Educational Considerations

Volume 39
Number 1 *Class Size and Student Achievement*

Article 4

9-1-2011

A Practical Method of Policy Analysis by Considering Productivity-Related Research

James L. Phelps

Recommended Citation

Phelps, James L. (2011) "A Practical Method of Policy Analysis by Considering Productivity-Related Research," *Educational Considerations*: Vol. 39: No. 1. https://doi.org/10.4148/0146-9282.1101

A Practical Method of Policy Analysis by Considering Productivity-Related Research

James L. Phelps

The basic notion underlying schooling is rather simple: Hire teachers to instruct students. From there, the tasks become more complicated. How many teachers should be employed? What assignments should the teachers be given, in the classroom or in a supporting role? What assistance should teachers receive from aides or volunteers? What role do administrators play? Schooling is even more than staffing: It includes the curriculum; methods of instruction, instructional materials, time of instruction, and home support including homework. All of these elements must combine into a unifying whole in order to achieve the desired educational goals. Goals other than achievement are important as well, e.g., staying in school, preparation for employment, and civic responsibility, just to name a few. However, the topic must be limited, so this discussion focuses only on the goal of student achievement.

Class size may be important in achievement, but it is not the only decision for policymakers. Class size plays a role, but the role is effectively fulfilled only when the other players are successful. Therefore, it is appropriate to address several questions: What goals are to be accomplished; what is the best distribution of personnel related to these goals; what roles do curriculum, instruction, time, and home support play; and how do the personnel work together effectively to achieve those goals? In the broadest sense, the fundamental question is: How are decisions made?

A Taxonomy of Class Size Decision Making

For the sake of discussion, three levels of decision making related to class size are presented. Generally speaking, there are three broad categories or levels:

(1) Professional and public opinion;
(2) A critical analysis of educational research evidence;
(3) A decision-making process, including: (a) establishing a set of clearly stated goals; (b) identifying a set of possible policy options to achieve the goals; (c) clearly stating the assumptions why each of the policy option would achieve the goals; and (d) evaluating each of the policy options to select the best alternative.

A case could be made that decision making based upon the first perspective is the most common. The premise of this article is to provide some rationale and ideas regarding how policymakers can move through the more sophisticated levels of the taxonomy—the critical analysis of educational research evidence and a structured decision making process. Undoubtedly, policymakers have intuitive answers to the complicated questions encompassing education, but the objective of good policymaking is to explicitly spell out those questions and underlying assumptions regarding the best answers.

- Will lower class sizes make a difference in student achievement? By how much?
- Will an increased number of other instructional staff have a beneficial impact on student achievement? By how much?
- Will effective instructional and organizational policies have a beneficial impact on achievement? If so, by how much?

The purpose of this discussion is to explore the policymaking process by exploring these issues through the research literature. The next article, "A Practical Method of Policy Analysis by Estimating Effect Size," further develops the issues raised here using data from Minnesota. The fourth article, "A Practical Method of Policy Analysis by Simulating Policy Options," suggests a method of policy analysis, based on the ideas and data from the previous articles, in order to investigate possible answers to the questions posed above.

This article is divided into three parts. In the first, Does Class Size Make a Difference: A Brief Overview of the Research,[1] a sampling of studies is presented. It should be noted that some research studies have included variables other than class size. The second section is titled, How Much of a Difference Does Class Size Make on Achievement? The 1978 meta-analysis of Glass and Smith suggested the possibility that achievement increases faster as class sizes become smaller. This study has influenced research and policy ever since. This section examines some of the issues concerning the nature of the relationship between class size and achievement. What is the magnitude of the relationship? What is the nature of the relationship, increasing as suggested by Glass and Smith, or some other pattern? This section notes that some other policy options might improve achievement either independently or in combination with lower class size. The third and final section closes with some observations.

Does Class Size Make a Difference? A Brief Overview of the Research Literature

Clearly, teachers and the public believe that small classes produce higher achievement. Whether their beliefs are supported by evidence is a separate question; nevertheless, beliefs have a major influence on the decision making process. Although the data are somewhat old, Robinson and Wittebols (1986) reported several polls indicating the magnitude of those beliefs. In most cases, lower

---

James L. Phelps holds a Ph.D. from the University of Michigan in Educational Administration. He served as Special Assistant to Governor William Milliken of Michigan and Deputy Superintendent in the Michigan Department of Education. Active in the American Education Finance Association, he served on the Board of Directors and as President. Since retirement, he spends a great deal of time devoted to music, composing and arranging, playing string bass in orchestras and chamber groups, as well as singing in two choirs. He resides with his wife, Julie, in East Lansing, Michigan.

class size was thought to be favorably associated with achievement, discipline, decreased drug use, decreased crime, and increased student motivation. There is little reason to think those beliefs have changed.

Hanushek (1989, 1998, 1999) has researched and written extensively on the issue of class size and its relationship to achievement.[2] He has offered evidence in four ways: (1) interpretation of historical aggregate data; (2) international comparisons; (3) econometric studies; and (4) analysis of controlled experiments. This overview follows the same structure.

Interpretation of Historical Aggregate Data

Substantially more teachers have been added to the U.S. system of education over time with little change in academic performance. Hanushek (1999) presented the changes in aggregate class size between 1960 and 1994, a reduction from about 27 to about 20. In contrast, the measure of achievement, NAEP (National Assessment of Education Progress), showed little change. The analysis went on to account for the changes in student population, changes in special education, and racial differences in achievement. Based upon his analysis, Hanushek (1999, 17-18) concluded:

> The available data and evidence suggest some uncertainty about the underlying forces related to families, school organization, class size, and achievement. Allowing for changes in family background and special education, however, it remains difficult to make a case for reduced class size from the aggregate data. A natural experiment in class size reduction has been going on for a long period of time, and overall achievement data do not suggest that it has been a productive policy to pursue. Nonetheless, the aggregate data are quite limited, restricted to a small number of performance observations over time and providing limited information about other fundamental changes that might affect school success (pp. 17-18).

International Comparisons

There is no systematic relationship between class size and achievement. The international analysis focuses on two examples. The first concerns the Third International Mathematics and Science study (TIMSS) for which the pupil-teacher ratios and achievement scores were correlated. The correlation was positive, higher ratios (more pupils in a classroom) were associated with higher performance, but thought to be a statistical artifact rather than persuasive evidence (Hanushek, 1998, 18).

The second analysis was a more systematic examination of international tests with 70 country-specific measures of pupil-teacher ratios and achievement. According to Hanushek and Kim (1995), the results were positive but statistically insignificant when controlled for parents' schooling. They added:

> Of course, there are many differences in schooling and societies of the sampled nations, so it would be inappropriate to make too much of these results. They do, however, underscore that the normal presumptions about the achievement effects of pupil-teacher ratio and class size are not found in the evidence (p. 19).
>
> Somewhat surprising, similar kinds of results are found if one looks across countries at the relationship between pupil-teacher ratios and student performance. While it is clearly difficult to develop standardized data across countries, to control for the many differences in populations and schools, and the like, there remains some appeal in looking across countries. The variation in class sizes and pupil-teacher ratios are larger than found within the U.S., leading to some hope that the effect of alternative intensities of teacher usage can be better understood. Even given the wide difference, there is no evidence that lower pupil-teacher ratios systematically lead to increased performance (p. 21).

In another study based on the TIMSS achievement measure, Woessmann and West (2002, 7) concluded:

> We estimate the effect of class size on student performance in 18 countries, combining school fixed effects and instrumental variables to identify random class-size variation between two adjacent grades within individual schools. Conventional estimates of class-size effects are shown to be severely biased by the non-random placement of students between and within schools. Smaller classes exhibit beneficial effects only in countries with relatively low teacher salaries. While we find sizable beneficial effects of smaller classes in Greece and Iceland, the possibility of even small effects is rejected in Japan and Singapore. In 11 countries, we rule out large class-size effects.

Econometric Studies

The number of econometric studies with statistically significant results are offset by an almost equal number of statistically insignificant studies. The econometrics studies are based on an input/output regression model controlled for socioeconomic status (SES)

Table 1
Distribution of Estimated Influence of Teacher-Pupil Ratio on Student Performance

School Level	Number of Estimates	Statistically Significant (%)		Statistically Insignificant (%)		
		Positive	Negative	Positive	Negative	Unknown
All schools	277	15	13	27	25	20
Elementary	136	13	20	25	20	23
Secondary	141	17	7	28	31	17

Source: Eric A. Hanushek, "The Evidence on Class Size," Occasional paper 98-1 (Rochester, NY: Wallis Institute of Political Economy, University of Rochester, 1998), 23, Table 5.

http://newprairiepress.org/edconsiderations/vol39/iss1/4
DOI: 10.4148/0146-9282.1101

Table 2
Krueger's Re-Analysis of Hanushek's Meta-analysis

Results (in Percentages)	*Hanushek: Estimates Weighted Equally*	*Krueger: Estimates Weighted by Inverse of Number of Estimates in Study*	*Krueger: Estimates Weighted by Citation Frequency*	*Krueger: Estimates Derived from Regression Analyses of Original Estimates*
Positive and Statistically Significant	14.8	14.4	30.6	33.5
Negative and Statistically Significant	13.4	10.3	7.1	8.0
Statistically Insignificant	71.9	61.2	62.3	58.4

Source: Alan B. Krueger, "Understanding the Magnitude and Effect of Class Size on Student Achievement," in *The Class Size Debate*, edited by Lawrence Mishel and Richard Rothstein (Washington DC: Economic Policy Institute, 2002), 7, Table 1-2.

Table 3.1
Class Size and Student Achievement: Studies Clustered by Grade Level

Grade Level	*Total Number of Studies*	*Studies Favoring Small Class Size*	
		Number	*Percent (%)*
K-3	22	11	50.0
4-8	21	8	38.1
9-12	22	4	18.2

Source: Glen E. Robinson, and J.H. Wittebols, *Class Size Research: A Related Cluster Analysis of Decision Making* (Arlington, VA: Educational Research Services, Inc., 1986), 67.

Table 3.2
Class Size and Student Achievement: Studies Clustered by Reading Achievement

Grade Level	*Total Number of Studies*	*Studies Favoring Small Class Size*	
		Number	*Percent (%)*
K-3	22	11	50.0
4-8	14	5	35.7
9-12	2	1	50.0

Source: Robinson and Wittebols (1986, 71).

Table 3.3
Class Size and Student Achievement: Studies Clustered by Mathematics Achievement

Grade Level	*Total Number of Studies*	*Studies Favoring Small Class Size*	
		Number	*Percent (%)*
K-3	14	5	35.7
4-8	15	6	40.0
9-12	17	0	0.0

Source: Robinson and Wittebols (1986, 80).

and other variables. The data for the studies are not identical in terms of achievement measures, unit of analysis (classroom or school), or measures of SES; thus, they are not always comparable. Some studies deal solely with class size while others include other aspects of education. In each case, there are differences of opinion regarding the method of analysis and conclusions. The evidence here is presented in the form of tables summarizing selected studies on class size (Tables 1, 2, and 3.1-3.3) and education policy studies (Tables 4-5) so that the reader can evaluate the merits of the conclusions.

Analysis of Controlled Experiments

Looking at the evidence one way, the conclusion seems to be class size does not make a difference, and, therefore, it should not be considered for further funding. Looking another way, the conclusion is that class size does make a difference and should be funded. Looking at the evidence a third way, it is reasonable to conclude that instructional quality and time make the largest difference and should be most heavily funded.

- According to Hanushek (1998, 25): "The economic evidence is clear. There is little reason to believe that smaller class sizes systematically yield higher student achievement. While some studies point in that direction, an almost equal number point in the opposite direction. Moreover, restricting attention to the best of these

Table 4
Production Function Studies

Inputs	*Statistically Significant*	*Statistically Insig-nificant*
Teacher Characteristics:		
Verbal achievement	12	3
Experience	24	5
SES background	6	1
Gender	1	0
Salary	17	1
Turnover rate	6	3
Employment status	1	0
Job satisfaction	2	1
Teacher personality	1	0
Professional preparation and academic training	18	11
NTE score	3	1
Policy and Administrative Arrangements:		
Class size	10	5
Pupil teacher ratio	13	6
Size of specific class	5	0
Specific staff to pupil ratio	4	0
Paraprofessional assistance for teachers	2	0
Teacher to administrator ratio	2	0
Number of special staff	3	1
Ability groups or tracking practices	6	2
Classroom atmosphere	1	0
Number of days of school	1	0

Source: Betty MacPhail-Wilcox and Richard A. King, "Production Functions Revisited in the Context of Educational Reform," *Journal of Education Finance* 12 (Fall 1986): 203-218, Tables 1-3.

Note: Facilities and fiscal characteristics from original table are not included here.

studies, including those with the most accurate measures of individual class sizes, merely strengthens the overall conclusion."

- According to Krueger (2002, 18): "In sum, all three of these alternatives to Hanushek's weighting scheme produce results that point in the opposite direction of his findings: all three find that smaller class sizes are positively related to performance, and that the pattern of results observed in the 59 studies is unlikely to have arisen by chance."[3]
- According to Robinson and Wittebols (1986, 197): "This research analysis dispels the idea of an 'optimum' class size covering all types of students, in all subject areas and at all grade levels. Students at different grade levels, in different subject areas, and at different levels of personal and academic development require different learning conditions in order for optimum gains in achievement to occur."
- According to MacPhail-Wilcox and King (1986, 220-222): "First, the characteristics of students...may contribute more to the learning process than any purchased resources. Second, teachers' socio-economic status, salary, experience, and verbal abilities are all related to pupils' achievement. Third, professional preparation of teachers is not consistently related to student achievement. Fourth, various indices show particularly strong relationship between student achievement and class size. Finally, levels of expenditures are closely related to student achievement."
- According to Hedges, Laine, and Greenwald (1994, 11): "Taken together, the effect size analyses suggest a pattern of substantially positive effects of global resource inputs (Per Pupil Expenditures) and for teacher experience. The effects of certain resource inputs (teacher salary, administrative inputs, and facilities) are typically positive, but not always. The typical effects of class size (expressed either as pupil/teacher ratio or teacher/pupil ratio) are decidedly mixed."

Each reader must evaluate these materials and statements based on the tables above and/or consult the original documents. The next section attempts to place these materials and conclusions into a larger context.

How Much of a Difference Does Class Size Make on Achievement?

In the previous section, the focus was on the statistical significance of the relationship between class size and achievement. The focus is now on the magnitude and nature of the relationship:

- What is the magnitude of the relationship—the rate of return—or what is commonly called effect size?
- What is the nature of the relationship—does the rate of return change?

These concepts are easily discerned when plotted. The slope of the line indicates the magnitude and the shape of the line indicates a change in the rate of return. There are two basic options for the shape of the line, linear or nonlinear. If linear, there is no change in the rate. If nonlinear, the shape either increases, decreases, or both increases and decreases.

DOI: 10.4148/0146-9282.1101

Table 5
Summary of the Production Function Coefficients Utilized in Hedges, Laine, and Greenwald (1994) Analysis

Input Variable	*Number of Estimates*	*Statistically Significant (%)*		*Statistically Insignificant (%)*		
		Positive (%)	*Negative (%)*	*Positive (%)*	*Negative (%)*	*Unknown (Number)*
Per Pupil Expenditure						
Hanushek	65	24	6	46	24	11
Reanalysis	55	24	5	45	25	
Combined significance	35	34	5	37	20	
Effect size estimation	38	27	3	53	18	
Teacher experience						
Hanushek	140	32	8	35	25	15
Reanalysis	131	30	5	40	25	
Combined significance	107	32	7	36	25	
Effect size estimation	57	26	4	46	25	
Teacher education						
Hanushek	113	11	7	41	42	113
Reanalysis	88	11	7	44	38	
Combined significance	68	12	7	51	29	
Effect size estimation	41	10	7	32	51	
Teacher salary						
Hanushek	69	24	9	36	31	24
Reanalysis		21	9	37	33	
Combined significance		23	12	42	23	
Effect size estimation		15	11	37	37	
Teacher-pupil ratio						
Hanushek	152	13	12	32	43	45
Reanalysis		10	13	38	38	
Combined significance		11	13	42	34	
Effect size estimation		9	10	30	51	

Source: Larry V. Hedges, Richard D. Laine, and Rob Greenwald, "Does Money Matter? A Meta-Analysis of Studies of the Effects of Differential School Inputs on Student Outcomes," *Educational Researcher* 23 (April 1994): 7, Table 1.

Note: Administrative inputs and facilities were included in the analysis of Hedges et al. (1994), but are not included here.

What Is Class Size?

There are two ways to measure the relationship between the number of pupils and the number of teachers: teacher/pupil ratio; and pupil/teacher ratio. Class size is considered the pupil/teacher ratio. The calculations result in different numerical ratios and have different policy implications. Simply put, school do not have the option of removing students from classroom to achieve a desirable class size, so the only option is to hire more teachers. Therefore, the teacher/pupil ratio is the appropriate policy measure of class size.

What Is Effect Size?

Effect size is the change in achievement measured in standard deviations. In general, effect size is reported under two circumstances. In controlled experiments, effect size is the difference of outcomes between the control and experimental groups measured in standard deviations. In econometric studies, effect size is usually the standard regression coefficient, or the rate of change in the outcome for one standard deviation change in the treatment.

Studies Estimating Magnitude and Shape of the Relationship Between Class Size and Achievement

Below, six studies, four using meta-analysis and two using a controlled experiment approach, are reviewed.

(1) Meta-analysis: Glass and Smith (1978). The research by Glass and Smith was influential in policymaking not because they concluded that class size made a difference in achievement but because they claimed that the influence became larger as classes got smaller. In essence, the effect size became larger as classes became smaller than about 15. To follow is a sampling of statements from other studies attesting to the influence of their proposition.

According to Hanushek (1998):

The design was heavily influenced by an earlier summary of research by Glass and Smith. That latter study combined the evidence from different experimental studies and suggested that student achievement was roughly consistent across class sizes until the class size got down to approximately 15-to-1. After 15-to-1, reductions in class size appeared to yield significant gains in student performance (p. 26).

Moreover, the original Glass and Smith (1978) analysis itself cast serious doubts on the potential for any improvement in student performance for this policy (p. 37).

According to Mosteller (1995, 115):

The Tennessee legislators and teachers were also aware of an investigation by Glass and colleagues which reviewed the vast literature on the effects of class size on learning using a special quantitative method called meta-analysis. The results of this investigation suggested that a class size of 15 or fewer would be needed to make a noticeable improvement in classroom performance. At the time of the Glass study, the effect of class size on performance was controversial because many studies in the literature differed in their outcomes.

The new methods used by Glass and his colleagues were not accepted by all professional groups. At the same time, there were ongoing discussions about the lesser cost and possibly equal effectiveness of placing paid teachers' aides in elementary classrooms. Because of the additional expense associated with a reduction in class size for early grades, members of the Tennessee legislature decided that any proposed innovation should be based on solid information and, therefore, authorized a four-year study of class size which would also examine the cost-effectiveness of teachers' aides. The legislature appropriated $3 million in the first year for a study of pupils in kindergarten and then appropriated similar amounts in subsequent years for the project, which carried the acronym STAR (for Student-Teacher Achievement Ratio).

According to Bohrenstedt and Stecher (2002, 22):

Among the most influential research was Glass and Smith's 1978 meta-analysis of 77 class size reduction studies, which concluded that "large [achievement] advantages [can be expected to occur] when class size is reduced below 20" (Glass and Smith, 1978, p. ii). In a 1982 follow-up report, Glass and associates reiterated the earlier findings and noted that of the more than 100 well-controlled comparisons, 81 percent favored smaller class sizes. They strongly suggested that class sizes needed to be reduced to fewer than 20 pupils for significant results to be observed (Glass et al., 1982).

(2) Meta-analysis: Phelps (2011). (See first article in this issue.) Phelps conducted a reanalysis of Glass and Smith and identified several flaws in assumptions and mathematics. He concluded that the data contained in the meta-analysis indicated a much different relationship between class size and achievement when the contrived methodology was removed. Specifically, Glass and Smith superimposed the squared term into the regression equation to obtain an artificial emphasis on class sizes below 15. Then, to correct for this imposition, they superimposed an entirely different equation on class sizes above 24. Plotting the data without the selection of a "preferred" regression equation,[4] the data showed a complex curve with high points at class sizes of 1, 33, and 64, inconsistent with the original conclusions. The reader is urged to review these findings.

Table 6
Median Regression Coefficients

Input Variable	*Number of Studies*	*Coefficient*
Pupil/teacher ratio		
All studies	45	0.0600
Achievement	22	0.0150
Teacher/pupil ratio		
All studies	24	-0.0010
Achievement	16	0.0176
Teacher education		
All studies	41	-.0200
Achievement	19	-.0300
Teacher experience		
All studies	57	.0700
Achievement	28	.0415
Teacher salary		
All studies	27	.0008
Achievement	12	-.0013
Per pupil expenditure		
All studies	38	.0014
Achievement	26	.0020

Source: Hedges et al. (1994, 11, Table 4).

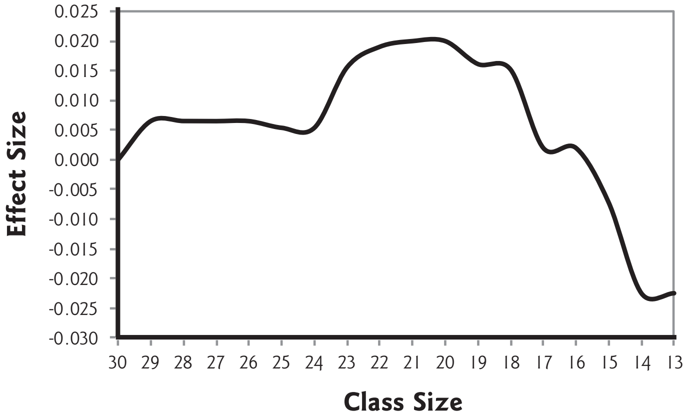
(3) Meta-analysis: Hedges, Laine, and Greenwald (1994). Hedges et al. estimated the relationship between several variables and student performance via standard regression coefficients: The amount of change in performance based on the change of an input. The study is a meta-analysis of other studies. Their motivation was to respond to the work of Hanushek (1989) and the implication that money does not matter. (See Table 6.)

Regarding the issue of class size, Hedges et al. (1994, 11) observed: "The typical effects of class size (expressed either as pupil/teacher ratio or teacher/pupil ratio) are decidedly mixed." This is consistent with the Hanushek analysis. Hedges et al. (1994, 11) included a per pupil expenditure variable (PPE) in their analysis and reached the following conclusion: "It [the result] suggests that an increase of PPE by $500 (approximately 10% of the national average) would be associated with a 0.7 standard deviation increase in student outcome."

(4) Meta-analysis: Addonizio and Phelps (2000). Addonizio and Phelps conducted a meta-analysis of four class size studies:

http://newprairiepress.org/edconsiderations/vol39/iss1/4
DOI: 10.4148/0146-9282.1101

Figure 1
Average Marginal Effect Size across All Subjects and Grades

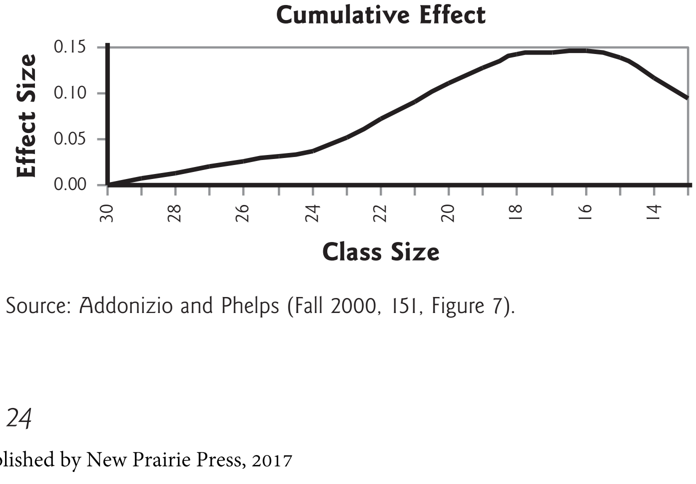


Source: Michael F. Addonizio and James L. Phelps, "Class Size and Student Performance, a Framework for Policy Analysis," *Journal of Education Finance* 26 (Fall 2000): 151, Figure 6.

Tennessee STAR, as reported by Mosteller (1995); Ferguson (1991), Ferguson and Ladd (1996), and Akerhielm (1995). The following is an excerpt from Addonizio and Phelps (2000, 150-154):

> The findings of four studies were summarized in a matrix with the individual outcomes from the studies as the rows, the class size intervals as the columns, and the marginal effects associated with class size changes as the cells. Of course, the cells contain the rates of change in the outcome only for the intervals of change reported in each study; therefore some cells are blank. The estimated effects can be plotted to indicate the general pattern of the effects on measured achievement over the entire range of class sizes. (See Figure 1.)
>
> Again, each cell in the matrix reports the marginal effect over the class size interval. In order to obtain an estimate of the cumulative effect across the range of intervals examined in each study, the average marginal rates of change for each interval are summed. (See Figure 2.)
>
> Finally, the functional relationship depicted in Figure 2 masks the substantial variation in findings across the studies.

Figure 2
Average Cumulative Effect Across Studies

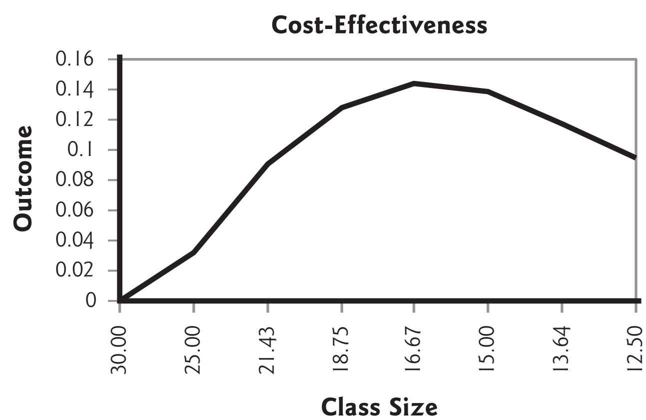


Source: Addonizio and Phelps (Fall 2000, 151, Figure 7).

(Quotation continued)

> These caveats raise questions regarding the appropriateness of combining the results as we have in an attempt to reach a general conclusion about the class size and student achievement relationship. With these caveats in mind, we find that achievement does rise as class size is reduced from about 30 to about 18.
>
> It is one thing to find a statistically significant relationship between class size and student achievement and quite another to determine that investment in smaller classes is a cost-effective strategy. This study has examined the estimated effect sizes of class size reductions from several published studies and will now derive a marginal cost function from these findings.
>
> The class size intervals—30, 29, etc.—provide the starting point for the cost-effectiveness analysis. Of course, the number of teachers necessary to reduce class sizes from 30 to 29 is not the same as reducing the class size from 29 to 28. Each successive incremental reduction in class size requires the hiring of an increasing number of teachers. For example, assuming 150 students in a grade, it would take 5 teachers to produce a class size of 30. By employing an additional teacher (making 6), the class size would then be 25, a reduction of 5. If a second teacher were added, the class size would then be 21.4, a marginal reduction of 3.6 students per classroom. Assuming a cost of $60,000 per teacher, we combine costs and estimated effects to derive a marginal cost curve for improving achievement through class size reductions.
>
> When the relationship between class size and outcomes is adjusted for this cost-effectiveness scale, the relationship looks like (Figure 3):

Figure 3
Cumulative Effect at Various Levels of Resources



Source: Addonizio and Phelps (Fall 2000, 153, Figure 8).

Published by New Prairie Press, 2017

(Quotation continued)

On the basis of our summary of the studies of the generalized relationship between class size and outcomes, the cost-effectiveness analysis indicates a modest gain in outcomes as class size is reduced from 30 about 16, after which the marginal gain falls off.

(5) Controlled Experiment: Mosteller (1995). In 1985, the state of Tennessee started a program to reduce class size in the early grades called STAR (Student Teacher Achievement Ratio). The controlled experiment was structured with two treatment groups and one control group. The control group was the regular-sized classes, and the treatment groups consisted of either smaller classes, or a regular-sized class with an aide. In both treatment groups, achievement was higher than the control group. (See Table 7.)

Mosteller (1995, 125-126) reached this conclusion:

Compelling evidence that smaller classes help, at least in early grades, and that the benefits derived from these smaller classes persist leaves open the possibility that additional or different educational devices could lead to still further gains. For example, applying to small classes the technique of within-class grouping in which the teacher handles each small group separately for short periods could strengthen the educational process (essentially a second-order use of small class size). The point is that small classes can be used jointly with other teaching techniques which may add further gains.

A follow-up study was conducted by Achilles et al. (1993) to assess the long-range benefits of the program. According to Mosteller (1995, 125):

In the Lasting Benefits Study,[5] a continuation of studies evaluated the performance of students from small classes as compared with the performance of students from regular-sized classes or regular-sized classes with an aide after all students had returned to regular-sized classes. The results always favored the students from smaller classes. One year later (1989–90), the effect sizes ranged from 0.11 to 0.16 (n = 4, 230) in the fourth grade, and then, in subsequent years, from 0.17 to 0.34 (n = 4, 639) in the fifth grade, from 0.14 to 0.26 (n = 4, 333) in the sixth grade, and from 0.08 to 0.16 (n = 4, 944) in the seventh grade... Thus, year after year, the students who were originally in smaller classes continued to perform better than the students from regular-sized classes with or without a teacher's aide.[6]

Interestingly, a summary of STAR results appears in *Capstone Report: What We Have Learned about Class Size Reduction in California* (Bohrenstedt and Stecher 2002), indicating the value they placed in the results in hope of a replication.[7]

Project STAR's major findings and those of other research to date include (Finn, 2002):

- Students in small classes performed better at all K–3 grade levels than did students in larger classes.
- Minority and inner city children gained more from reduced classes than their white and nonurban school peers; indeed, the effects were two to three times as great.
- Teacher morale was higher in smaller than in larger classes.
- Teachers spent more time on direct instruction and less on classroom management in smaller versus larger classes. Students in smaller classes were more engaged in learning than were students in large classes.
- The earlier and longer the participation in small classes, the greater the effect on achievement.
- Students in small K–3 classes did better academically in grades 4, 6, and 8 than did students in larger K–3 classes.
- The more years students spent in small K–3 classes, the longer-lasting the benefits in later years of schooling.
- Students who had been in small K–3 classes were more likely to graduate from high school, to take college admissions examinations, and, in general, to take courses that prepared them for college than were those who had been in larger K–3 classes. Furthermore, these effects were stronger for minority students, thereby helping close the college preparation gap between African American and white students.

Table 7
Tennessee Class Size Study Summary of Effect Sizes in First Grade

	SAT Reading	*BSF Reading*	*SAT Math*	*BSF Math*
Small class vs. regular-sized class without an aide	.30	.25	.32	.15
Regular-sized classes with an aide compared with regular-sized classes without and aide	.14	.08	.10	.05

Source: Jeremy D. Finn, and Charles M. Achilles, (1990), "Answers and Questions About Class Size: A Statewide Experiment," *American Educational Research Journal* 27 (3): 557-577, Table 5. In Frederick Mosteller, "The Tennessee Study of Class Size in the Early Grades," *Future of Children* 5 (Summer/Fall 1995): 121, Table 2.

Not everyone reached the same conclusions. Hanushek (1998) argued that the effects in the Tennessee STAR project occurred primarily in kindergarten and first grade and that there was no evidence that additional years of class size reduction contributed incrementally to the effect of small classes in the early years. He acknowledged that the effects were greater for minority and disadvantaged students but then argued, "...the effects appear small relative to costs of programs and alternative policy approaches" (p. 31).

In 1999, Hanushek also took issue with the methodology of the Tennessee STAR project, stating:

While random-assignment experiments have considerable conceptual appeal, the validity and reliability of results depends crucially on a number of design and implementation issues. This paper reviews the major experiment in class size reduction-Tennessee's Project STAR-and puts the results in the context of existing nonexperimental evidence about

http://newprairiepress.org/edconsiderations/vol39/iss1/4
DOI: 10.4148/0146-9282.1101

(Quotation continued)

> class size. The nonexperimental evidence uniformly indicates no consistent improvement in achievement with class size reductions. This evidence comes from very different sources and methodologies, making the consistency of results quite striking. The experimental evidence from the STAR experiment is typically cited as providing strong support of current policy proposals to reduce class size. Detailed review of the evidence, however uncovers a number of important design and implementation issues that suggest considerable uncertainty about the magnitude of any treatment effects. Moreover there is reason to believe that the commonly cited results are biased upwards. Ignoring consideration of the uncertainties and possible biases in the experiment, the results show effects that are limited to very large (and expensive) reductions in kindergarten or possibly first grade class sizes. 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 is found in the STAR results (p. 43).

Krueger (2000) countered Hanushek's cost-ineffectiveness argument by showing that there may be significant long-term learning differentials for Tennessee STAR students who were in small versus large classes given that they were more likely to take courses and entrance examinations that rendered them more college ready and, therefore, more job-prepared.

(6) <u>Controlled Experiment: Bohrenstedt and Stecher (1999; 2002)</u>. According to Bohrenstedt and Stecher (2002, 4):

> A task force assembled by the California Department of Education, called for among other reforms, smaller classes—a move strongly favored not only by the teachers' unions, but also by parents and teachers. California elementary schools had the largest class size in the country—averaging 29 students. Evidence from the Tennessee STAR experiment had shown rather clearly that elementary students in the primary grades did better academically when in small versus larger classes in K–3, and the difference was greatest for inner-city and minority students...A law was passed in July 1996. The law provided districts with $650 per student for each K–3 classroom with 20 or fewer students, providing they first reduced all first grade classes in a school, followed by all second grades and finally by either kindergarten or third grade classes. The cost to the state in the first year was roughly $1 billion dollars and in the current year, roughly $1.6 billion.

In the first report of the CSR Research Consortium (Bohrenstedt and Stecher 1999, 18), there were indications of achievement gain in the smaller classes: "The 'effect size' of the difference between students in smaller and larger classes was nearly 0.1 or one-tenth of a standard deviation. That is equivalent to a 2 to 3 point gain on average in the scale score on the Stanford Achievement Test." The major findings, taken in part from the final CSR report (Bohrenstedt and Stecher 2002, 5-8), are summarized as follows [italics in the original]:

1. *Implementation of CSR occurred rapidly, although it lagged in schools serving minority and low-income students...*
2. *Our analyses of the relationship of CSR to student achievement was inconclusive.* Student achievement has

(Quotation continued)

> been increasing since the first administration of the SAT-9 in 1997, but we could find only limited evidence linking these gains to CSR. We found a positive association in 1998 between third-grade class size and SAT-9 scores after controlling for differences in student and school characteristics. However, the size of this CSR effect was small. In the following year, 1998–99, these positive differences persisted when students who had been in reduced size third-grade classes moved to the fourth grade and regular size classes. The spring 1999 SAT-9 results showed that fourth-grade students who had been in reduced size third-grade classes scored higher than those who had not been in such classes. By 2001, CSR implementation was nearly complete, and as a result we could not examine differences in SAT-9 scores between students who were and were not in reduced size classes. Instead, we tracked achievement gains between cohorts of students with incrementally different patterns of CSR exposure to CSR from kindergarten through third grade.
>
> Although both overall exposure to CSR and statewide average test scores increased across cohorts, the magnitude of the changes in test scores did not track with the incremental changes in CSR. Thus, attribution of gains in scores to CSR is not warranted. More refined school-level analyses also failed to find meaningful differences in second- or third-grade scores of students with an additional year of CSR exposure in first grade compared to students who participated only in grades 2 and 3. We could not determine whether our ability to link CSR to achievement was due to weakness of the effect of incremental differences in CSR or to design limitations (or a combination of both). We were also limited in our ability to determine how much of the recent gain in achievement was attributable to CSR and how much was linked to other initiatives.

3. *CSR was associated with declines in teacher qualifications and a more inequitable distribution of credentialed teachers.* Reducing class size required an enormous increase in the number of K–3 teachers in California...To meet the increased demand for teachers, many districts hired teachers without full credentials...Most of the uncredentialed teachers were hired by schools serving the most disadvantaged students, in part because these schools were slower to implement CSR, and more certificated teachers had already been hired elsewhere. In 2000–01, more than one in five K–3 teachers were not fully credentialed in schools with high percentages of low-income, EL, minority, or Hispanic students (primarily large and urban).
4. *CSR had only a modest effect on teacher mobility.* One of the fears was that class-size reduction would result in two types of teacher mobility—teachers from urban schools moving into suburban schools and upper grade elementary teachers moving into K–3. While there was some initial increase, the effect was small and soon disappeared...

(Quotation continued)

5. *CSR implementation did not affect special education identification or placement...*
6. *Students in reduced size third-grade classes received more individual attention, but similar instruction and curriculum.* Compared to teachers with larger classes, teachers of reduced size classes were more likely to say they know what each student knows and can do, that they provide feedback on writing assignments within one day, that they give more individual attention to students, and are able to meet the instructional needs of all students. Teachers in reduced size classes also reported fewer behavior problems and reported that students were more likely to complete the lesson for the day and less likely to be "off task" for more than 5 minutes. But teachers in both reduced and non-reduced size third-grade classes reported spending similar amounts of time and covering similar amounts of curriculum in language arts and in mathematics.
7. *Parents liked reduced size classes.* Based on survey results, parents of third-grade students in reduced size classes rated selected features of their child's education higher than did parents of children in non-reduced size classes. The differences in rating of classroom size were particularly pronounced, with parents of children in reduced size classes reporting satisfaction levels far higher than parents of children in regular size classes. However, parents of children in both reduced and non-reduced size classes expressed equal satisfaction with the qualifications of their children's teachers.
8. *Classroom space and dollars were taken from other programs to support CSR.* Most districts in our statewide sample reported incurring operating costs for CSR that exceeded state payments for it, and these funding problems persisted, or even worsened, in recent years. Districts attempted to overcome budget shortfalls created by CSR by reducing funds for facility maintenance and administrative services. About one-third of such districts also reduced resources for professional development, computer programs, or libraries. To be able to implement the program, many schools reported having to reallocate full-sized classrooms that had been designated for special education back to K-3 classrooms, thereby forcing special education classes to use alternative spaces. CSR implementation also preempted space from such uses as music and arts, athletics, and childcare programs.
9. *In spite of budget shortfalls districts are not projecting CSR cutbacks for 2002–03...*Some [districts] did indicate, however, that cuts to the CSR program were a possibility and would continue to be discussed as their budgets were developed. However, it would be a "last resort" change given the popularity of CSR with parents and teachers.

Effect Size Estimates for Instructional Policy Options

There are few studies estimating the effect size for instructional policy options. Walberg (1984) compiled a comprehensive list of estimated effects in three categories: Student aptitudes; instructional quality and time; and home, peer, class morale, and media. (See Tables 8 and 9.) A class size effect was estimated at .09; however, no class size interval was provided to calculate a rate of change. Walberg (1984, 25) concluded: "Syntheses of educational and psychological research shows that improving the amount and quality of instruction can result in vastly more effective and efficient academic learning. Educators can do even more by also enlisting families as partners, and engaging them directly and indirectly in their efforts."[8]

Table 8
Instructional Quality and Time Effects on Learning

Method	*Effect Size*
Reinforcement	1.17
Acceleration	1.00
Reading training	0.97
Cues and feedback	0.97
Science mastery learning	0.81
Cooperative learning	0.76
Reading experiments	0.60
Personalized instruction	0.57
Adoptive instruction	0.45
Tutoring	0.40
Individualized science	0.35
Higher order questioning	0.34
Diagnostic prescriptive methods	0.33
Individualized instruction	0.32
Individualized mathematics	0.32
New science curricula	0.31
Teachers expectations	0.28
Computer assisted instruction	0.24
Sequenced lessons	0.24
Advance organizers	0.23
New mathematics curricula	0.18
Inquiry biology	0.16
Homogenous grouping	0.10
Class size	0.09
Programmed instruction	-0.03
Mainstreaming	-0.12
Instructional time	0.38

Source: Herbert J. Walberg, "Improving the Productivity of America's Schools," *Educational Leadership* 41 (May 1984): 24, Figure 3.

http://newprairiepress.org/edconsiderations/vol39/iss1/4
DOI: 10.4148/0146-9282.1101

Table 9
Home, Peer, Class Morale, and Media Effects

Method	*Effect Size*
Graded homework	0.79
Class morale	0.60
Home interventions	0.50
Home environment	0.37
Assigned homework	0.28
Socioeconomic status	0.25
Peer group	0.24
Television	-0.05

Source: Walberg (May 1984, 24, Figure 4).

Effect Size Based on Organizational Effectiveness

Levin (1997) made a case for improving achievement by increasing the effectiveness of school operations. He identified five areas for attention: (1) Commitment to a clear purpose with measurable outcomes; (2) incentives linked to the success of meeting the outcomes; (3) access to useful information for decision-making; (4) flexibility to meet changing conditions; and (5) use of productive technology. Accordingly, efforts towards effectiveness were more likely to improve achievement than increased resource allocations.

Phelps (2009) estimated the effect size of school effectiveness by inspecting the residuals of a production function. The research question was whether schools consistently performed better than their predicted achievement levels when controlled for socioeconomic status (SES), staffing quantity, staff qualifications, and instructional materials. The answer was yes. Over the four-year period, schools consistently either overperformed or underperformed on the achievement expectation. The effect size was measured in terms of the amount of statistical variance explained by averaging the residual. SES explained about 55%, and school and district effectiveness about 27%, supporting Levin's contention.

Other references to this general issue include: (1) In *Cost-Effectiveness and Educational Policy*, Levin and McEwan (2002) addressed many of these issues in great detail; (2) In *Measuring School Performance and Efficiency: Implications for Practice and Research*, Stiefel, Schwartz, Rubenstein, and Zabel (2005) addressed the issues of effectiveness; and (3) In *Making Schools Work: Improving Performance and Controlling Cost*, Hanushek et al. (1994) provided practical alternatives for school improvement.

Some Observations

These questions remain unanswered: (1) Is adding staff a good investment? (2)Will effective instructional and organizational policies produce better achievement results? (3)How should policymakers decide between adding staff or changing instructional and organizational policies?

Hedges et al. (1994, 11) made the following observation:

> It might seem odd that the effect of global resources inputs (PPE) are so clearly positive while the effects for the components are less consistently positive. However, this is not at all contradictory. This pattern of results is consistent with the idea that resources matter, but allocation of resources to a specific area (such as reducing class size or improving facilities) may not be helpful in all situations. That is, local circumstances may determine which resource inputs are most effective, and local authorities utilize discretion in wisely allocating global resources among the areas most in need.

Maybe Hedges et al. are correct: Local circumstances should determine the effective policy options, and uniform statewide or national policies are likely to be ineffective. This might explain why the beneficial effects of statewide policies are difficult to measure and why some schools tend to be associated with higher academic achievement and others are not, even when adjusted for SES and resources.

The Decision-Making Taxonomy

The natural sciences provide many examples where the identification of a unifying structure leads to a new paradigm--a new way to think about the subject, a new way to think about research, and a new way to think about decision-making. To name just a few: the Periodic Table in chemistry; DNA in biology and chemistry; and Gravity, Relativity, and Quantum Mechanics in physics. At the beginning of this article, a decision-making taxonomy was suggested with these underlying questions:

A. Does the research fit into a unifying structure where the evidence and conclusions can be compared and evaluated?

B. Does the research fit into a unifying structure valuable in a decision-making process?

Based on the review of research, below are some observations regarding the decision-making taxonomy.

(1) <u>Professional and public opinion regarding class size</u>. Professional and public opinion matter! The reader is encouraged to reread the Bohrenstedt and Stecher (2002) regarding public opinion. The public is willing to sacrifice other programs to keep lower class sizes in light of budgetary difficulties—even when smaller classes produced no apparent results and at substantial costs. Also reread the section giving credit to the research of Glass and Smith for investing in class size reduction. People believe lower class size works and tend to believe research supporting that position.

Teachers and parents of children in school clearly favor lower class size. Perhaps they see themselves as the beneficiaries of the policy. Legislatures, board members, administrators, and parents without children in school tend to be less enthusiastic, probably because they are more responsible for the funding of a class size policy. Public education is a political entity relying on public opinion. If the public opinion is not accurately informed and changed, moving away from lowering class size to other more cost-effective policies will indeed be difficult. In light of the evidence, a change in opinion is appropriate. A change in the heavy reliance on public opinion by decision makers might also be appropriate. The answers to the underlying questions: A=No; B=No.

(2) A critical analysis of educational research evidence regarding class size: What is statistically significant? Without doubt, the econometric research on class size is mixed. The many meta-analyses show a balance of positive and negative effect signs and a balance of significant and insignificant results. It seems as if the analysis is analogous to a partly filled glass of water: Some see it half-full, and some see it half-empty. Policymakers are in the same position regarding a class size decision; it comes down to personal and public preferences. While the econometric studies were valuable at one time, that time may have passed. More comprehensive research would be more valuable for decision-makers. The answers to the underlying questions: A=No; B=No.

(3) A critical analysis of educational research evidence regarding class size: What is the nature of the relationship? Glass and Smith (1978) contend class size makes a substantial difference in achievement, but only when the classes are smaller then about 15; there, achievement steadily increased as classes become smaller. Phelps, in the first article in this issue, refuted Glass and Smith by identifying shortcomings in their analytical method and by reanalyzing their data with less prejudiced means. The result of the reanalysis shows a pattern of increasing and decreasing benefits to scale, a confusing pattern difficult to interpret or defend. In another meta-analysis, Addonizio and Phelps (2011) found a diminishing returns point where further reductions in class size produced little or no additional gain. This finding was directly the opposite that of Glass and Smith.

There is no clear indication as to the nature of the impact of class size on achievement. In most cases, the assumption is that the relationship is constant—benefits continue for every reduction in class size. But maybe that assumption is incorrect. There are many illustrations where "some" is "good," but "more" either does not add any benefit or could cause harm. It is possible—indeed likely—there are circumstances where there is a benefit threshold, and it is prudent to move to other policy areas when the threshold is reached. The answers to the underlying questions: A=Maybe; B=Maybe.

(4) A critical analysis of educational research evidence regarding class size: What is the magnitude of the relationship? Hedges et al. (1994) found no consistent effect size associated with reducing class size, but found a positive and strong effect size with per pupil expenditures, citing the standard regression coefficients as evidence. Their conclusions were curious:

- The amount of money made a difference, but when spent in the most usual ways, it did not.
- The estimated improvement in achievement for an additional $500 was the same for all schools.
- The estimated improvement in achievement for an additional $500 was the same for every increment of $500, i.e., an increase of $1,500 would produce three time the results of $500.

Here is a thought experiment. Take a hypothetical classroom with 20 pupils and a teacher with a salary of $60,000. The teacher is given $500 per pupil (a total of $10,000) to improve achievement, as suggested by Hedges et al. However, the condition is that achievement must improve by .7 standard deviations or the teacher will forfeit $10,000 of their salary. To make the conditions fairer, the teacher selects his or her students, either high-achieving or average-achieving.[9] What are the chances of the teacher being successful? Would a reasonable teacher accept these conditions?

Hedges et al.'s conclusion regarding the achievement result of a $500 investment is a reasonable interpretation of the standard partial regression coefficient, but these findings are in conflict with the conclusion stated earlier: Benefits accrue based on individual school decisions. The implication of the Hedges et al.'s proposition is that all schools will get the same results with the same additional expenditures, but this is not the case. The regression line is not actually a line; it is a three-dimensional distribution with the average of the distribution being the regression line; [10] that is to say, at any expenditure level, half of the schools will do better than what the line predicts and, half will not do as well. To express it another way, some schools are more effective than others in how they spend money. Economists call this efficiency.[11] It stands to reason if the ineffective schools spend the new money in the old way, there is little chance the predicted achievement gain will be realized, but if they spend the new money in a more effective way, the gains could be larger. This scenario raises an unusual dilemma. What if the ineffective schools would have spent the previous money more effectively? Surely their achievement scores would be higher. With this interpretation of the regression statistics, the logical answer is not to spend more money but to spend the existing money more wisely. Hedges et al.'s own analysis demonstrated the areas where schools spend money with no achievement benefit--teacher education, teacher salary, and administrative inputs. A case could be made that additional money could be helpful in making the effective changes in the school instructional programs or in the operations of the organization. Economists call these "opportunity costs." As suggested by Levin (1997) and measured by Phelps (2009), these opportunities are likely to be substantially larger than what would accrue with more resources. The answers to the underlying questions: A=Likely; B=Likely.

There is another consideration in the Hedge et al.'s interpretation. It is unlikely that the top-performing schools will accrue the same benefit as the lowest-performing schools with the same dollar amount and the same degree of effectiveness—there is a performance ceiling effect. Because there is an upper limit to achievement tests, high performing schools have larger numbers of students near or at the test ceiling; they have no room to improve. Another example of a ceiling is teacher experience. The interpretation of standard partial regression coefficients is that for every additional year of experience achievement will increase by the same amount—only if the teachers do not exhibit the same behavior each year. Clearly experience matters because as new teachers gain experience they change their behavior, but after a period of time, say five years, the changes are minimal. There is a behavior ceiling unless there is a change in the operations of the school or the instructional program. It is doubtful whether a prudent teacher, knowing the other interpretations of the statistics, would accept the thought experiment challenge. The moral: Don't always bet on the standard partial regression statistics! The answers to the underlying questions: A=Likely; B=Likely.

(5) A critical analysis of educational research evidence regarding class size: What do controlled experiments say about the magnitude of the relationship? The analysis of the Tennessee controlled experiment found positive and substantial benefits with effect size around a standard deviation, or effect size, of .25 for the smaller classes and .09 for regular classes with an aide (Achilles et al., 1993). The results for mathematics were about .04 lower than for

http://newprairiepress.org/edconsiderations/vol39/iss1/4
DOI: 10.4148/0146-9282.1101

reading. On the other hand, the analysis of the California controlled experiment found no achievement gain attributable to the reduction in class size (Bohrenstedt and Stecher, 2002), although there was an effect size of about .10 reported in an early analysis (Bohrenstedt and Stecher, 1999). There were not enough instructional or organizational data collected to explain why the results might be different in these situations. Surely, the different results were not due to the difference in location or time period. There must have been different circumstances. Were there differences in the instructional programs or the operations of the organizations?

While the controlled experiments estimated effect size, it is not the same measure as reported in the econometric studies. The experiments reported the effect difference between treatment and control groups while the econometric studies reported an effect rate of change, or a change in achievement for a given change in class size. The answers to the underlying questions: A=Unclear; B=Unclear.

(6) A critical analysis of educational research evidence regarding class size: What is the cost-benefit relationship? There is no disputing the fact that lowering class size is costly. Most of the econometric analyses do not focus on this point. Levin (1997) and Phelps (2009) demonstrated the concepts, methods, and benefits of cost-effectiveness analysis. The answers to the underlying questions: A=Likely; B=Likely.

(7) A critical analysis of educational research evidence: What is the magnitude of the relationship between achievement and instructional policy options? Walberg (1984 suggested that instructional and time policies have a major influence on achievement.His estimates of effect size raised several puzzling questions:

- Because the effect size estimates were substantially larger than those of class size, why is there so much emphasis on lowering class size?
- If the instructional and time benefits were so large, why don't schools implement these policies?
- If schools implemented the instructional and time policies and they were of the suggested magnitude, why aren't the results apparent in the improvement of overall achievement in the U.S.?
- Is it possible the effect sizes were overestimated?

There is an underlying impression that each of the instructional and time policy options operate independently—substantial achievement gains will be realized with each action taken—because the policy options are unique and additive. That impression is most likely false. More likely, there is a commonality among these instructional policy options suggesting they work together rather than separately and, as a result, there is a ceiling to their overall contribution. Actually, this notion is inherent in the nature of achievement testing and in the regression formulation. There is a ceiling to achievement tests, the perfect score. No matter the effect sizes, they cannot add up to perfect scores for all students because the tests are made to identify differences among students. Without variance in the tests, they would serve no useful purpose. There is a test ceiling with built-in variance. Regarding regression, if the instructional and time policy variables are correlated, and they surely are, they share a common variance. As a result, as variables are added, their contribution to the total explanation is increasingly smaller—the basis of stepwise regression. The answers to the underlying questions: A=Likely; B=Likely.

(8) A critical analysis of educational research evidence: What is the magnitude of the relationship between achievement and organizational policy options? Levin (1997) suggested that effective operation of the school has more to do with improving achievement than the allocation of resources. Phelps (2009), following up on the Levin proposition, estimated the effect size of instructional and organization effectiveness to be substantially higher than that for the allocation of resources. Their work supports the idea that effective utilization of the resources is more important than the amount of the resources, counter to the Hedges et al. (1994) proposition. The implications are enormous. There are many ineffective schools due to their operations, not due to the level of resources or SES. Conversely, there are many effective schools due to their operations, not due to the level of resources or SES. This important conclusion is repeated: The effect size attributable to effectiveness is large, substantially larger than what can be attributed to class size or any other resource policy. In other words, the success of implementing any resource policy is more dependent on the level of effectiveness than the policy itself.

Is it possible to determine what effective schools are doing and provide the knowledge to the others? Unfortunately, there is little research as to the reasons for the effectiveness. However, it is possible to include the concept of effectiveness in the policy analysis process. The answers to the underlying questions: A=Likely; B=Likely.

(9) A decision-making process including: Establishing a set of clearly stated goals; identifying a set of possible policy options to achieve the goals; clearly stating the assumptions why each of the policy option would achieve the goals; and evaluating each of the policy options to select the best alternative. If the above statement reflects the highest category on the suggested decision-making taxonomy, then existing research is scant. Without a clear statement of the underlying assumptions regarding the potential benefits of the competing alternatives and a practical decision-making model, what remains are personal preferences. These preferences morph, as Hedges 1994) suggested, into local discretion. In many cases, this process clearly works, as measured by the results; but, in other cases, it clearly does not, and a closer look at the decision-making process seems warranted.

The difference between level one and level three of the decision-making taxonomy, and the reasons why level one is the most common, is captured in the following quote from Schrage (1991, 305):

> The advantage and perhaps the major motivation for using "seat-of-the-pants" decision making is that it obscures the assumptions made in arriving at a decision. If no one knows the assumptions upon which you based your decisions, then even though they may be uneasy with the decision they will have a difficult time criticizing your assumptions or decisions.

What is missing in the research review is an integrated and comprehensive paradigm capable of accommodating the seemingly unrelated research and dissimilar numerical estimates into a unified structure conducive to policy analysis and decision-making.

Kuhn (1970), author of *The Structure of Scientific Revolutions*, is noted for his thoughts regarding paradigms. He set two essential characteristics: The work was "sufficiently unprecedented," from competing modes of research, and "sufficiently open-ended with all sorts of problems to resolve" (p. 10). He continued to describe the characteristics as including theory, mathematical laws, applications, instrumentation, and rules for future research. Later, Kuhn (1970, 15) made an observation which appears to summarize the previously reviewed research:

> In the absence of a paradigm or some candidate for paradigm, all of the facts that could possibly pertain to the development of a given science are likely to seem equally relevant. As a result, early fact-gathering is a far more nearly random activity than the one that subsequent scientific development makes familiar. Furthermore, in the absence of a reason for seeking some particular form of more recondite information, early fact-gathering is usually restricted to the wealth of data that lie ready to hand.

The nine points identified above are a modest attempt at building a conceptual base for such a policy analysis paradigm. The following articles in this issue will combine the various estimates of effect sizes into a coherent structure (theory and laws); build a rationale (theory) and analytical method (laws) to accommodate the ceiling and effectiveness effects; and demonstrate an integrated and comprehensive policy analysis paradigm (instrumentation and application).

References

Achilles, C.M., B.A. Nye, J.B. Zaharias, and B.D. Fulton. "The Lasting Benefits Study (LBS) in Grades 4 and 5 (1990–1991): A Legacy from Tennessee's Four-year (K–3) Class-size Study (1985–1989)." Project STAR. Paper presented at the North Carolina Association for Research in Education, Greensboro, North Carolina, January 14, 1993.

Addonizio, Michael F., and James L. Phelps. "Class Size and Student Performance: A Framework for Policy Analysis." *Journal of Education Finance* 26 (Fall 2000): 135-156.

Akerhielm, Karen. "Does Class Size Matter?" *Economics of Education Review* 14 (June 1995): 229-241.

Barnett, W. Steven. "Benefits of Compensatory Preschool Education," *Journal of Human Resources* 27 (Spring 1992): 279-312.

Bloom, Benjamin S. "The Search for Methods of Groups Instruction and as Effective as One-to-One Tutoring." *Educational Leadership* 4 (May 1984): 4-17.

Bohrenstedt, George W., and Brian M. Stecher, eds. *Class Size Reduction in California: Early Evaluation Findings, 1996-98.* Sacramento, CA: CSR Research Consortium, California Department of Education, June 1999. http://www.classize.org/techreport/index.htm.

Bohrenstedt, George W., and Brian M. Stecher, eds. *Capstone Report: What We Have Learned about Class Size Reduction in California*. Sacramento, CA: CSR Research Consortium, California Department of Education, August 2002. http://www.classize.org/techreport/index-02.htm.

Ferguson, Ronald F. "Paying for Public Education: New Evidence on How and Why Money Matters." *Harvard Journal of Legislation* 28 (Summer 1991): 465–498.

Ferguson, Ronald F., and Helen F. Ladd. "How and Why Money Matters: An Analysis of Alabama Schools." In *Holding Schools Accountable: Performance-Based Reform in Education*, edited by Helen F. Ladd, 265-298. Washington, DC: The Brookings Institution, 1996.

Finn, Jeremy D. "Class-Size Reduction in Grades K-3." In *School Reform Proposals: The Research Evidence*, edited by Alex Molnar, 27-48. Greenwich, CT: Information Age Publishing, 2002.

Glass, Gene V., and Mary Lee Smith. *Meta-analysis of Research on the Relationship of Class-size and Achievement*. San Francisco, CA: Far West Laboratory for Educational Research and Development, 1978.

Hanushek, Eric A. "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 (Summer 1999): 143-163.

____________. "The Evidence on Class Size." Occasional paper 98-1. Rochester, NY: University of Rochester, Wallis Institute of Political Economy, February 1998. http://edpro.stanford.edu/Hanushek/files_det.asp?FileId=114.

____________. "The Impact of Differential Expenditures on School Performance." *Educational Researcher* 18 (May 1989): 45-65.

Hanushek, Eric, A., and Dongwook Kim. "Schooling, Labor Force Quality, and Economic Growth." NBER Working Paper No. 5399. Cambridge, MA: National Bureau of Economic Research, 1995.

Hanushek, Eric A., with others. *Making Schools Work: Improving Performance and Controlling Cost*. Washington, DC: The Brookings Institution, 1994.

Hedges, Larry V., Richard D. Laine, and Rob Greenwald. "Does Money Matter? A Meta-Analysis of Studies of the Effects of Differential School Inputs on Student Outcomes." *Educational Researcher* 23 (April 1994): 5-14.

Kuhn, Thomas S. *The Structure of Scientific Revolutions*. Chicago, IL: University of Chicago Press, 1970.

Krueger, Alan B. "Understanding the Magnitude and Effect of Class Size on Student Achievement." In *The Class Size Debate*, edited by Lawrence Mishel and Richard Rothstein, 7-35. Washington DC: Economic Policy Institute, 2002. http://edpro.stanford.edu/hanushek/admin/pages/files/uploads/classsizedebate.full%20volume.pdf.

Levin, Henry M. "Cost-Effectiveness and Educational Policy." *Educational Evaluation and Policy Analysis* 10 (Spring 1988): 51-61.

____________. "Raising School Productivity: An X-Efficiency Approach." *Economics of Education Review* 16 (June 1997): 303-311.

Levin, Henry M., and Patrick J. McEwan, eds. *Cost-Effectiveness and Educational Policy*. Larchmont, NY: Eye on Education, 2002.

MacPhail-Wilcox, Betty, and Richard A. King. "Production Functions Revisited in the Context of Educational Reform." *Journal of Education Finance* 12 (Fall 1986): 191-222.

Marzano, Robert J., Barbara B. Gaddy, and Ceri Dean. *What Works in Classroom Instruction*. Aurora, CO: Mid-Continent Research for Education and Learning (McREL), August 2000.

DOI: 10.4148/0146-9282.1101

Mosteller, Frederick. "The Tennessee Study of Class Size in the Early Grades." *Future of Children* 5 (Summer/Fall 1995): 113-127. http://www.princeton.edu/futureofchildren/publications/docs/05_02_08.pdf.

Nye, B.A., J.B. Zaharias, B.D. Fulton, et al. *The Lasting Benefits 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*. Seventh grade technical report. Nashville, TN: Center of Excellence for Research in Basic Skills, Tennessee State University, 1994. Cited in Frederick Mosteller, "The Tennessee Study of Class Size in the Early Grades." *Future of Children* 5 (Summer/Fall 1995): 113-127.

Phelps, James L. "Measuring and Reporting School and District Effectiveness." *Educational Considerations* 36 (Spring 2009): 40-52.

Robinson, Glen E., and J. H. Wittebols. *Class Size Research: A Related Cluster Analysis of Decision Making*. Arlington, VA: Educational Research Services, Inc., 1986.

Schrage, Linus E. *Lindo: An Optimization Modeling System*. San Francisco, CA: Scientific Press, 1991.

Stiefel, Leanna, Amy Ellen Schwartz, Ross Rubenstein, and Jeffrey Zabel, ed. *Measuring School Performance and Efficiency: Implications for Practice and Research*. Larchmont, NY: Eye on Education, 2005.

Walberg, Herbert J. "Improving the Productivity of America's Schools." *Educational Leadership* 41 (May 1984): 19-27.

Woessmann, Ludger, and Martin R. West. "Class-Size Effects in School Systems Around the World: Evidence from Between-Grade Variation in TIMSS." Cambridge, MA: Harvard University, Education Policy and Governance, 2002. http://www.eric.ed.gov/PDFS/ED467039.pdf.

Endnotes

[1] Wherever possible, the original material is presented so that the reader can judge the significance firsthand.

[2] For a complete list of Hanushek's publications, see http://edpro.stanford.edu/Hanushek/content.asp?contentId=81.

[3] Krueger (2002, 16) went on the estimate the amount of variance explained by class size to be 0.08.

[4] This analytical method was used by Addonizio and Phelps (2000) and is described later in this section.

[5] C.M. Achilles, B.A. Nye, J.B. Zaharias, and B.D. Fulton, "The Lasting Benefits Study (LBS) in Grades 4 and 5 (1990–1991): A Legacy from Tennessee's Four-Year (K–3) Class-Size Study (1985–1989), Project STAR, paper presented at the North Carolina Association for Research in Education. Greensboro, North Carolina, January 14, 1993.

[6] B.A. Nye, J.B. Zaharias, and B.D. Fulton, et al. *The Lasting Benefits 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*. Seventh grade technical report. Nashville, TN: Center of Excellence for Research in Basic Skills, Tennessee State University, 1994.

[7] Finn, a coauthor on Tennessee STAR project publications, served on the CSR advisory panel (http://www.classize.org/advpanel/index.htm), so it is reasonable to assume he participated in preparing this summary.

[8] Following are some other studies regarding instructional effect sizes: (1) In *What Works in Classroom Instruction*, Marzano, Gaddy, and Dean (2000) provided effect size estimates similar to those of Walberg (1984), but provided a full description of the instructional conditions; (2) In "The Search for Methods of Groups Instruction and as Effective as One-to-One Tutoring," Bloom (1984) provided the effect sizes for instructional methods of mastery learning and tutorial instruction all with a consistent class size of 1 to 30; (3) In "Benefits of Compensatory Preschool Education," Barnett (1992, 297) estimated the effect size of preschool programs at .75; and (4) In *Capstone Report: What We Have Learned about Class Size Reduction in California*, Bohrenstedt, George W., and Brian M. Stecher (2002) included references to instructional policy options other than class size reduction.

[9] Starting at the average, the 50th percentile, a .7 improvement would raise the standing to the 75th; starting at the 75th, the improvement would be to the 95th; starting at the 95th, the improvement would be to the 99th. As the starting point gets higher, the percentile gains gets smaller.

[10] The standard error of estimate is the parameter of the three-dimensional distribution.

[11] The efficiency portion of the residual is separated from the random error portion by averaging over time the residual for each observation. In econometrics, this is known as the fixed effect.