Class size and student achievement in the United States: A meta-analysis

Article *in* KEDI journal of educational policy · January 2009

In-Soo Shin
Dongguk University

Jae Young Chung
Ewha Womans University

KJEP 6:2 (2009), pp. 3-19

Class size and student achievement in the United States: A meta-analysis

In-Soo Shin
Yon-Sei Educational Research Center, Korea
Jae Young Chung
Ministry of Education, Science and Technology, Korea

Abstract

The purpose of this meta-analysis is to review the effect of class size on student achievement. Class size reduction (CSR) is one of the most important and interesting educational issues in the world, especially since it is one of the most expensive choices in educational policy. In this meta-analysis, 17 studies (8 published and 9 unpublished) were analyzed using the random-effects model. The results of this review suggest that student achievement in small classes is better than that of large classes by .20 standard deviations.

A fixed-effects categorical analysis was also conducted to find the sources of variance and moderator variables that predict the effects of CSR. First, effect sizes were higher in published studies than in unpublished studies. Second, in terms of school subjects, the results of CSR were generally positive. Third, the effect of CSR on student achievement was larger in elementary schools than in secondary schools. Fourth, the results of CSR were generally positive, except in 10^{th} grade. Fifth, the results of CSR are mixed, but generally positive by location of states.

Keywords: class size, class size reduction (CSR), student achievement, research synthesis, meta-analysis

Introduction

Class size reduction (CSR) to increase student achievement is an approach that has been tried, debated, and analyzed for several decades. CSR is one of the most important educational and interesting issues in the world today, just as it has been for many of the past hundred years (Webb, 2003). Many countries have tried to improve their public education systems and raise school effectiveness through CSR, and the growing interest in the productivity of public schools has fueled a renewed interest in class size as a potential policy lever for improving student achievement (Rice, 1999). Nearly all educational systems need improvement; however, there is no consensus on the course that should be followed and on the need for allocating additional expenditures to produce the desired improvements (Hedges, Laine, & Greenwald, 1994).

Since CSR assumes that smaller classrooms equal better students, it presents an economic obstacle in education. In fact, CSR policy is one of the most expensive policy choices in the educational policy arena because it requires more classrooms and teachers. The core question is: is a small classroom better? Because more classrooms are needed, CSR may require more schools, which leads to an increase in construction costs. Because more classrooms are needed, more teachers are needed, which further increases costs. Brewer, Krop, Gill and Reichardt (1999) estimated the operational costs of nationwide CSR programs under various policy alternatives; these costs are estimated to vary from about $2 billion per year to over $11 billion per year. These estimates could be further by increased teacher salaries and so on.

In the United States (U.S.), four states have conducted representative statewide CSR experiments: Indiana, Tennessee, California, and Wisconsin. The CSR policy in Indiana is called PRIME TIME, and it involves the reduction of Pupil Teacher Ratios (PTR) in grades K-3. PRIME TIME commenced as a pilot program in 1981 with an aim to improve the quality of the early schooling experience (Kiger, 2000). Begun in 1985, Tennessee's CSR policy was based on the Student-Teacher Achievement Ratio (STAR). STAR involved three types of class size: small (a classroom of 13-17), regular (22-26 students), and regular plus a teacher's aide. STAR is the most famous CSR policy in the U.S., not only because of its scale, but also because it was studied using an experimental design. Begun in the 1996-97 school year, Wisconsin's CSR policy is called the Student Achievement Guarantee in Education (SAGE). The program was designed as a 5-year K-3 pilot project. CSR policy in California was also adopted during the 1996-97 school year. California's CSR reform measure aimed at reducing class size from 28 to a maximum of 20 students. The objective of this meta-analysis is to review the effect of class size on student achievement and to conduct a systematic review, evaluation, and synthesis of the body of relevant studies. Although the relationship between class size and student achievement has been studied for a long time, the findings are inconsistent among the primary studies. We have investigated studies from the 1990s because previous meta-analyses on this topic have examined earlier work (e.g., Glass & Smith, 1979). This meta-analysis will provide important information to researchers and policy makers.

The class size literature

The history of class size studies and policies in the U.S. is quite long, beginning in the 1930s; therefore, many researchers have synthesized studies on class size and student achievement. Blake (1954) analyzed 85 class size studies dating to before 1950, which dealt with elementary and secondary school students. From those 85 studies, 35 indicated that small classes were better, 18 indicated that large classes were better, and 32 indicated that the data did not support either conclusion. Only 22 of the 85 previously accepted studies met Blake's minimum requirements. Of these studies, 16 favored small classes, 3 favored large classes, and 3 were inconclusive. He used a vote-counting method, however, the vote-counting method is problematic (Hedges & Olkin, 1980).

Using meta-analytic techniques, Glass and Smith (1979) analyzed the results of 77 studies pertaining to the relationship between class size and achievement. Overall, they concluded that "reduced class-size can be expected to produce increased academic achievement" (p. 8). However, Hedges and Stock (1983), who reanalyzed the Glass and Smith data through improved meta-analytic techniques, questioned these conclusions on statistical grounds. Educational Research Services (ERS) (1980) also challenged Glass and Smith's assertions concerning the strong effects of reduced class size. Of the 76 studies cited by Glass and Smith, for example, the ERS asserted that only 14 were well controlled. Furthermore, according to the ERS, 73% of the database for the Glass and Smith conclusions was drawn from only four of these 14 studies.

Research reviews conducted after the Glass and Smith compilation have also produced controversial results. McGiverin, Gilman, and Tillitske (1989) asserted that literature reviewers have largely concluded that the class-size evidence is inconclusive relative to academic benefits (e.g., Berger, 1982; Bourke, 1986; Haddad, 1978; Helmick & Wasem, 1985; Mayhew, 1983; Shapson, Wright, Easen, & Fitzgerald, 1980; Vignocchi, 1980). Robinson and Wittebols (1986) reviewed 100 class-size-research studies conducted from 1950 to 1985 using a related cluster analysis approach that grouped similar kinds of studies together. They concluded that "the effects of smaller classes on student learning, attitudes, and behavior decreases as grade level increases" and "the clearest evidence of positive effects is in the primary grades, particularly kindergarten through third grade, and that reducing class size is especially promising for disadvantaged and minority students" (Robinson & Wittebols, 1986, pp. 18-19). They also explained that if teachers did not change their teaching methods and classroom management in the smaller classes, positive effects could be smaller.

After reviewing research on class size and student achievement, Hanushek (1999) concluded that research showed, "no support for smaller reductions in class size (i.e., reductions resulting in class sizes greater than 13-17 students) or for reductions in later grades if found in the STAR results" (p. 143). However, this conclusion was again based on a 'vote-counting' table of the results of studies, which as noted above is problematic. We would not emphasize these vote counts.

Table 1. Percentage distribution of effect of class size on student performance, based on value-added models of individual student performance

	Statistically significant			Statistically insignificant		
	No. of estimates	Positive (%)	Negative (%)	Positive (%)	Negative (%)	Unknown sign (%)
All value-added studies	78	12	8	21	26	35
Value-added studies within a single state	23	4	13	30	39	13

Note. A positive sign implies that smaller classes enhance student performance.
Source. 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.*

According to the class-size literature, relations between class size and academic achievement have been disputed until now, and they continue to be debated by both educators and researchers. As the ERS (1980) concluded, the ongoing discussions and studies have served to "only confuse the class size issue" (p. 241).

Operational definitions

Class size refers to the actual number of students per class. The definition of class size should be distinguished from teacher-pupil ratio; however, the pupil/teacher ratio can be calculated by dividing the total number of students by the total number of teachers. The teacher-pupil ratio is a concept related to class size, but in practice, it can be very different. For example, in a class with a large number of students, the teacher-pupil ratio could be small if there are many teacher aides, extra-curricular teachers, and so on. This research used class size as our focal concept rather than the teacher-pupil ratio.

The outcome of student achievement can be measured in various ways. For example, it can be measured by retention rates, graduation rates, standardized test scores, and performance test scores. In the long run, wages earned in the labor market are the ultimate measure of success. This research synthesis, however, used standardized test scores as a measure of student achievement.

Methods

The purpose of this study is to analyze the effectiveness of CSR. To better understand the effect of CSR on student achievement, meta-analysis, the quantitative synthesis of research findings, is need. What we know of the outcome achievements of CSR has been learned through 'vote-counting' analyses and class-size meta-analyses conducted 30 years ago. Therefore, meta-analysis can be a useful method for evaluating overall class size studies' findings because the findings of primary studies are sometimes contradictory. The benefit of this meta-analysis is that it gives us the direction and amount of achievement difference between small class and large class students.

Data collection

Our comprehensive search located studies using five databases: Education Resources Information Center (ERIC), Institute of Science Information (ISI), JSTOR (short for Journal Storage), PsycINFO, and University Microfilms International (UMI). We searched for studies reporting the effects of CSR on student achievement. The key terms used for the search were class size, class size reduction, student achievement, and class size reduction effect on achievement. We investigated literatures spanning approximately 20 years, from 1989 to 2008.

We sought primary studies with samples of K-12 students. Studies of college and pre-kindergarten students were excluded. The treatment group was students in a small class, and the control group was students in a regular or large class. The effect size computed for this meta-analysis was the standardized mean difference. Studies written in English and conducted the U.S. were included. Ultimately, the literature search yielded 129 studies: 79 Journal papers, 43 conference papers or reports, and 7 dissertations.

Data evaluation

Characteristics of studies are coded to find potential moderating variables between CSR and student achievement. These characteristics are represented as follows: (a) published or not, (b) grade of students, (c) school subject, (d) location of states (States), and (e) study ID. Ultimately, 17 studies were included for analysis: 6 published and 11 unpublished.

Table 2. Sample studies of this meta-analysis

Author	Year	Source	Random Assign	Data	Subject	Publication	Grade
Achilles	1994	Conference paper	1	STAR	R, M	0	K-3
Butler & Handley	1989	Conference Paper	0	MS	R	0	1, 2
Davis	2000	Dissertation	1	Georgia	R, M	0	1
Finn & Achilles	1990	AERJ	1	STAR	R, M, SS	1	1
Finn & Achilles	1999	EEPA	1	STAR	R, M	1	K-3
Finn, et al.	1989	PJE	1	STAR	R, M, S, SS	1	1
Goldstein, et al.	1998	BERJ	1	STAR	R, M	1	K,1
Haenn	2002	Conference paper	1	NC	R	0	K,1
Johnston, el al.	1990	Report	1	STAR	R, M	0	K-3
Lapsley, et al.	2001	Conference paper	1	IN	R, M	0	4
Lapsley, et al.	2002	Conference paper	1	IN	R, M	0	4
Monlar, et al	1995	EEPA	0	WI	R, M	1	1, 2
Mostellar	1995	Future of children	1	STAR	R, M	1	1
Nye, et al.	1992	Report	1	STAR	R, M, SS	0	5
Peake, K.	2001	Dissertation	0	SC	R	0	2,3
Scudder	2002	Conference paper	1	NC	R	0	1
Wikins	2002	Conference paper	0	WV	SAT, ACT	0	10

Computation of effect sizes

The d statistic is an effect-size indicator that represents the standardized mean difference between small and large class sizes. The equation that we used is as follows:

$$d_i = (X_{small-size} - \overline{X}_{large-size}) / S_{pooled}$$

$$\text{where, } S_{\text{pooled}} = \sqrt{\frac{(n^S - 1)S_S^2 + (n^L - 1)S_L^2}{(n^S - 1) + (n^L - 1)}},$$

n^S and n^L are group sample sizes, and S_S^2 and S_L^2 are group variances.

We computed a total of 120 effect sizes and an average of 7.1 effect sizes per study. Multiple outcomes exist within a study; therefore, reviewers should be careful about the dependence of outcomes within a study. We conducted shifting unit of analysis (Cooper, 1998) and discussed this issue in conclusion section.

Data analyses

Weighted analyses developed by Hedges and Olkin (1985) were used. We used SAS to analyze the above standardized mean difference effect size. The fixed-effects model, random-effects model, and fixed-effects categorical analysis also can be used. To estimate the overall effect size in meta-analysis, reviewers should examine whether or not the effect sizes are homogenous. The results of the homogeneity test will provide information for model selection for analysis and the motivation for categorical analysis to find moderator variables.

The effect sizes of the primary studies were heterogeneous (Q = 719.4, df = 119, p < .05); therefore, we estimated the overall effect size using the random-effects model and compared the effect sizes according to the characteristics of each study (e.g., published or not, grade of students, school subject, and location).

Results

Description of overall effects

The results of the homogeneity test were significant, so we also conducted random-effects model analysis. The random-effects variance (Cooper & Hedges, 1994) is as follows:

$$\hat{\sigma}_\theta^2 = s^2(d) - (1/k)\sum_{i=1}^{k} v_i ,$$

where

$$s^2(d) = \sum_{i=1}^{k} \frac{(d_i - \bar{d})^2}{(k-1)}, \quad v_i = \frac{n_i^e + n_i^c}{n_i^e n_i^c} + \frac{d_i^2}{2(n_i^e + n_i^c)} .$$

The result was that student achievement is better in a small class size than in a large size class by .20 standard deviations. This indicates that students learn more effectively in smaller classes.

Table 3. The overall result of meta-analysis using random effect model

k	-95%CI	*ES*	+95%CI	*SE*
120	.18	.20	.22	.0104

k: Number of effect sizes; *Q*: Homogeneity statistic; *ES*: Effect size; *SE*: Standard error

Fixed-effect categorical analysis

Fixed-effects categorical analysis was conducted to find the source of variability and moderator, which affects the direction and amount of the relation between student achievement and CSR. Categorical variables are as follows: published or not, grade of students, school subject, state, and random design or not.

Publication

In the analysis of fixed-effects categorical analysis by publication status, the mean effect size in the published papers (.21) is higher than that of unpublished papers (.15). Generally speaking, publication bias represents the known difference in the statistical significance of published versus unpublished studies. The results of published studies have a tendency of being large with more positive effects than unpublished studies. In this meta-analysis, reviewers can assume that there is a publication bias with regard to class size studies.

Table 4. The effect sizes by publication status

Publication	*k*	*Q*	*p*-value	-95%CI	*ES*	+95%CI	*SE*
0	44	341.9	$< .05$	.11	.13	.15	.0091
1	76	301.5	$< .05$	.21	.21	.22	.0041

k: Number of effect sizes; *Q*: Homogeneity statistic; *ES*: Effect size; *SE*: Standard error

School subject

In the analysis of fixed-effects categorical analysis by school subjects, the results of CSR are generally positive. In particular, the mean effect sizes of social science (.20), math (.20), and reading (.19) are positive, respectively. However, the effect sizes of writing (-.09), science (.15), ACT (.18), and SAT (-.29), respectively, have some limitation when generalizing the results because they have only small number of effect sizes. The reason for a small number of effect sizes for certain subjects is that statewide tests mainly include reading and mathematics.

Table 5. The effect sizes by school subjects

Subject	*k*	*Q*	*p*-value	-95%CI	*ES*	+95%CI	*SE*
Reading	58	429.4	< .05	.18	.19	.21	.0055
Writing	1	.0	-	-.28	-.09	.09	.0940
Math	34	114.6	< .05	.19	.20	.21	.0062
Science	2	.9	.3	.09	.15	.20	.0265
Social Science	9	26.1	< .05	.18	.20	.23	.0129
ACT	3	.3	.9	-.24	.18	.60	.2152
SAT	3	.0	1.0	-.71	-.29	.14	.2157

k: Number of effect sizes; *Q*: Homogeneity statistic; *ES*: Effect size; *SE*: Standard error

School level and grade of students

In the analysis of fixed-effects categorical analysis by school level, the result of synthesizing elementary schools (.20) is better than secondary schools (-.05). This shows that CSR works better in the elementary school level. We can suppose individualized teaching is more important at the elementary school level; however, the effect size of the secondary school level has some limitation when directly comparing it with the with elementary school level because the numbers of effect size of secondary schools are much less than those of elementary schools. The cause of a small number of effect sizes for secondary schools is that the state governments' CSR experiments were mainly conducted at the elementary school level.

Table 6. The effect sizes by school level

School Level	*k*	*Q*	*p*-value	-95%CI	*ES*	+95%CI	*SE*
Elementary	114	714.0	< .05	.19	.20	.21	.0037
Secondary	6	2.6	.8	-.35	-.05	.25	.1523

k: Number of effect sizes; *Q*: Homogeneity statistic; *ES*: Effect size; *SE*: Standard error

In the analysis of fixed-effects categorical analysis by grade of students, the results of CSR are generally positive, except in 10th grade. Only the mean effect size of 10th grade is negative (-.05). The mean effect sizes of 1st (.23), 2nd (.22), and 5th (.21) grades are relatively higher than the other grades. The numbers of effect size of K, 1st, 2nd, and 3rd grade are larger; therefore, their results are more reliable than the other grades.

Table 7. The effect sizes by grade of students

Grade	*k*	*Q*	*p*-value	-95%CI	*ES*	+95%CI	*SE*
K	12	10.4	.5	.15	.17	.20	.0112
1	48	159.0	< .05	.22	.24	.25	.0062
2	14	167.0	< .05	.18	.20	.23	.0127
3	12	142.0	< .05	.14	.16	.18	.0083
4	8	1.4	1.0	.10	.13	.16	.0137
5	7	7.3	.3	.17	.20	.24	.0137
10	6	2.6	.8	-.35	-.05	.25	.1523

k: Number of effect sizes; *Q*: Homogeneity statistic; *ES*: Effect size; *SE*: Standard error

We performed fixed-effects regression analysis by grade, for the purpose of analyzing the relationship between the mean effect size and the grade of students. In the result of regression, the intercept (.22) is positive, but the coefficient (-.01) is negative. This indicates that CSR effect decreases as the grade increases.

Table 8. The results of fixed-effects regression analysis by grade of students

Standard Parameter	Estimate	Error	*t*-value	*p*-value
Intercept	.22	.02	14.64	< .0001
Grade	-.01	.01	-1.76	.08

Location of States

There are four representative statewide class size reduction experiments: Indiana PRIME TIME, Tennessee STAR, California, and Wisconsin SAGE. Of these, STAR is the most famous CSR policy in the U.S., not only because of its scale, but also because of its well-controlled experimental design. We could say that Tennessee STAR studies are prevailing in terms of quantity and the quality of experimental design. Of all 121 effect sizes, the number of effect sizes of Tennessee is 78, North Carolina is 9, and the number of effect sizes of Indiana and South Carolina is 8 each.

In the analysis of fixed-effects categorical analysis by location of state, the mean effect sizes of Mississippi (.32), North Carolina (.26), Tennessee (.22), Wisconsin (.20), and Indiana (.14) are positive, respectively. This means that CSR worked in those states. On the other hand, the effect sizes of South Carolina (-.64), Georgia (-.07), and West Virginia (-.05) are negative, respectively. This means that the student achievement of small classrooms was lower than that of large classrooms in those states. The effect size of Mississippi (.32) is the largest, but it has some limitations when generalizing the results because Mississippi has only two effect sizes.

Table 9. The effect sizes by location

State	*k*	*Q*	p-value	-95%CI	*ES*	+95%CI	*SE*
GA	3	.6	.8	-.18	-.07	.04	.0543
IN	8	49.0	< .05	.12	.14	.16	.0098
MS	2	2.1	.1	.21	.32	.42	.0550
NC	9	7.5	.5	.16	.26	.36	.0519
SC	8	7.1	.4	-.75	-.64	-.54	.0541
TN	78	300.8	< .05	.21	.22	.22	.0042
WI	6	22.1	< .05	.17	.20	.23	.0165
WV	6	2.6	.8	-.35	-.05	.25	.1523

k: Number of effect sizes; *Q*: Homogeneity statistic; *ES*: Effect size; *SE*: Standard error; *GA*: Georgia; *IN*: Indiana; *MS*: Mississippi; *NC*: North Carolina; *SC*: South Carolina; *TN*: Tennessee, *WI*: Wisconsin; *WV*: West Virginia

Research design

Research design is a very important factor of good measurement and the results of research depend on the method of comparison. To investigate the difference of the results by research design, we divided primary studies as random assignment or not. In the analysis of fixed-effects categorical analysis by research design, the result of synthesizing random assignment (.20) is better.

Table 10. The effect sizes by random assignment of students

Random Assignment	*k*	*Q*	*p*-value	-95%CI	*ES*	+95%CI	*SE*
0	21	244.6	< .05	.08	.11	.14	.0162
1	90	336.3	< .05	.19	.20	.21	.0040

k: Number of effect size; *Q*: Homogeneity statistic; *ES*: Effect size; *SE*: Standard error

Discussion

In a meta-analysis, reviewers assume that every primary study is independent, but if a study has multiple effect sizes, within the study dependence exists. In this meta-analysis, each study has an average of 6.4 effect sizes. If a study has multiple effect sizes, the same sample can be repeatedly used. Repeated use of the same sample is, however, a violation of independent assumption. If we choose an effect size among multiple effect sizes within a study, not to violate independent assumption, this will cause loss of information.

Cooper's (1998) proposed 'shifting unit of analysis.' When researchers estimate a total effect size, they effectively use the study as a unit; however, when researchers estimate effect sizes for sub-groups, they can use the effect size as a unit. When meta-analysts study the effect of class-size on student achievement, there are reading, mathematics, and science achievement outcomes for the same samples. When the analyst measures effect size for each subject area, the unit of analysis is the effect size for each subject; however, when the meta-analyst measures the overall effect size, the researcher will use the study as the unit of analysis for the independence assumption, thus averaging across the subject areas. Shifting units of analysis is a good strategy because no information is lost and the independence assumption is not violated.

This meta-analysis has two dependences. Within the study, dependence exists because each study has 6.4 effect sizes; between study dependence exists because there were 8 STAR studies among 17 studies and there were 78 effect sizes among 120 effect sizes. Shin (2009) proposed that source of 'data' can be a unit of analysis when between studies dependence exists.

Table 10 shows the different results based on the random-effects model. Overall effect size is .20 when treating each 'effect size' independently, but overall effect size is .17 when the unit of analysis is 'study'. If a state was used as a unit of analysis, the effect size is .08. STAR has so many studies in class size researches; however, it is the most controlled experiment among class size studies, and its studies are not the same as each other (Shin, 2009). Reviewers can use effect size as a unit of analysis for this meta-analysis.

Table 11. Dependence and unit of analysis

Analysis Unit	*k*	*Q*	*p*-value	-95%CI	*ES*	+95%CI	*SE*
ES	120	313.1	< .05	.18	.20	.22	.0104
Study	17	106.4	< .05	.12	.17	.21	.0210
State	8	93.9	< .05	.03	.08	.14	.0285

k: Number of effect size; *Q*: Homogeneity statistic; *ES*: Effect size; *SE*: Standard error

Limitations

Although the results of this meta-analysis are clear, we have some limitations. First, in terms of data evaluation, we found too many Tennessee STAR studies, which have 78 effect sizes. We worry about the dependence issue. When the reviewer first retrieved the class size papers, the reviewer was concerned about the same-author and same-data studies' dependence (Shin, 2009). While the reviewer conducted the data analysis, there were multiple effect sizes for several subject test scores, and this causes dependences within study. Between study dependence and within study dependence should be considered in class size meta-analysis. This meta-analysis used 'study' and 'state' as a unit of analysis in Table 11.

Second, many other factors in addition to class size could affect student achievement. Student achievement can be impacted by a family's socio-economic status, a student's IQ, a teacher's skill level, school culture, and so on. In this study, we only analyzed the relation between class size and achievement. Meta-analysis cannot control the research model and design; on the other hand, meta-analysis totally relies on primary studies. We found that well controlled studies have a larger effect in Table 10.

Third, for the purpose of analyzing the effect of CSR correctly, we need to measure the long-term effects of class size. However, most of the class size experiments and studies are small scale and short term. Consequently, it is difficult to determine if the special environment caused the class size effects. It is not easy to be convinced by the results of class size studies, and synthesizing these studies has some limitation. This meta-analysis only synthesized the standardized mean difference of achievements between the students in small versus large classes. The mean change difference effect estimate (Becker, 1988) is the only possible method to measure the longitudinal effect of class size.

Fourth, this meta-analysis synthesized studies of the United States. We will conduct meta-analysis of class size studies conducted in other countries, compared to the United States as a future research project.

Conclusion and implications

Overall, the student achievement of the small classes is better than that of students in large classes by .20 standard deviations. Using fixed-effects categorical analysis, we tried to find the source of variance. First, the results of published papers are stronger than unpublished papers. Second, in terms of school subjects, the results of CSR are generally positive. Third, the results for elementary schools are better than secondary schools. Fourth, the results of CSR are generally positive, except in 10th grade. Fifth, in terms of school state, the results of CSR are mixed.

In this meta-analysis, the effects of smaller classes on student achievement decreases as grade level increases. CSR effect on achievement is better in the early stage of studentship, especially in grades K-3. Therefore, policy makers should focus on this critical time period. If they have limited educational funds, elementary school students would be a better investment.

Many researchers concluded that CSR works better on minority and disadvantaged students. Finn and Achilles (1999) reported minority effects that are approximately double those for white students in grades K-3. Robinson and Wittebols (1986) reported that reducing class size is especially promising for disadvantaged and minority students. Therefore, CSR can be a good educational policy option for minority and disadvantaged students, especially minority and disadvantaged students in grades K-3.

In the result of this meta-analysis, CSR affects student achievement positively. However, this meta-analysis could not project the proper size of class because there is little information in primary studies. In future studies, researchers need to study the appropriate size of class. In population decreasing countries such as Japan, Korea, and EU countries, class size has been reduced naturally. In other developing countries, CSR is one of the most expensive policy choices. CSR is not one best way to improve all public school systems. There are so many reforms that need to occur for the purpose of reinventing public education and raising the effectiveness of school systems. Policy makers must consider the ideal or reasonable class size

Address for correspondence

Jae Young Chung
Deputy Director
Ministry of Education, Science and Technology
77-6 Sejongno, Jongno-gu
Seoul 110-760, Korea
Tel: 82 2 2100 6457
Fax: 82 2 2100 6455
Email: cjy@mest.go.kr

References

References marked with an asterisk indicate studies included in the meta-analysis.

*Achilles, C. M. (1994, November). *The multiple benefit of class size research: A review of STAR's legacy, subsidiary and ancillary studies*. Paper presented at the annual meeting of the Mid-South Educational Research Association, Nashville, TN.

Becker, B. J. (1988). Synthesizing standardized mean-change measures. *British Journal of Mathematical and Statistical Psychology, 41*, 257-278.

Berger, M. A. (1982, April). *Class size is not the issue. National Institute of Education, Washington, DC*. Paper presented at the annual meeting of the National School Boards Association, Atlanta, GA.

Blake, H. V. (1954). *Class size: A summary of selected studies in elementary and secondary schools*. Unpublished dissertation, Teachers College, Columbia University.

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

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.

*Butler, J. M., & Handley, H. M. (1989, November). *Differences in achievement for first and second graders associated with reduction in class size*. Paper presented at the Mid-South Educational Research Association Annual Conference, Little Rock, AR.

Cooper, H. (1998). *Synthesizing research: A guide for literature reviews* (3rd ed.). Thousand Oaks, CA: Sage.

Cooper, H., & Hedges, L. V. (Eds.). (1994). The handbook of research synthesis. New York: Russell Sage Foundation.

*Davis, F. (2000). The effect of class size reduction on student achievement and teacher attitude in first grade. Unpublished doctoral dissertation, Georgia State University.

Educational Research Services. (1980). Class size research: A critique of recent analyses. *Phi Delta Kappan, 62*, 239-244.

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

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

Glass, G. V., & Smith, M. L. (1979). Meta-analysis of research on class size and

student achievement. *Educational Evaluation and Policy Analysis, 1*(1), 2-15.
*Goldstein, H., & Blatchford, P. (1998). Class size and educational achievement: A review of methodology with particular reference to study design. *British Educational Research Journal, 24*(3), 255-268.
Haddad, W. D. (1978). *Educational effects of class size (World Bank Staff Working Paper No. 280)*. Washington, DC: World Bank.
*Haenn, J. F. (2002, April). *Class size and student success: Comparing the results of five elementary schools using small class sizes*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
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.
Hedges, L. V., & Olkin, I. (1980). Vote-counting methods in research synthesis. *Psychological Bulletin, 88*, 359-369.
Hedges, L. V., & Stock, W. (1983). The effects of class size: An examination of rival hypotheses. *American Educational Research Journal, 20*, 63-85.
Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. New York: Academic Press.
Hedges, L. V., Laine, R. D., & Greenwald, R. (1994). Does money matter? A meta-analysis of studies of the effects of differential school inputs on student outcomes. *Educational Researcher, 3*, 5-13.
Helmick, E., & Wasem, L. (1985). *Class sizes for kindergarten and primary grades: A review of the research*. Springfield: Illinois State Board of Education, Department of Planning, Research and Evaluation.
Kiger, D. M. (2000). *A case study evaluation of a fortified class size reduction program*. Unpublished doctoral dissertation, Northern Illinois University.
*Johnston, J., Bain, H. P., Fulton, B. D., Zaharias, J. B., Achilles, C. M., Lintz, M. N., et al. (1990). *The state of Tennessee's student/teacher achievement ratio (STAR) project: Final summary report 1985-1990*. Nashville, TN: Tennessee Department of Education.
*Lapsely, D. K., & Daytner, K. M. (2001). *Indiana's "class size reduction" initiative: Teacher perspectives on training, implementation and pedagogy*. Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.
*Lapsely, D. K., Daytner, K. M., Kelly, K., & Maxwell, S. E. (2002, April). *Teacher aides, class size and student achievement: A preliminary evaluation of Indiana's prime time*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
Mayhew, H. C. (1983). *Related research on class size*. Morehead, KY: Morehead State University.
McGiverin, J., Gilman, D., & Tillitske, C. (1989). A meta-analysis of the relation between class size and achievement. *The Elementary School of Journal, 90*(1), 47

-56
*Monlar, A., Smith, P., Zahorik, J., Palmer, A., 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.
*Mostellar, J. (1995). The Tennessee study of class size in the early school grades. *The Future of Children, 5*(2), 113-127.
*Nye, B. A., Zaharisa, J. B., Fulton, B. D., Wallenhorst, M. P., Achilles, C. M., & Hooper, R. (1992). *The lasting benefit study: A continuing analysis of the effect of small class size in kindergarten through third grade on student achievement test scores in subsequent grade levels: Fifth grade technical report*. Nashville, Center of excellence for research in basic skills, Tennessee State University.
*Peake, K. (2001). *The effect of class size: A study of second and third grade student achievement in the school district of Greenville county, South Carolina*. Unpublished doctoral dissertation, University of South Carolina.
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.
Robinson, G. E., & Wittebols, J. H. (1986). *Class size research: A related cluster analysis for decision-making*. Arlington, VA: Education Research Service.
*Scudder, D. F. (2002, April). *An evaluation of the federal class-size reduction program in Wake County, North Carolina*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
Shapson, S. M., Wright, E. N., Eason, G., & Fitzgerald, J. (1980). An experimental study of the effects of class size. *American Educational Research Journal, 17*, 141-142.
Shin, I. S. (2009). *Between study dependence in meta-analysis*. Unpublished doctoral dissertation, Florida State University, Tallahassee, FL.
Vignocchie, N. (1980). *Research on the effects of class size on academic achievement*. Washington, DC: U.S. Department of Health, Education, & Welfare, National Institute of Education.
Webb, C. E. (2003). *The effects of a class size reduction on elementary student reading and achievement in a Midwestern urban school district*. Unpublished doctoral dissertation, The University of Iowa.
*Wikins, W. E. (2002, March). *No simple solution: Do smaller classes, more experienced and educated teachers, and per pupil expenditure really make a difference?* Paper presented at the Annual International Conference of the Society of Educators and Scholars, San Antonio, TX.