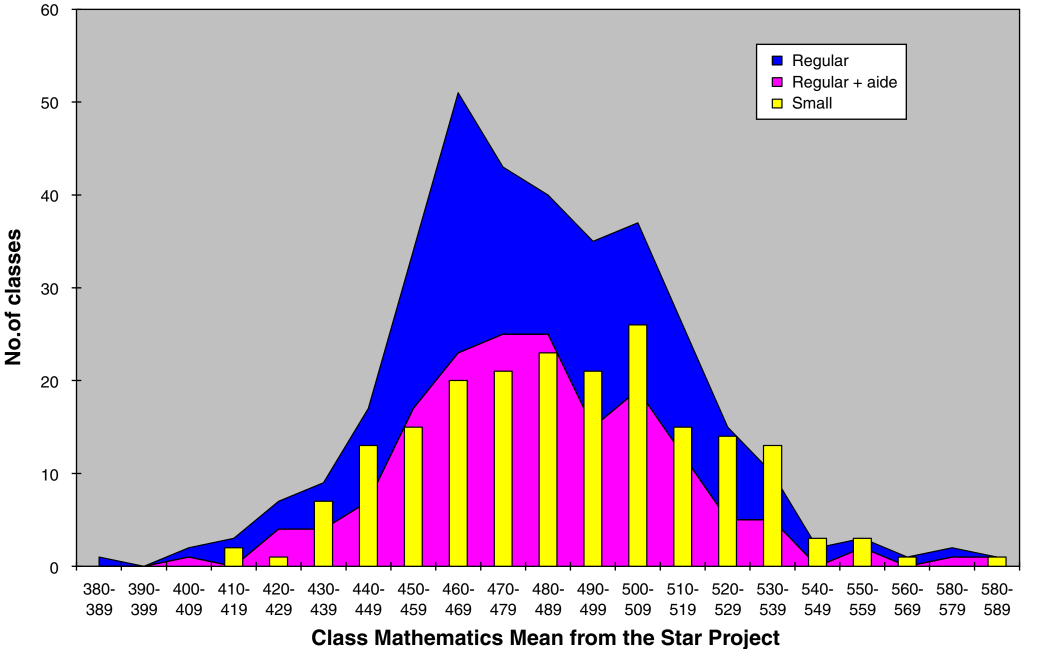

Fig. 10. Means of STAR classes on mathematics.

percent of student-initiated questions, the percent of student-initiated comments, the percent of students off-task, or in time waiting for help in reading or mathematics. They also reported that there were statistically significant differences between the percent of academic contacts but the effects were trivial (7.3% compared with 4.1%), and in the percent of students definitely on-task (87.4% compared with 83.2%). In the Wisconsin SAGE (Student Achievement Guarantee in Education) program, Maier, Molnar, Percy, Smith, and Zahorik (1997) reported that the percent of on-task behavior increased about 8%, and active learning behavior increased 6% (and moreover these were teacher ratings). Small indeed.

For social behaviors, in Project STAR there were statistically significant differences in the percent of behavioral contacts but the differences were small (7.3% compared with 11.4%), in the "amount of disruptive behavior" (1.6% vs. 2.0%) and the "amount of inappropriate behavior" (1.6% vs. 2.1%) (Evertson & Folger, 1989). Similarly, in the California class size reduction initiative (Stecher & Bohrnstedt, 2000), statistically significant differences were found in the percentage of students "engaged in exclusionary behavior (i.e., forming cliques)" during the most recent school day but this difference was small (19% of students in reduced classes compared to 26% of students in non-reduced classes) and on "disrupted the work of other students" (19% of students in reduced classes compared to 23% of students in non-reduced classes). All small differences.

In the Project STAR findings of classroom behaviors, Evertson and Folger (1989) reported that students in smaller classes initiated more contacts with the teacher for purposes of clarification, gave more answers to questions that were open to the whole class, more often contacted the teacher privately for help, were more on-task, and spent less time

Table 3
Arguments supporting and not supporting reductions in class size

	Argument	Supported	Not supported
Student	On task behavior	Cahen et al. (1983)	Blatchford et al. (2005), Bourke (1986), Shapson et al. (1980)
	Amount of individual attention		Bourke (1986), Hargreaves et al. (1998)
	Amount of individual work		Blatchford et al. (2005)
	Interactions more task related	Achilles et al. (1994), Blatchford et al. (2002)	
	Help other students complete work		Blatchford et al. (2005), Rice (1999), Stasz and Stecher (2000)
	Interactions more passive (listening)	Blatchford et al. (2002)	
	Student more on-task time	Blatchford et al. (2002), Cahen et al. (1983), Chase et al. (1986), Maier et al. (1997)	Evertson and Folger (1989), Shapson et al. (1980), Stasz and Stecher (2000)
Student to student	More pro-social behavior/less disruptions	Blatchford et al. (2002), Evertson and Folger (1989), Molnar et al. (1999), Stasz and Stecher (2000)	Chase et al. (1986)
	More interactions with other children		Blatchford et al. (2002), Stecher and Bohrnstedt (2000)
Student to teacher	More teacher interaction with students	Blatchford et al. (2002), Bourke (1986)	
	Student initiated questions		Bourke (1986), Evertson and Folger (1989), Stasz and Stecher (2000)
	Seeking assistance for school work	Blatchford et al. (2002), Johnston (1990)	
Teacher to student	More positive and less negative comments	Krueger (2003)	
	Greater monitoring of whole class	Hargreaves et al. (1998)	
	Less time on administration	Bourke (1986), Hargreaves et al. (1998), Molnar et al. (1999), Stasz and Stecher (2000)	
	Time spent waiting for help		Cahen et al. (1983), Evertson and Folger (1989)

Table 3 (*continued*)

	Argument	Supported	Not supported
	Less whole class teaching		Bourke (1986), Evertson and Folger (1989)
	Greater coverage of material		Betts and Shkolnik, (1999), Shapson et al. (1980)

waiting for the next assignment (see also Achilles, Kiser-Kling, Owen, & Aust, 1994; Kiser-Kling, 1995). This was despite teachers not changing their fundamental teaching strategies when given a small class (Finn & Achilles, 1999). Finn, Gerber, Achilles, and Boyd-Zaharias (2001) found that teachers in STAR small classes spent increased time in direct instruction, although less time on managerial organizational tasks (Evertson & Folger, 1989; Achilles et al., 1994).

In a counterargument to claims of increased individual attention, Buckingham (2003) has determined that if half each 5-h school day is spent directly addressing the class, and the other half on individual attention, each child would hypothetically receive 6 min of individual instruction in a class of 25 and 7.5 min of individual instruction in a class of 20. That is, an extra 1.5 min per day of teacher's time. Buckingham noted that these calculations may be simplistic, but indicate the insubstantial change in individual attention that a 20% reduction in class size brings.

In Canada, Shapson, Wright, Easom, and Fitzgerald (1980) examined 62 fourth- and fifth-grade classes in which students were randomly assigned to class sizes of 16, 23, 30, or 37 students. Classroom observations revealed virtually no consistent differences in classroom practices in smaller versus larger classes. For example, there were no differences between class size and student engagement or "classroom atmosphere". Similarly in Australia, Bourke (1986) found no relationship between class size and student engagement, but teachers in large classes spent more time on classroom management. However, time on classroom management was, by itself, unrelated to achievement. He found more whole-class teaching and not more individualization in smaller classes. Smaller classes were also associated with fewer teacher–student interactions, fewer student questions, less teacher lecturing, and more probing and waiting for responses when teachers asked students questions.

In a national survey covering similar grade levels, Betts and Shkolnik (1999) examined 2170 mathematics classes and students and from grades 7 to 12 in the late 1980s through to the early 1990s. They found no association between class size and text coverage and, correspondingly, no more time devoted to new material in classes of one size or another. However, smaller classes tended to spend more time on review. Conversely, "teachers appear to react to increases in class size by guarding the time they devote to new material, while cutting back on review time" (p. 200). Indeed, they demonstrated that teachers could make small classes "considerably more effective if they did not reduce group instruction to the extent that they do" (Betts & Shkolnik, 1999, p. 209). Overall, they argued that, because teachers reallocate their time to such a small extent, this "may explain why it has been so hard in most past research to identify a positive and significant impact of class size reduction on student achievement" (p. 209).

If the above is not convincing of a small differences in the practices of teachers, consider another interpretation. Let us say Blatchford, Bassett, & Brown (2005) are correct and that "in small classes children were more likely to interact with their teachers, more one-to-one teaching took place, children were more often the focus of teacher's attention, more teaching took place overall, and children more often attended to their teachers, … and individual children in small classes received more interactions with their teachers of a task-related nature" (p. 8). Then why is the overall effect-size a tiny 0.13?

4.3. Teachers tend to use the same teaching methods regardless of class size

In a fascinating history of teaching over the past 150 years, Cuban and Tyack (1995) have noted that about 80% of teachers have not changed the fundamentals of classroom teaching. The most common teaching method has remained a didactic method that attends to the "grammar of schooling" (see also Tyack & Tobin, 1994) or what Sarason (1982) calls the school "regularities". The "grammar of teaching" involving group monitoring, much teacher talk, didactic instruction, whole class and occasionally group teaching. Yair (2000) monitored many classrooms and reported that about 80% of time is spent with the teacher talking. The claim is that this method of teaching is efficient, well-rehearsed, and functional and has stood the test of time as an effective method of instruction for most teachers. A further 15% of the teachers aim to be much more efficient in their teaching but the fundamental grammar of teaching is not that different and only about 5% could be considered truly innovative and departing from this fundamental model of teaching. Given this history of teaching, perhaps we should not be so surprised that the overwhelming conclusion to be drawn from the effects of reducing class size is that the nature of the teaching does not change. Teachers teach similarly in classes of 15 as in classes of 30, and the grammar of schooling does not change (for the teacher or student). We therefore should not be surprised to find that teachers teach and students learn similarly regardless of class size.

Glass, Cahen, Smith, and Filby (1982) also reported that the nature of the instruction rarely changed when classes were reduced from 40 to 20 students. A poor/good teacher with 30 students may remain a poor/good teacher even only with 20 students. In the Shapson et al. (1980) Canadian study, the teachers expressed more positive attitudes with the smaller classes and were more pleased with the ease of managing and teaching in the smaller-class setting. However, "the observation of classroom process variables revealed very few effects of class size. Class size did not affect the amount of time teachers spent talking about course content or classroom routines. Nor did it affect the choice of audience for teachers' verbal interactions; that is, when they changed class sizes, teachers did not alter the proportion of their time spent interacting with the whole class, with groups, or with individual pupils" (Shapson et al., pp. 149–150). No differences were found in student satisfaction or affective measures, teacher activities, subject emphasis, classroom atmosphere, or the quality measures.

Even in smaller classes teachers continue to teach as if there were 30 students in front of them. In a sense, why should they not: What has worked for them (successfully at least in their eyes) in one context is surely adaptable for many other situations. Finn and Pannozzo (2004) claimed that there is not much evidence showing change in instructional practices in smaller classes (see also Evertson & Folger, 1989; Shapson et al., 1980), and the overall structure of lessons, teaching practices, or content coverage is similar (e.g., Bohrnstedt, Stecher, & Wiley, 2000; Cahen, Filby, McCutcheon, & Kyle, 1983; Molnar et al., 1999; Stasz & Stecher, 2000; Wang & Stull, 2000). After reducing class sizes in California, Stasz and Stecher (2000) concluded: "This study's findings are also consistent with research that suggests teaching practice is resistant to change and that teachers adapt their practices slowly and marginally as new materials and techniques are introduced" (p. 29).

Cahen et al. (1983, p. 119) found that teachers relied on a similar format for most lessons regardless of class size. Both before and after the class size change, teachers most often gave brief introductions and directions for lessons rather than incorporating motivational discussions, demonstrations, or other activities. They did not provide opportunities for students to become involved in expressive activities or in small-group or individual projects. There were no changes in instructional content or in the way content was presented. In quantitative comparisons, the only instructional differences observed were that students spent more time academically engaged, and less time off-task, in the smaller classes. Even a summer professional development institute did not lead to instructional modifications among teachers in smaller classes. Evertson and Randolph (1989, p. 102) concluded, "our findings show that teaching practices did not change substantially regardless of class type assignment or training condition."

Hargreaves et al. (1998) also concluded from their British study of classroom interaction differences that the successful teachers of the larger classes had "difficulty in maximizing the opportunities offered in the small class setting, largely because they were unfamiliar with having to cope with such small numbers" (p. 791). They observed seven pairs of teachers over 28 teaching sessions: half in classes of 22–32 and the others in classes of 14. There were more task and challenging questions in smaller classes and more routine/class management questions in larger classes.

Stasz and Stecher (2000) surveyed teachers from 158 small (about 20 students per class) and 236 larger (30 students) K-3 classes from the California Class Size Reduction (CSR) initiative. They found few differences in classrooms, and concluded that "these data revealed no clear patterns that distinguish non-reduced and reduced size classes" (p. 319) on content coverage or time on topics, on time working with individual students, although in smaller classes they spent more time with the poorer readers, and discussing individual students' personal concerns. There were no differences in time allocations for completing paperwork, assessing and monitoring students, or reviewing homework (Table 4).

One remedy to this lack of change in teaching methods to optimize the new opportunities provided by fewer students per class would be to provide extensive professional development. The evidence, however, is not positive. In Project STAR, Ehrenberg, Brewer, Gamoran, and Willms (2001) found that few teachers modified their classroom practices in different size classes even after attending a professional development program.

In the SAGE study, teachers were interviewed and regardless of class size, teachers identified the same factors as most effective in their classrooms: the power of teaching time, individualization, student engagement, content coverage, and hands-on activities (Molnar et al., 1999). The teachers of the 15:1 classes claimed that the major benefits of smaller classes were their knowledge of students, student's knowledge of their peers, reduction in class discipline problems, and more time for helping students acquire common content or skills. Although not supported by the evidence, teachers claimed that "students are moving through content at a much faster pace … much farther along in textbooks, sometimes even using textbooks that are usually reserved for the next higher grade, …. were (more) able to expand and deepen students' learning, … able to add breadth to the content in terms of

Table 4
Differences in time allocations for various tasks for larger and smaller classes

Dimensions	Larger classes		Smaller classes		
	Mean	SD	Mean	SD	Effect-size
Disciplining students	1.6	0.51	1.4	0.51	−0.39
Monitoring/assessing progress	2.8	0.53	2.9	0.58	0.18
Doing paperwork	3.1	0.52	3.0	0.57	0.18
Diagnosing individual learning needs	2.0	0.55	2.1	0.68	0.16
Reviewing homework	2.1	0.52	2.2	0.64	0.17
Providing individual feedback	2.3	0.50	2.4	0.60	0.18
Addressing individual student's personal concerns	1.8	0.47	1.9	0.59	0.19
Discussing student-initiated topics not part of the planned lesson	1.6	0.48	1.6	0.51	0.00

new topics of interest to the students, including greater attention to inquiry and personal learning skills, and they were able to dwell on a topic and pursue it in depth'' (Molnar et al., 1999, p. 175). The authors concluded that the major change was NOT the adoption of more student-centered teaching: ''Teachers do not suddenly permit students to set goals or decide on learning activities, nor do they install a problem-solving approach rich with resources and manipulatives'' (p. 176). The same techniques and methods were used in regular and reduced classes; instead the teacher-centered approach that they had been using is recycled for fewer students.

Rice (1999) completed a similar analysis but used the National Education Longitudinal Study (NELS) database. She used a sample of 12,000 eighth-grade students from the 1998 NELS based on 24,599 randomly selected students in 4932 mathematics and 3828 science classes. Small classes were defined as less then 20 and large classes greater than 20. The effect of reducing large to small classes was 0.04 for mathematics and −0.08 for science. She found that class size does not appear to influence the instructional strategies in science or mathematics classes. The effects were more pronounced in classes of higher-ability students, suggesting that the teachers do not change their instructional practices for lower-achieving students no matter what the class size. Reducing class size led to more time spent working with small groups and using innovative instructional practices, and more time devoted to whole-group discussion (although the effect was tiny once classes exceeded 20 students), but no differences in time spent on administrative tasks. Of interest, they found that teachers needed to spend more time planning for smaller than larger group instruction and as small group instruction occurs more often in smaller classes, the time for planning can go up.

In the Shapson et al. (1980) Canadian study, the only variables that showed differences relating to class size were teachers' opinions and attitudes. Observations showed that class size did not affect the amount of time teachers spent talking about course content, on classroom routines, the choice of audience for teachers' verbal interactions (whole class, group, individuals), and they concluded that there were ''virtually no changes in methods of instruction used by teachers in the different class sizes'' (Shapson et al., 1980, p. 150). While there were small differences favoring smaller classes for mathematics-concepts there were no differences in reading, vocabulary, mathematics-problem solving, or in those subjects that have been claimed as more sensitive to class size such as art and composition.

There were no effects for students' attitudes toward school, for their self-concepts, or for their participation in classroom tasks.

In Bourke's (1986) Australian study, larger classes more often used class groupings, had more student questions (usually seeking help or clarification), but in smaller classes there was more teacher follow-up of questions and probing with greater wait time before teachers answered. Smaller classes were more likely to have more homework and assignments, and more direct interaction, whereas larger classes had more interactions overall between teachers and students. The extent to which teachers individualized their methods was not related to class size, and in smaller classes the teachers more often taught the class as a whole.

4.4. The influences of peer effects

The most likely explanation for the increases in achievement in smaller classes is that smaller classes enhance engagement and reduce inattention and withdrawal behaviors (Finn & Achilles, 1999; Finn, Pannozzo, & Voelkl, 1995; Lamborn, Brown, Mounts, & Steinberg, 1992; McFadden, Marsh, Price, & Hwang, 1992). Finn et al. (2001) undertook the most comprehensive review of potential reasons to account for the small overall effect-size when reducing class sizes. Their claim was that the major effects of smaller class sizes related to the "visibility of the individual". In smaller classes there is increased pressure to participate: "students in a small class can't easily avoid being noticed and the teacher cannot readily ignore any pupil(s) even if s/he would like to" (p. 10). As a consequence they are more likely to take responsibility for their learning, be less involved in social loafing, have a greater sense of belonging, and higher levels of group cohesiveness. Thus, if any effects accrue from smaller classes, the reasons relate to what students rather than teachers do in smaller versus larger classes. However we need to be cautious about this explanation given that Blatchford et al. (2005) argued against Finn et al.'s (2001) interpretation, noting that there was no effect of class size on student on- or off-task behavior.

5. A resolution to this seeming paradox of reducing class sizes and tiny effects on increasing student learning

The major claims so far are that the effects of reducing class size are tiny, which should not be surprising when teachers appear to not change their teaching behaviors even when provided with opportunities in smaller classes to teach in different ways. I wish to advance a series of arguments that may assist in moving the literature forward away from "What are the effects of reducing class sizes?" and "Why the small effect-sizes?" to "What are the conditions for optimizing student learning in smaller classes?" This move is primarily because of the overwhelming press by teachers and parents seeking to reduce class-size despite the overwhelming mass of evidence that it does not have a major effect on learning outcomes. If such a press is successful, as it clearly has been, then it may be worthwhile contemplating the conditions that may optimize student learning in small classes.

The first observation is that the concept of "student learning" is often assumed but not necessarily agreed. For some it means achievement in curricula domains (such as reading, mathematics), for others it means retaining interest in learning whatever the subject, and for others it means on task behavior regardless of any changes in test scores. Across and within cultures the notions defining student learning can be contested (Alexander, 2000). For most of the evidence reviewed in this article, student learning relates to achievement scores (certainly this is the case for the more empirical studies and for the meta-analyses) and/or desirable attributes of behavior (such as the values and dispositions towards involving the student in learning activities).

5.1. The relationships between class size and the concept of excellent teaching

There appears to be a relationship between the notion of excellent teaching and various class sizes. For classes of 80+ it is probably necessary to assume that individual students are already self-regulated to learn and the major tasks for teachers are to provide content; interpretation of this content; and assess students on the facility to absorb, and (slightly) transform this content into their words and beliefs (via structured essays or multiple choice exams). A perusal of student evaluations of teaching of such classes (most evident at the university level) shows the high desirability of organized lectures and lecturers, clear expectations of the exam system, provision of notes and resources, and a well signposted guided tour through text books, syllabi and assessments. At the tertiary level, there is plenty of evidence supporting the conclusion that size of class does not make a difference (Eash & Bennett, 1964; Feldhusen, 1963; Marsh, Overall, & Kesler, 1979; Maxwell, & Lopus, 1995; McKeachie, 1962); and where there are differences in learning outcomes they favor larger classes; although when students and teachers are interviewed they prefer smaller classes. A further and critical issue with these class sizes, is that the concept of "student learning" is more related to scores on achievement tests (assignments, end of course assessments) and less related to developing values, dispositions or learning strategies.

When classes move to the 30–80 size, the concept of excellent teaching is the close following of scripts and chalk/whiteboard lessons, no toleration of deviant behavior in the class, over-learning the rules of classroom behavior, more rigid forms of discipline that allow for little deviance, copying and high amounts of rote learning, straight rows, all walking through the lessons at the same pace (see Cortazzi & Jin, 2001).

In classes of 20–30 grouping becomes possible. There is more opportunity to group students according to ability, to encourage peer interactions, to allow for different proficiencies of self-regulation, and some tailoring of curricula to students (either in topic or pace). There is already a wealth of literature as to the profile of excellent teachers and how they differ from experienced teachers in this size of class (e.g., Berliner, 1987, 1988; Borko & Livingstone, 1990; Chi, Glaser, & Farr, 1988; Hattie, Clinton, Thompson, & Schmitt-Davis, 1996; Housner & Griffey, 1985; Krabbe, 1989; Leinhardt, 1983; Livingston & Borko, 1989; Ropo, 1987; Shanteau, 1992; Sternberg & Horvath, 1995; Strahan, 1989; Swanson, O'Connor, & Cooney, 1990; Tudor, 1992; Van der Mars, Vogler, Darst, & Cusimano, 1991; Westerman, 1991; Yekovich, Thompson, & Walker, 1991).

The concept of excellence in classes of 15–20 requires a shift in conception to the development of individuals, allowing students to become teachers (or self-regulators, self-evaluators, or self-learners), generating opportunity for richer tasks in the learning process, and having teachers become more co-learners with their students. Unlike the largest class sizes, the concept of student learning is more likely to shift and be expressed in terms of student values, dispositions, and learning strategies that allow them to become more autonomous lifelong learners.

The argument is that moving from one "level" of class size to another requires a shift in the concept of excellence of teaching—a move from direct (most often transmission) teaching of students (at 80 +) through attending to teaching and learning (at 20–80), to co-working with a cohort of individual students teaching and learning together (Chan, 2005). The shift required by teachers is not merely to adapt their methods as they move across the levels, but a major re-conceptualization of what it means to be excellent as a teacher at the various levels of class size.

Alexander (2000) has undertaken an extensive comparative study between classrooms in five countries (England, USA, India, Russia, and France). The contrast across these systems helps identify the nature of the dimensions whereby teaching may differ across the levels: particularly among those from India with class sizes ranging between 36 to 70 and the others between 20 and 30. In a different context, Alton-Lee (2005) used a synthesis of literature to relate various teaching methods to outcomes. I have used these two sources to suggest some key attributes that could be used as markers of the differences between teaching in classes of 20–30 or 15–20 students (Table 5). As classes move from larger to smaller classes then there CAN be different dynamics. Perhaps the most defensive educational approach would be to ensure many of these are not one versus the other, but one and the other.

To assist in confronting the paradox that the countries with the larger class sizes are among those with the higher achievements (Galton, 1998), it is likely that this challenge and "tightness of the specification of the curricula" is often lost as classes become smaller. Often accompanying the move to smaller classes there is an increase in the language of "teacher professionalism" but the mistake may be that too many options are given to teachers to "interpret" the curriculum. Such multiple interpretations can lead to lower expectations, poorer understanding of aspects of curriculum content, and teacher preferences of what and when to teach concepts in the curricula—leading to many students not being taught the rigors of all aspects of the curriculum; and more importantly, there is a major decrease in teachers' sharing conceptions of progression across the years and implementing curricula which means students often "cruise", "drift", or may not be aware that they have missed important knowledge and understanding. It is noted how much more prescriptive and demanding are the curricula in the countries placed highest in the international comparisons in contrast to curricula from Western countries with lower class sizes.

Another conjecture is that it is possible that the reason why the greatest effects of reducing class sizes are located at the Kindergarten and first year levels of schooling may be because teachers of these classes are more likely to adopt the teaching principles and behaviors on the right-hand side of Table 5, and certainly measure their excellence as a function of these attributes. It may not be the age of the children, but the concept of teaching. Ehrenberg et al. (2001) noted:

> Kindergarten and first-grade teachers in particular are especially likely to use small groups, hands-on projects, and rely on personal relationships with students, in contrast to teachers of older children whose instruction consists largely of whole-group lecture, recitation, and seatwork. Under this scenario, smaller classes would be more productive in the early grades but make little difference for achievement later on. (p. 24)

Table 5
Dimensions that may be evident as classes vary from larger to smaller sizes

Dimension	Larger classes	Smaller classes
Curriculum		
Curriculum	Tight and prescribed	Loose and diversely interpreted
Classroom interactions		
Focus of habits	The rite of rote	The rite to roam
Centre of class gravity	Whole class	Individuals in class
Interaction with teacher	Teacher to class	Individual to teacher
Management	For discipline and compliance	For learning and self-regulation
An overt culture of	"Niceness" such that peers are reluctant to challenge and contradict peers	"Inquiry" as students encouraged to challenge and contradict peers
The lesson		
Central part of the lesson	A task or sequence of tasks per lesson	On-going tasks over many sessions
Time frames	Fixed time frames	No specified time frames
Nature of tasks	Incremental, practice, and revision	Restructuring, applying, and enrichment
Routines	Ritualized and routinized	Flexible and negotiated
Rules	Not problematic as clear	Problematic as unclear and negotiated
Assessment	Prerogative of the teacher	Prerogative of the student and teacher
Teaching		
Nature of delivery	Transmission based	Activity based
Teacher as	Facilitator and regulator	Coach
Interaction with teacher	Teacher to class or group	Teacher to individual
Talking	Teacher talking	Student talking
Unit of task	Black/white board	Worksheets
Discipline	Well known and disciplinary	Higher levels of unpredictability
Learning dependent on teacher's ability	To teach groups	To equip learners to learn
Activities		
Involvement in the tasks	Passive to active	Active to active
Level of complexity	Low to adequate cognitive demands	High cognitive demands
Learning task	Procedural and propositional (acquisition of information and understanding)	Conceptual (knowledge of ideas and principles)
Nature of knowledge	Ritual understanding	Principled understanding
Nature of knowing	Knowing that	Knowing how
End point for thinking	Process as a means to an end	Process is an end
Student		
Diversity	Allow for individual tasks	Allow for individual interactions
Student answers	Right or wrong	Received and accepted as part of understanding
Peer learning	De-emphasized	Intensified
Culture resides primarily in the	Group	Individual

It is likely that some teachers are more adept at teaching with the rightmost attributes from Table 5, and maybe an optimum school could consist of students moving in and out of smaller classes with teachers teaching in this manner. It is unlikely that all teachers would be excellent at such teaching in smaller classes, and it is certainly the case that in those instances of excellence of teaching in smaller classes there is just as high a work load, levels of stress and coping, and work demands as teaching in larger classes (see Glass & Smith, 1978). An alternative more radical solution (suggested by a reviewer) would be to assign to small classes those promising new teachers who had received pre-service education that emphasized the practices believed to be both most effective for student learning and best suited to small classes. After they had developed skill in these smaller classes, fostered by mentoring by experienced teachers who also used these practices, these young teachers could be given progressively larger classes and encouraged to maintain with necessary adaptations the practices that they had mastered in the small classes. These could achieve two purposes: to ensure that the best practice of small classes is transferred into the more normal 20–30 sizes of class, and also to reduce the socializing effects of many experienced teachers who are more expert in transmission teaching (usually associated with the larger classes as noted above) who pressure many new teachers to perpetuate the 'transmission' status quo.

The major criterion for assigning teachers, whatever size of class, needs to move from contingencies focused on working conditions to a focus on student learning outcomes. Hanushek (1997b) has argued that the incentive structures of schools tend not to emphasize student performance, but instead our system is more geared to discussions on working conditions for teachers and students, teacher satisfaction and their belief they are doing a good job, potential approval or disapproval of parents and principals, and school objectives become framed in terms of fairness, "which is viewed as dictating that all classes be uniformly reduced. Thus, reductions in class size are seldom done differentially with an objective of boosting achievement" (p. 305). Further, he suggested that the outcomes of the effects of reducing class size might by very different if schools faced a different set of incentives (maximizing achievement) and were more selective in when and how they reduced class sizes.

5.2. The role of teacher quality

It is not sufficient to merely claim that it is teaching quality that matters. For example, Rivkin, Hanushek, and Kain (2000) claimed that variation in teacher quality swamped any competing factors such as class size (see also Hanushek, 1992). Although the influences nearer the top of Table 2 are related to teacher quality, it is sobering to balance this with noting that many of the attributes near the bottom are also teacher related influences. Further, as shown from the 1996 Californian example, merely hiring extra teachers, without attention to quality merely mimics the outcomes as if the class sizes had not changed (Stecher & Bohrnstedt, 2000).

Smith, Baker, Hattie, and Bond (in review) investigated the classroom practices, beliefs, and actions of 65 teachers who had passed the National Board for Professional Teaching Standards (NBPTS) certification (1 SE above the cut score) and who had not passed (1 SE below the cut score). This NBPTS model aims to identify highly accomplished teachers (in classes of 20–30) and there is a growing body of evidence in support of the identification methods, and the effects of these teachers, and the acceptance of the public and fellow teachers to this model (Vandevoort, Amrein-Beardsley, & Berliner, 2004). There were substantial differences across the 13 dimensions with an average effect-size of 0.55 (Fig. 11). As important, the effects of the depth of learning were appreciably different. Seventy-four percent of student work samples in the classes of NBPTS certified teachers were judged to reflect a level of deeper understanding, that is Relational or Extended Abstract, and 26% reflected a more surface understanding. This compares with 29% of the work samples of non-NBPTS certified teachers so classified as deep and 71% as surface (Fig. 12).

The question is which of these attributes is likely to be affected by class size and which are more attributes of teachers regardless of the size of class. Table 6 identifies those aspects of teacher quality influenced and less likely to be influenced by class size. Most are not influenced. It is likely that smaller classes will allow teachers to have a more multi-dimensional perspective of classes (as less must be happening with 15 than 25 students), have more opportunity to anticipate difficulties that students may be experiencing, monitor and provide feedback more frequently and at more depth, and have a higher probability of emphasizing students' responsibility for their own learning (which is also the case with class sizes of 80+). But there is no guarantee that this is so—indeed many of the attributes identified in Table 2 can and do occur regardless of the size of class.

6. Conclusions

When addressing the issue of reducing class size, it seems important to investigate the underlying motivations for teachers and parents. For example, the synthesis of meta-analyses indicated that the presence of disruptive students (even one of them) in a

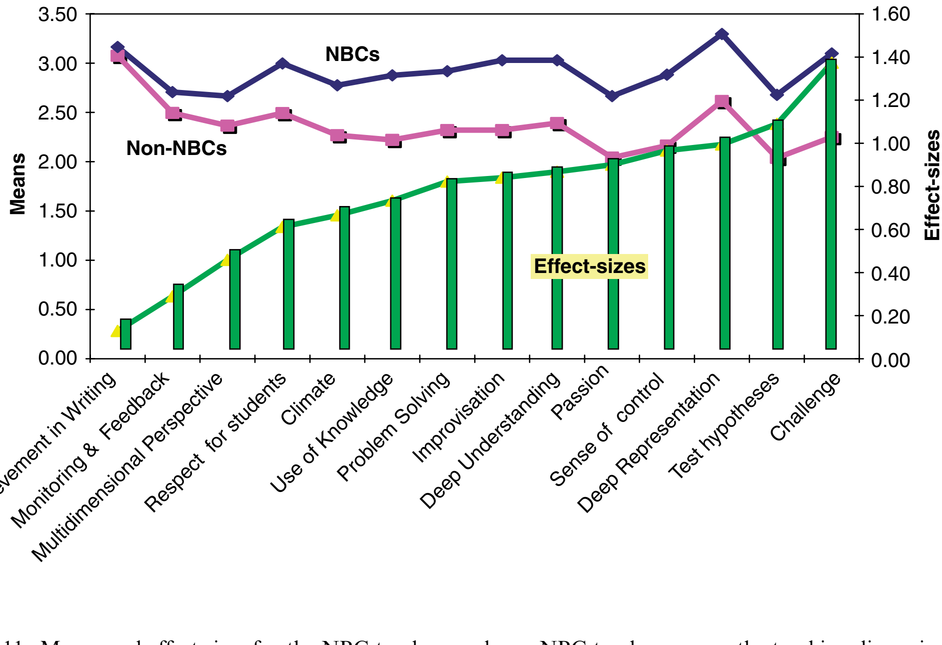

Fig. 11. Means and effect-sizes for the NBC teachers and non-NBC teachers across the teaching dimensions.

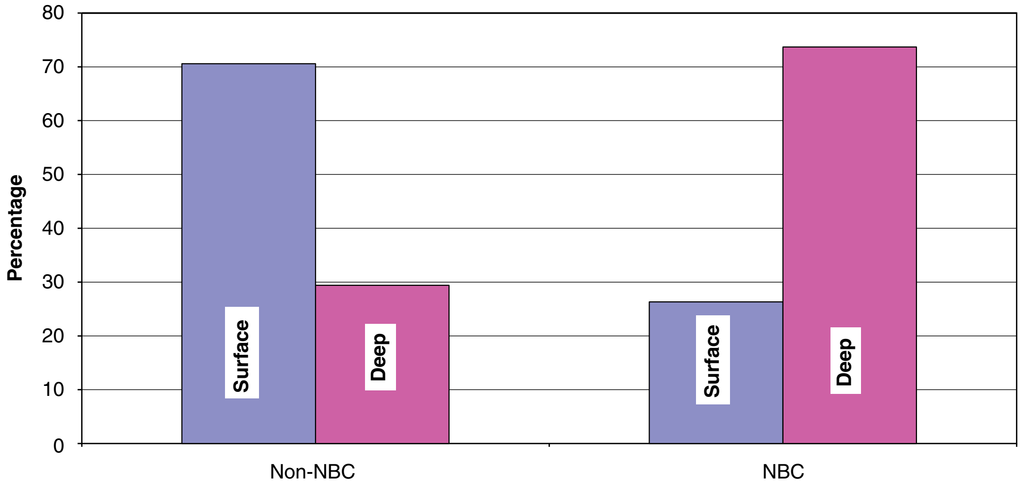

Fig. 12. Percentage of student work assigned surface or deep for the non-NBC and the NBC teachers.

Table 6
Attributes of teachers influenced by class size

Not influenced by class size	Somewhat influenced by class size
Respect for students	Multidimensional perspective of classes
Ensure students understand and follow teachers' management procedures	Anticipate difficulties students may experience
High levels of sense of control	Monitoring and provide feedback
Ability to identify and use the most relevant information	Emphasize students' responsibility for their own learning
Enhance student self-efficacy	
Display passion	
Engage in problem solving about their teaching	
Test hypotheses about student learning	
Deep representation of pedagogical knowledge	
High levels of anticipation and improvisation	
Motivate students towards more challenging goals	
Positive, engaging and rule-clear classroom climate	
Sensitive to the task demands and social situations of the classroom	
Emphasize student efficacy and not indulge in blaming behaviors	
Higher achievement outcomes	
Greater depth of learning	

class has the effect of decreasing achievement by 0.79—which is enormous. When I ask teachers if they would choose between a class size of 15 when I choose the students, or a reduction of 5 from their current class and they choose the students, they nearly always prefer the latter. For many teachers, it is the presence of a few disruptive students that often lead them to desiring smaller classes. There is a question also about the optimal class size; although there seems to be some "magic" in the literature and among policy makers around a class size of 15. When asked in a survey as to the optimal class size, New Zealand secondary teachers claimed that 16 was optimal for Year 13 (the final year of school), 19 for Year 12, 21 for Year 11, and 23 for Years 9–10 (which is not that different from the actual class sizes in NZ secondary schools) (Post-Primary Teachers Association (PPTA), 2005).

The motivation of parents is easier to understand. Their interest is in their own children ($n \cong 1$ or so), and the presence of other children is mostly immaterial to their sense of achievement growth of their child; although there is much evidence that they choose schools on the basis of the type of friends they wish their children to have (Wilkinson, Townsend, Parr, & Hattie, 2002). As schools are forced to cope with more and more students who are disengaged, disinterested, and discontented, these pressures for exclusive learning incubators will increase. Similarly, minimizing the number of disruptive and disobeying students (although some teacher has to get them) may be an effective counterbalance to the stressors experienced by teachers in whatever size of class. In many cases, the claims for reduction in class size seems more related to teacher and student working conditions, whereas the claims against reduction in class size are more related to the lack of evidence of the effects on student learning.

The argument in this paper is that those teaching practices that are conducive to successful learning are more likely to occur in smaller rather than larger classes, and these practices do not actually occur more in smaller classes because teachers have been prepared to, and indeed do, work with larger classes using more transmission practices and therefore they are not so equipped to adopt the more effective practices when they are given smaller classes. (A related argument is that research reviews and/or meta-analyses need not be based on finding a single effect-size which is somehow expected to speak for itself, but that previous research findings can be developed within a theoretically informed model of how the effects may be improved.)

It is conjectured that a way forward in this class size literature is to investigate the classroom and curricula attributes in various levels of class size, the nature of teaching excellence at each of these levels, and the necessity to dramatically alter the concept of teaching excellence when class sizes are changed. Without changing the teaching and ensuring rigor in the curriculum delivery then the effects of this most expensive policy is likely to be close to zero. Maybe identifying teachers who can adopt the nature of teaching that is excellent for smaller classes is the first step, as such teachers could make major and positive effects on student learning: whether this is from 50 to 30, or from 30 to 15. These teachers could work within current schools, although issues of fairness and workload for teachers and students would need to be addressed, and the contingencies for all teachers would need to move from working conditions alone to also embrace positive student learning outcomes in a very public defensible manner.

One of the dilemmas for policy makers, however, is that educational research such as outlined in this paper has had little effect on the lobbying by parents, teacher groups, and politicians for the reduction of class sizes (Achilles, Krieger, Finn, & Sharp, 2003). Further, it is common practice for private "elite" schools to advertise smaller classes as a bonus and feature of their schools, with clear market research showing that this is an attractive feature. Teachers and parents are more convinced that if the working conditions for teaching and learning appear optimal then it is "logical" that benefits must follow, and they seem less convinced of the evidence of the effects of these working conditions, often becoming dismissive of researchers who show evidence to the contrary. If a legislature decides to pour the millions (and cumulatively over a short time, billions) of dollars into reducing class sizes, there are two strategies that may add value to the learning outcomes well beyond the expected 0.13 standard deviation increase in learning outcomes (achievement or non-achievement). First, it is important to identify those teachers who employ excellent teaching methods optimal for smaller classes as outlined above, and provide them with smaller-sized classes (perhaps with many different cohorts of students per week). This introduces the concept of "specialist teachers of small classes." At the same time it is important to acknowledge the excellence of teachers of larger-sized classes, and it likely that there are major issues of "fairness" that will need to be addressed. Second, it is important to identify 3–5 major innovations (e.g., reducing class sizes, reducing class sizes for teachers who teach in a manner as noted above, providing a day per week free for teacher planning and marking, introduction of reciprocal teaching and maximal feedback, etc. See Darling-Hammond and Miles (1998), and Odden (1990) for further possibilities), and set up a comparative research design to assess the effects of these innovations on student learning. Random assignment (as in Project STAR), comparative benefits (as outlined by Levin (1988) and Jamison et al. (1974) noted above), and the use of effect-sizes relative to costs and teacher workload, etc., could be powerful elements, particularly if the study is conducted as a public debate about the benefits of these innovations. Such a debate would inform parents and politicians of the relative effectiveness of the various innovations and it is extremely likely that they could be convinced by the results, thereby moving from seeking different conditions of learning to desiring optimal "effects" on their children. Perhaps the major way forward for any innovation that is as costly as reducing class-sizes is to move the debate from asking does it positively influence student achievement (as Table 2 above showed that almost everything can pass this test) to asking does it positively influence student achievement more than other interventions? The contingencies need to move from focusing not only on working conditions for teachers and students to also focusing outcomes deriving from these working conditions (Hanushek, 2005).

Different systems have other factors that may allow for many more positive effects even when the class sizes are larger. In China for example, Jin and Cortazzi (1998) have argued that the physical conditions of current schools do not permit smaller classes, teachers specialize in primary and high schools and find it difficult to imagine how Western primary teachers teach the entire range of curricula to one class, they teach fewer lessons per week (primary have a maximum of 720 min teaching per week often teaching the same lesson to different classes, whereas Western teachers have more like 1200 min a week with the same class) and thus greater preparation, marking, and teacher-interaction time, and make more visits to other teachers' classrooms and to student homes (Chen, 2005; Ward et al., 1995). There is no doubt that many of the factors identified as the more powerful influences on student learning outcomes (from Table 2) are more likely to occur in class sizes that are smaller; but this is merely a probability statement as they can and do have influence in larger classes.

The issue of class size also needs to be placed into the wider social and cultural domain of any educational system. In the more highly individualistic countries the focus is more on the individual or "autonomous" child, the notion that they are "ready" to move forward at their pace, and there is an imperative to value cultural and individual diversity. Small classes are often seen as more likely to allow these desires to be realized. In the more highly collectivist countries the focus is more on the child in harmony and working in the group, the notion that good teaching is that which outpaces development (Vygotsky, 1978), and the imperative to minimize individual differences in achievement. Class size seems less relevant to these desires.

The International Association for the Evaluation of Educational Achievement (IEA) Third Mathematics and Science Study (TIMSS) noted the importance of ‘‘pedagogical flow’’ that is characteristic of the teaching in each country (Schmidt et al., 1996). This ‘‘flow’’ derives from socialization practices, cultural values about education, and so on, within each educational system. As Biggs (1993, 1998) noted, large class size, apparent authoritarianism, and exam-orientation may co-exist, and changing one aspect (e.g., class size) may need concomitant changes in other aspects of the system. Slightly modifying Bigg's metaphor: What is true of the cuddly English bunny when transported to New Zealand where it wreaked havoc, applies to educational ecosystems (or kudzu from Japan to the USA).

The major issue in any educational system aiming to most expeditiously enhance student outcomes is to focus on those outcomes as driving the system, to ensure that those outcomes are worthwhile and shared by the various constituents, and to then create optimal conditions to maximize the probability that those outcomes can be realized in a cost-, time- and human-effective manner. Reducing class sizes may be but one means to these ends, although it does not appear to be among the most effective of all policies that could be introduced.

References

Achilles, C. M. (1999). *Let's put kids first finally: Getting class size right*. Thousand Oaks, CA: Corwin Press.

Achilles, C. M. (2002). *Class size in the early grades, from inquiry to practice*. Bloomington, IN: Phi Delta Kappa.

Achilles, C. M., & Finn, J. D. (2000). Should class size be a cornerstone for educational policy? In M. C. Wang, & J. D. Finn (Eds.), *How small classes help teachers do their best* (pp. 299–324). Philadelphia, PA: Temple University Center for Research in Human Development in Education.

Achilles, C. M., Finn, J. D., & Pate-Bain, H. (2002). Measuring class size: Let me count the ways. *Educational Leadership*, *59*(5), 24–26.

Achilles, C. M., Kiser-Kling, J., Owen, J., & Aust, A. (1994). *Success starts small: Life in a small class*. Greenboro, North Carolina: University of North Carolina.

Achilles, C. M., Krieger, J. D., Finn, J. D., & Sharp, M. (2003, February). School improvement should rely on reliable, scientific evidence. Why did ‘‘No Child Left Behind’’ leave class size behind? In *Paper presented at the conference within a convention*. New Orleans, LA: American Association of School Administrators (ERIC Reproduction Document No. ED032464).

Akerhielm, K. (1995). Does class size matter? *Economics of Education Review*, *14*(3), 229–241.

Alexander, R. (2000). *Culture and pedagogy: International comparisons in primary education*. Oxford, UK: Blackwell.

Alton-Lee, A. (2005). *Overview of key messages from quality teaching for diverse students in schooling: best evidence synthesis*. Wellington, New Zealand: Ministry of Education. Retrieved 5 May 2005 from 〈www.minedu.govt.nz/goto/bestevidencesynthesis/〉.

Angrist, J. D., & Lavy, V. (1999). Using Maimonides' rule to estimate the effect of class size on scholastic achievement. *The Quarterly Journal of Economics*, *114*, 533–575.

Berliner, D. C. (1987, April). Expert and novice interpretations of classroom data. In *Paper presented at the annual meeting of the American Educational Research Association*, Washington, DC.

Berliner, D. C. (1988, February). The development of expertise in pedagogy. In *Charles W. Hunt Memorial Lecture presented at the annual meeting of the American Association of Colleges for Teacher Education*. New Orleans, LA, February 17–20, 1988. (ERIC Reproduction Document No. ED298122).

Betts, J. R., & Shkolnik, J. L. (1999). The behavioral effects of variations in class size: The case of math teachers. *Educational Evaluation and Policy Analysis*, *21*(2), 193–213.

Biggs, J. B. (1993). From theory to practice: A cognitive systems approach. *Higher Education Research and Development*, *12*, 73–86.

Biggs, J. B. (1998). Learning from Confucian heritage: So size doesn't matter? *International Journal of Educational Research*, *29*(8), 723–738.

Blake, H. (1954). *Class size: A summary of selected studies in elementary and secondary public schools*. Unpublished doctoral dissertation, Columbia University, New York.

Blatchford, P., Bassett, P., & Brown, P. (2005). Teachers' and pupils' behavior in large and small classes: A systematic observation study of pupils aged 10 and 11 years. *Journal of Educational Psychology*, *97*(3), 454–467.

Blatchford, P., Bassett, P., Goldstein, H., & Martin, C. (2003). Are class size differences related to pupils' educational progress and classroom processes? Findings from the institute of education class size study of children aged 5–7 years. *British Educational Research Journal*, *29*(5), 709–730.

Blatchford, P., Goldstein, H., Martin, C., & Browne, W. (2002). A study of class size effects in English school reception year classes. *British Educational Research Journal*, *28*(2), 169–185.

Bohrnstedt, G., Stecher, B., & Wiley, E. (2000). The California class size reduction evaluation: Lessons learned. In M. C. Wang, & J. D. Finn (Eds.), *How small classes help teachers do their best* (pp. 201–226). Philadelphia, PA: Temple University Center for Research in Human Development and Education.

Boozer, M. A., & Rouse, C. (1995). *Intraschool variation in class size: Patterns and implications*. NBER working paper no. 5144, June 1995. Retrieved on 20 October 2005 from 〈http://papers.nber.org/papers/w5144〉.

Borko, H., & Livingstone, C. (1990). Cognition and improvisation: Differences in mathematics instruction by expert and novice teachers. *American Educational Research Journal*, *26*, 473–498.

Bourke, S. (1986). How smaller is better: Some relationships between class size, teaching practices, and student achievement. *American Educational Research Journal*, *23*(4), 558–571.

Boyd-Zaharias, J., & Pate-Bain, H. (2000). Early and new findings from Tennessee's Project STAR. In M. C. Wang, & J. D. Finn (Eds.), *How small classes help teachers do their best* (pp. 65–97). Philadelphia, PA: Laboratory for Student Success at Temple University Center for Research in Human Development and Education.

Brewer, D. J., Krop, C., Gill, B. P., & Reichardt, R. (1999). Estimating the cost of national class size reductions under different policy alternatives. *Educational Evaluation and Policy Analysis*, *21*(2), 179–192.

Buckingham, J. (2003). Class size and teacher quality. *Educational Research for Policy and Practice*, *2*, 71–86.

Cahen, L. S., Filby, N., McCutcheon, G., & Kyle, D. W. (1983). *Class size and instruction*. New York: Longman.

Chan, C. (2005, May). Are small class better? Or what makes small class better? In *Paper presented at the conference on learning effectiveness and class size*, University of Hong Kong, Hong Kong.

Chase, C. I., Mueller, D. J., & Walden, J. D. (1986). *PRIME TIME: Its impact on instruction and achievement (Final report)*. Indianapolis, IN: Indiana Department of Education.

Chen, J. (2005). *Teacher evaluation in China*. Unpublished master's thesis, University of Auckland, Auckland, New Zealand.

Chi, M. T. H., Glaser, R., & Farr, M. J. (Eds.). (1988). *The nature of expertise*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Cortazzi, M., & Jin, L. (2001). Large classes in China: 'Good' teachers and interactions. In D. A. Watkins, & J. B. Biggs (Eds.), *Teaching the Chinese learner: Psychological and pedagogical perspectives* (pp. 115–134). Hong Kong: Hong Kong Comparative Educational Research Center.

Cram, B. M. (1968). *An investigation of the influence of class size upon academic attainment and student satisfaction*. Doctoral Dissertation, Arizona State University. Dissertation Abstracts International, 29/04A, 1066.

Cuban, L., & Tyack, D. B. (1995). *Tinkering toward utopia: A century of public school reform*. Cambridge, MA: Harvard University Press.

Darling-Hammond, L., & Miles, K. H. (1998). Rethinking the allocation of teaching resources: Some lessons from high-performing schools. *Educational Evaluation and Policy Analysis*, *20*(1), 9–29.

Douglass, H. R., & Parkhurst, A. J. (1940). Size of class and teaching load. *Review of Educational Research*, *10*(3), 216–221.

Dustmann, C., Rajah, N., & van Soest, A. (2003). Class size, education, and wages. *The Economic Journal*, *113*(February), F99–F120.

Eash, M. J., & Bennett, C. M. (1964). The effects of class size on achievement and attitudes. *American Educational Research Journal*, *1*(4), 229–239.

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 (whole issue).

Evertson, C. M., & Folger, J. K. (1989, March). *Small class, large class: What do teachers do differently?* Paper at American Educational Research Association, San Francisco.

Evertson, C. M., & Randolph, C. H. (1989). Teaching practices and class size: A new look at an old issue. *Peabody Journal of Education*, *67*, 85–105.

Feldhusen, J. F. (1963). The effects of small and large group instruction on learning of subject matter, attitudes, and interests. *Journal of Psychology*, *55*, 257–262.

Finn, J. D. (1998). *Class size and students at risk: What is known? What is next?*. Washington, DC: US Department of Education.

Finn, J. D., & Achilles, C. M. (1990). Answers and questions about class size: A statewide experiment. *American Educational Research Journal*, *27*(3), 557–577.

Finn, J. D., & Achilles, C. N. (1999). Tennessee's class size study: Findings, implications, misconceptions. *Educational Evaluation and Policy Analysis*, *21*(2), 97–109.

Finn, J. D., Folger, J., & Cox, D. (1991). Measuring participation among elementary grade students. *Educational and Psychological Measurement*, *51*, 393–402.

Finn, J. D., Fulton, D., Zaharias, J., & Nye, B. A. (1989). Carry-over effects of small classes. *Peabody Journal of Education*, *67*(1), 75–84.

Finn, J. D., Gerber, S. B., Achilles, C. M., & Boyd-Zaharias, J. (2001). The enduring effects of small classes. *Teachers College Record*, *103*(2), 145–183.

Finn, J. D., Gerber, S. B., & Boyd-Zaharias, J. (2005). Small classes in the early grades, academic achievement, and graduating from high school. *Journal of Educational Psychology*, *94*, 214–223.

Finn, J. D., & Pannozzo, G. M. (2004). Classroom organization and student behavior in kindergarten. *Journal of Educational Research*, *98*(2), 79–92.

Finn, J. D., Pannozzo, G. M., & Voelkl, K. E. (1995). Disruptive and inattentive-withdrawn behavior and achievement among fourth graders. *The Elementary School Journal*, *95*(5), 421–434.

Fleming, C. M. (1959). Class size as a variable in the teaching situation. *Educational Research*, *1*, 35–48.

Fuller, B. (1987). What school factors raise achievement in the third world? *Review of Educational Research*, *57*(3), 255–292.

Galton, M. (1998). Class size: A critical comment on the research. *International Journal of Educational Research*, *29*, 809–818.

Glass, G. V., Cahen, L. S., Smith, M. K., & Filby, N. N. (1982). *School class size: Research and policy*. Beverly Hills, CA: Sage.

Glass, G. V., & Smith, M. L. (1978). *Meta-analysis of research on the relationship of class size and achievement*. San Francisco: Far West Laboratory for Educational Research & Development.

Goldstein, H., Yang, M., Omar, R., Tumer, R., & Thompson, S. (2000). Meta-analysis using multilevel models with an application to the study of class size effects. *Applied Statistical*, *49*(3), 399–412.

Greenwald, R., Hedges, L. V., & Laine, R. D. (1996). The effect of school resources on student achievement. *Review of Educational Research*, *66*(3), 361–396.

Grissmer, D. W. (1999). Class size effects: Assessing the evidence, its policy implications, and future research agenda. *Educational Evaluation and Policy Analysis*, *21*(2), 231–248.

Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature*, *24*(September), 1141–1177.

Hanushek, E. A. (1992). The trade off between child quantity and child quality. *Journal of Political Economy*, *100*, 84–117.

Hanushek, E. A. (1997a). Assessing the effects of school resources on student performance: An update. *Educational Evaluation and Policy Analysis*, *19*(2), 141–164.

Hanushek, E. A. (1997b). Outcomes, incentives, and beliefs: Reflections on analysis of the economics of schools. *Educational Evaluation and Policy Analysis*, *19*(4), 301–308.

Hanushek, E. A. (1998). *The evidence on class size* (Occasional Paper Number 98-1). Rochester, NY: W. Allen Wallis Institute of Political Economy University of Rochester. Retrieved November 28, 2005, from http://edpro.stanford.edu/hanushek/admin/pages/files/uploads/evidence.size.mayer-pererson.pdf.

Hanushek, E. A. (1999). Some findings from an independent investigation of the Tennessee STAR experiment and from other investigations of class size effects. *Educational Evaluation and Policy Analysis*, *21*(2), 143–163.

Hanushek, E. A. (2003). The failure of input-based schooling policies. *The Economic Journal*, *113*(February), F64–F98.

Hanushek, E. A. (2005, May). Alternative policy options: Class size reductions and other possibilities. In *Keynote Presentation to the Conference on learning effectiveness and class size*. University of Hong Kong, Hong Kong.

Hanushek, E. A., Rivkin, S. G., & Taylor, L. L. (1996). Aggregation and the estimated effects of school resources. *Review of Economics and Statistics*, *78*, 611–627.

Hargreaves, L., Galton, M. & Pell, A. (1997, March). The effects of major changes in class size on teacher–pupil interaction in elementary school classes in England. Does research merely confirm the obvious? In *Paper presented at the American Educational Research Association annual conference*, Chicago.

Hargreaves, L., Galton, M., & Pell, A. (1998). The effects of changes in class size on teacher–pupil interactions. *International Journal of Educational Research*, *29*, 779–795.

Harker, R. (2004, November). Class size and student attainments: Research and strategic implementation. In *Paper presented at the New Zealand Association for Research in Education*, Wellington, New Zealand.

Hattie, J. A., Clinton, J. C., Thompson, M., & Schmitt-Davis, H. (1996). *Identifying expert teachers*. Chapel Hill, NC: North Carolina Association for Research in Education.

Hedges, L. V., & Stock, W. (1983). The effects of class size: An examination of rival hypotheses. *American Educational Research Journal*, *20*, 63–85.

Heyneman, S., Jamison, D., & Montenegro, X. (1983). Textbooks in the Philippines: Evaluation of the pedagogical impact of a nationwide investment. *Educational Evaluation and Policy Analysis*, *6*, 139–150.

Housner, L. D., & Griffey, D. C. (1985). Teacher cognition: Differences in planning and interactive decision making between experienced and inexperienced teachers. *Research Quarterly for Exercise & Sport*, *56*(1), 45–53.

Hoxby, C. M. (1998). *The effects of class size and composition on student achievement: New evidence from natural population variation*. NBER Working Paper Series 6869. Retrieved on 2 May 2005 from ⟨http://www.nber.org/papers/w6869⟩.

Hoxby, C. M. (2000). The effects of class size on student achievement: New evidence from population variation. *Quarterly Journal of Economics*, *115*(4), 1239–1285.

Jamison, D. T. (1982). Reduced class size and other alternatives for improving schools: an economist's view. In G. V. Glass, L. S. Cahen, M. L. Smith, & N. N. Filby (Eds.), *School class size, research and policy* (pp. 116–129). Beverly Hills, CA: Sage.

Jamison, D. T., Suppes, P., & Wells, S. (1974). The effectiveness of alternative instructional media: A survey. *Review of Educational Research*, *44*(1), 1–67.

Jepsen, C., & Rivkin, S. (2002). *What is the trade-off between smaller classes and teacher quality?* NBER Working Paper Series 9205. Retrieved on 2 May 2005 from ⟨http://www.nber.org/papers/w9205⟩.

Jin, L., & Cortazzi, M. (1998). Dimensions of dialogue: Large classes in China. *International Journal of Educational Research*, *29*, 739–761.

Johnson, D., Jensen, B., Feeny, S. & Methakullawat, B. (2004, August). Multivariate analysis of performance of Victorian schools. In *Paper presented in making schools better submit conference*, Melbourne Institute of Applied Economics and Social Research, Melbourne, Victoria, Australia.

Johnston, J. M. (1990). *Effects of class size on classroom processes and teacher behaviors in kindergarten through third grade*. ERIC Document Reproduction Service No. ED321848.

Kiser-Kling, K. (1995). *Life in a small teacher–pupil ratio class*. Unpublished doctoral dissertation. University of North Carolina, Greensboro, North Carolina.

Krabbe, M. A. (1989, March). A comparison of experienced and novice teachers' routines and procedures during set and discussion instructional activity segments. In *Paper presented at the American Educational Research Association annual meeting*, San Francisco, CA.

Krueger, A. B. (1999). Experimental estimates of education production functions'. *Quarterly Journal of Economics*, *114*(2), 497–532.

Krueger, A. B. (2003). Economic considerations and class size. *The Economic Journal*, *113*(February), F34–F63.

Lamborn, S. D., Brown, B. B., Mounts, N. S., & Steinberg, L. (1992). Putting school in perspective: The influence of family, peers, extracurricular participation, and part-time work on academic engagement. In F. M. Newman (Ed.), *Student engagement and achievement in American secondary schools* (pp. 153–181). New York: Teachers College Press.

Legislative Council Panel on Education (2004, February). *Study on effective strategies of class and group teaching in primary schools.* Hong Kong Legislative Council Paper No. CB(2)1282/03-04(01). Hong Kong: Legislative Council Panel on Education.

Leinhardt, G. (1983). *Novice and expert knowledge of individual students' achievement*. Pittsburgh, PA: Learning Research and Development Center, University of Pittsburgh.

Levin, H. M. (1988). Cost-effectiveness and educational policy. *Educational Evaluation and Policy Analysis*, *10*(1), 51–69.

Lewit, E. M., & Baker, L. S. (1997). Class size. *Future Child*, *7*(3), 112–121.

Lindahl, M. (2001). *Home* Versus *school learning: A new approach to estimating the effect of class size on achievement.* IZA Discussion Paper Number 261. Retrieved 2 May 2005 from 〈http://netec.mcc.ac.uk/WoPEc/data/Papers/izaizadpsdp261.html〉.

Livingston, C., & Borko, H. (1989). Expert-novice differences in teaching: A cognitive analysis and implications for teacher education. *Journal for Research in Mathematics Education*, *21*, 372–387.

Ludwig, J., & Bossi, L. J. (1999). The puzzling case of school resources and student achievement. *Educational Evaluation and Policy Analysis*, *21*(4), 385–403.

Maier, P., Molnar, A., Percy, S., Smith, P., & Zahorik, J. (1997). *First-year results of the Student Achievement Guarantee in Education Program*. Milwaukee, WI: SAGE Evaluation Team Center for Urban Initiatives and Research, University of Wisconsin-Milwaukee.

Malloy, L., & Gilman, D. (1989). The cumulative effects on basic skills achievement of Indiana's PRIME TIME—A state sponsored program of reduced class size. *Contemporary Education*, *60*, 169–172.

Marsh, H. W., Overall, J. U., & Kesler, S. P. (1979). Class size, students' evaluations, and instructional effectiveness. *American Educational Research Journal*, *16*(1), 57–69.

Maxwell, N. L., & Lopus, J. S. (1995). A cost effectiveness analysis of large and small classes in the university. *Educational Evaluation and Policy Analysis*, *17*(2), 167–178.

McFadden, A. C., Marsh, G. E., Price, B. J., & Hwang, Y. (1992). A study of race and gender bias in the punishment of school children. *Education and Treatment of Children*, *15*, 140–146.

McGiverin, J., Gilman, D., & Tillitski, C. (1989). A meta-analysis of the relation between class size and achievement. *Elementary School Journal*, *90*, 47–56.

McKeachie, W. J. (1962). Procedures and techniques in teaching: A survey of experimental studies. In N. Sanford (Ed.), *The American College* (pp. 312–364). New York: Wiley.

Molnar, A. (1999). *Vouchers, class size reduction, and student achievement: Considering the evidence*. Bloomington, IN: Phi Delta Kappa.

Molnar, A., Smith, P., Zahorik, J., Palmer, P., Halbach, A., & Ehrle, K. (1999). Evaluating the SAGE program: A pilot program in targeted pupil–teacher reduction in Wisconsin. *Educational Evaluation and Policy Analysis*, *21*(2), 165–177.

Mueller, D. J., Chase, C. I., & Walden, J. D. (1988). Effects of reduced class size in primary classes. *Educational Leadership*, *45*(7), 48–53.

Nye, B., Hedges, L. V., & Konstantopoulos, S. (1999). The long-term effects of small classes: A five-year follow-up of the Tennessee class size experiment. *Educational Evaluation and Policy Analysis*, *21*(2), 127–142.

Nye, B., Hedges, L. V., & Konstantopoulos, S. (2000). The effects of small classes on academic achievement: The results of the Tennessee class size experiment. *American Educational Research Journal*, *37*(1), 123–151.

Odden, A. (1990). Class size and student achievement: Research-based policy alternatives. *Educational Evaluation and Policy Analysis*, *12*(2), 213–227.

Pate-Bain, H., Boyd-Zaharias, J. B., Cain, V.A., Word, E., & Binkley, M. E. (1997, September). *The Student/Teacher Achievement Ratio (STAR) Project: STAR Follow-up Studies, 1996–1997.* Prepared by Health and Education Research Operative Services, Inc. (HEROS).

Post Primary Teachers Association (2005, September). Class size, educational performance and teacher workload: Report prepared by the Class Size Taskforce. In *Paper presented at the annual conference of the Post-primary Teachers Association*, Wellington, New Zealand.

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.

Ritter, G. W., & Boruch, R. F. (1999). The political and institutional origins of a randomized controlled trial on elementary school class size: Tennessee's project STAR. *Educational Evaluation and Policy Analysis*, *21*(2), 111–125.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2000). *Teachers, schools, and academic achievement*. Working Paper No. 6691 (revised). Cambridge, MA: National Bureau of Economic Research.

Ropo, E. (1987, April). Teachers' conceptions of teaching and teaching behavior: Some differences between expert and novice teachers. In *Paper presented at the annual meeting of the American Educational Research Association*, Washington, DC (ERIC Document Reproduction No. ED287824).

Sarason, S. B. (1982). *The culture of the school and the problem of change* (2nd ed.). Boston: Allyn and Bacon.

Schmidt, W., et al. (1996). *A summary of characterizing pedagogical flow: An investigation of mathematics and science teaching in six countries*. London: Kluwer.

Shanteau, J. (1992). How much information does an expert use? Is it relevant? *Acta Psychologica*, *81*, 75–86.

Shapson, S. M., Wright, E. N., Easom, G., & Fitzgerald, J. (1980). An experimental study of the effects of class size. *American Educational Research Journal*, *17*(2), 141–152.

Sitkei, E. G., (1968). *The effects of class size: A review of the research*. Research Report 4, Research Study Series 1967–68. Los Angeles County Superintendent of Schools, Division of Research and Pupil Personnel Services (ERIC Document Reproduction Service No. ED043124).

Slavin, R. (1989). Class size and student achievement: Small effects of small classes. *Educational Psychology*, *24*, 99–110.

Smith, M. L., & Glass, G. V. (1980). Meta-analysis of research on class size and its relationship to attitudes and instruction. *American Educational Research Journal*, *17*(4), 419–433.

Smith, T. W., Baker, W. K., Hattie, J., & Bond, L. A construct and consequential validity study of the certification system of the National Board for Professional Teaching Standards, in review.

Stasz, C., & Stecher, B. M. (2000). Teaching mathematics and language arts in reduced size and non-reduced size classrooms. *Educational Evaluation and Policy Analysis*, *22*(4), 313–329.

Stecher, B. M., & Bohrnstedt, G. W. (Eds.). (2000). *Class size reduction in California: The 1998–99 evaluation findings*. Sacramento, CA: California Department of Education.

Sternberg, R. J., & Horvath, J. A. (1995). A prototype view of expert teaching. *Educational Researcher*, *24*(6), 9–17.

Strahan, D. (1989). How experienced and novice teachers frame their views of instruction: An analysis of semantic ordered trees. *Teaching and Teacher Education*, *5*(1), 53–67.

Swanson, H. L., O'Connor, J. E., & Cooney, J. B. (1990). An information processing analysis of expert and novice teachers' problem solving. *American Educational Research Journal*, *27*, 533–556.

Tomlinson, T. M. (1990). Class size and public policy: The plot thickens. *Contemporary Education*, *62*(1), 17–23.

Tudor, M. (1992). Expert and novice differences in strategies to problem solve an environmental issue. *Contemporary Educational Psychology*, *17*, 329–339.

Tyack, D., & Tobin, W. (1994). The "grammar" of schooling: Why has it been so hard to change? *American Educational Research Journal*, *31*(3), 453–479.

Urquiola, M. (2000). *Identifying class size effects in developing countries: Evidence from rural schools in Bolivia*. Washington: World Bank.

van der Mars, H., Vogler, E. W., Darst, P. W., & Cusimano, B. E. (1991, April). Novice and expert physical education teachers: They think and decide differently … But do they behave differently? In *Presented at the AERA national conference*, Chicago, IL (ERIC Document Reproduction Service No. ED336354).

Vandevoort, L. G., Amrein-Beardsley, A., & Berliner, D. C. (2004). National board certified teachers and their students' achievement. *Education Policy Analysis Archives, 12*(46). Retrieved 5 May 2005 from ⟨http://epaa.asu.edu/epaa/v12n46/⟩.

Vygotsky, L. S. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.

Wang, M., & Stull, J. (2000). School characteristics and classroom practice: Smaller versus larger classrooms. *The Center on Education in the Inner Cities Review*, *9*(2), 21–22.

Ward, T., Barr, B., Chai, M., Hua, D., Kong, X., & Lu, H. (1995). *A crane among the chickens? Evaluating a training program for English teachers in China*. Shanghai: Shanghai Foreign Language Education Press.

Wenglinsky, H. (1997). *When money matters: How educational expenditures improve student performance and when they don't*. Princeton, NJ: Educational Testing Service, Policy Information Center.

Westerman, D. (1991). Expert and novice teacher decision making. *Journal of Teacher Education*, *42*(4), 292–305.

Wilkinson, I. A. G., Townsend, M. A. R., Parr, J. M., & Hattie, J. A. C. (2002). Modeling and maximizing peer effects in school. *International Journal of Educational Research*, *37*(5), 521–535.

Word, E., Johnston, J., Bain, H. P., Fulton, B. D., Zaharies, J. B., Lintz, M. N., et al. (1990). *Student/Teacher Achievement Ratio (STAR), Tennessee's K-3 class size study: Final Summary Report, 1985–1990*. Nashville, TN: Tennessee State Department of Education.

Wößmann, L., & West, M. R. (2002). *Class-size effects in school systems around the world: Evidence from between-grade variation in TIMSS*. Retrieved on 5 May 2005 from 〈ftp://repec.iza.org/RePEc/Discussionpaper/dp485.pdf〉.

Yair, G. (2000). Educational battlefields in America: The tug-of-war over students' engagement with instruction. *Sociology of Education*, *73*, 247–269.

Yekovich, F., Thompson, M., & Walker, C. (1991). Generation and verification of inferences by expert and trained non-experts. *American Educational Research Journal*, *28*(1), 189–209.