

Perspectives on Class Size Reduction

Elizabeth Graue, Denise Oen, Kelly Hatch, Kalpana Rao, Erica Fadali
University of Wisconsin, Madison

A paper presented at the symposium *Early Childhood Policy in Practice: The Case of Class Size Reduction* at the annual meeting of the American Educational Research Association, Tuesday, April 12, 2005, Montreal, Canada.

Primary contact information: Elizabeth Graue, Professor, Wisconsin Center for Education Research, University of Wisconsin Madison, 1025 W. Johnson, Madison, WI 53706, 608 262-7435, graue@education.wisc.edu.

Support for this paper was provided by the Wisconsin Department of Public Instruction.

Gov. Jeb Bush on Monday abandoned his effort to repeal Florida's popular class size amendment and unveiled a scaled-back plan that also increases teacher pay.

Conceding that repealing the amendment would be difficult, Bush asked lawmakers to put a measure before voters that would essentially freeze the caps on class sizes to the current standard. That would give school districts more flexibility, he said. (James & Johnson, 2005)

The governor's proposal to increase state funding for public education wouldn't solve all of the Madison School District's budget woes, but it would provide significant help in dealing with an estimated shortfall of \$6 million to \$8 million next year, Superintendent Art Rainwater said.

The district would gain an added \$1.3 million for the Student Achievement Guarantee in Education program, which limits class sizes in kindergarten through third grade to an average of 15 students.

Doyle wants to raise the reimbursement for low-income SAGE students from \$2,000 to \$2,500 per pupil. Madison has 2,656 low-income SAGE students and a total 5,620 students in the SAGE program. (Cullen, 2005)

Class size reduction (CSR) is one of the most frequently researched and policy implemented reforms in recent educational history. It has been studied by a wide range of researchers representing interests ranging from economics to administration and has focused on short term and long term outcomes for students, costs relative to other reforms, teacher attitudes and practices. In the last 10 years, 40 states have implemented CSR programs, the federal government had a short term program that infused teachers into the employment ranks to reduce class size, and a number of reviews have suggested that children learn more in smaller classes, their teachers are more satisfied, and discipline problems are reduced. Anchored by the only large scale randomized experiment in fieldbased educational research (Jeremy D. Finn & Achilles, 1990), the knowledge base on class size reduction is voluminous, highly varied in quality and foci.

In our work analyzing the literature on class size reduction we found that a number of substantive reviews had already been completed on the subject. One might wonder why researchers would undertake another review. In fact, we were told just that by Alex Molnar who strongly suggested we were just reinventing the wheel – that we knew what we needed to know ten years ago.

Despite this advice, we can think of several reasons to revisit this literature. The first is a pragmatic one in relation to policy. Given the difficult economic context that has many states and local governments cutting programs and reallocating the funding to other initiatives, it is important to assess the knowledge base about the efficacy of class size reduction so that good decisions can be made. The second reason might be seen as more rhetorical, or political. We consider a review to be much more and much less than attempting to find out what the literature “says.” The literature has no voice – it is researchers who take a position and speak about topics in their area of interest. We follow the idea, first suggested by Grant & Graue (Grant & Graue, 1999) that reviews represent conversations in the educational research community – as such they are sites of interaction that often present multiple perspectives on a topic. To reflect this approach, we structure our review in a slightly different way than most reviews. We take an interpretive approach to research syntheses, exploring how different researchers have taken up a topic, operationalized it in research methods, and argued for a particular understanding of the outcomes of the study. From this perspective, a research synthesis is an analysis of the discourse of the research community on a topic. We add to this a layer of inquiry that is not typically part of a review. We present researchers’ perspectives on our knowledge base through analysis of interviews with a sample of researchers who have worked on the topic of class size reduction.

This exploration of *understandings of understandings* is a methodological instantiation of the interpretive status we use here.

Our review is organized around two questions:

- How has *class size reduction* been conceptualized in the literature?
- What methods and measures have shaped our knowledge?

Methods

We should be clear about the catalyst for this review. As part of an evaluation of the Student Achievement Guarantee in Education, we recognized that having a firm grasp of the literature would be important. Given the number of reviews that had been done previously and the mixed messages coming out of those reviews, we decided to take a mixed method approach to exploring the literature. We began in a traditional way, doing a systematic search of the literature. But one of the first that we found is that what is “there” depends on “where” you look. With the changes in ERIC it is rarely the case that one can find a single database that covers the range of journals of interest in education. Therefore, multiple searches across multiple databases were required. We did searches of the major electronic databases (Education Full text, ERIC, EBSCO, Ingenta, JSTOR and Proquest) using the descriptor class size reduction. Initially we searched for any types of written text (journals, reports, popular press) and found close to 400 pieces. We narrowed our search to peer reviewed published works to make the task more manageable. We then began working from bibliography to bibliography, searching out published work that our initial searches had missed.

We worked to describe the literature by focusing on 4 general categories of the literature. We summarized the research related to the three major policy initiatives for class size reduction—STAR in Tennessee, SAGE in Wisconsin, and the California class size reduction

program. We described the parameters of the policy, the methods used to evaluate the programs and the suggested results. In addition we analyzed the synthetic work on class size reduction, reading and analyzing reviews on the topic. Finally, we identified key researchers who have contributed to the scholarship on class size and interviewed individuals who represented varied perspectives and expertise. Our interviews included scholars instrumental in evaluation of the major class size initiatives: Charles Achilles (STAR), Alex Molnar (SAGE), Joan McRobbie (California). In addition, we talked with people who had synthesized the CSR literature: Bruce Biddle, David Grissmer, and Adam Gamoran. We interviewed an economist who analyzes the cost of class size relative to other investments in education – Eric Hanushek and a researcher who did secondary and follow on analysis of SAGE data – Norman Webb. We recognize that our interview participants represent a sample of the research on CSR. It is important to say that some researchers declined participation. Therefore, silences come for a variety of reasons and we are comforted in the idea that we do not have to talk with everyone to have some sense of the conversations related to CSR.

Major class size initiatives

STAR

Tennessee's Student Teacher Achievement Ratio project began as a pilot in 1984 and was mandated in 1985. A large scale, randomized within-school experiment (Nye, Hedges, & Konstantopoulos, 2001), pupils entering kindergarten in participating schools were randomly assigned to a small class (13-17), a full size class (22-25) or a full size class with a full time teacher aide. Teachers were also assigned at random to the classrooms. Pupils were kept in the same condition for up to 4 years, with a new teacher assigned at random each year (Finn, 2002). The four year, \$13 million project was completed in 1990.

Students in small classes performed better on all achievement measures, and the positive benefits of smaller classes were cumulative. In kindergarten, slightly more than half of the small classes were in the top scoring 10% of STAR classes. By third grade, that proportion had grown to 78% (Bain et al, 1992). Small class advantages were greater children in poverty and of color. The gap between black and white students were reduced in the smaller classes, with African American students in smaller classes 4%ile points below white students but those in large classes were behind by 14%ile points. While there was no evidence that smaller classes improved student motivation, small classes had fewer discipline problems, more questions and greater participation and concentration. Teacher quality, including practices developed through professional development, had little effect on student outcomes. The addition of teacher aides to regular sized classes did not improve achievement (Achilles, 1997, Biddle et al., 2002, Finn, 1998,2002). Charles Achilles explains why:

Teacher aides are not a help in the classroom at early grades except as required by legislation and/or to help teachers out with materials and things. They shouldn't be around kids unless they are well-trained and then there is some question because the teacher sends troubled kids to the teacher aide. (Achilles interview)

While the STAR project appeared to be very positive, there were several criticisms of the study. The primary concerns about the project were that the student sample wasn't representative of the U.S. population (few Hispanic, Native American and Non-English speaking students) and that the study suffered from the Hawthorne effect when news about the greater achievement of small classes leaked out during the project. Critics argued that it was likely that this may have led participating teachers to work harder to get positive results (Gilman & Kiger, 2003). Another critique of the project was that students were randomly assigned not randomly selected and that the assignments changed over time. Only volunteers were part of the random assignment pool and schools had to be large enough to accommodate at least three classes in

each grade. Hanushek (1999) argued that the results of the study depended fundamentally on the choice of teachers. There was little description of how teachers were randomly assigned to treatment groups. The huge differences generally found among teachers could dramatically influence the results. Because of the ongoing need to assign new teachers to the various treatment groups across the 4 years, it seems entirely plausible (in Hanushek's estimation) that elements of teacher and principal preferences for different classes entered into the process. He also criticized the original researchers for their lack of consideration of attrition of the treatment groups. Of the initial experimental group starting in kindergarten, only 48% remained in the experiment for the entire 4 years. In our interview, Eric Hanushek described the kind of study that should be done to alleviate concerns:

A more careful description of the randomization process and how they randomized teachers and students. More careful assessment of achievement differences before and after reduced class size; allowing for some variation in the number of years that people are in different treatment categories; trying to assess better any variation in teacher quality that lie behind this; trying to follow the attrition from the program better; trying to maintain the study design that keeps people in their particular treatment groups. (Hanushek interview)

Bruce Biddle recognized the weaknesses in the study but suggests that in the context of a practice based field like education, STAR might be a good enough study to inform policy:

The issue is that within a field setting, with human beings, this one was a pretty good piece of research. It was done on a state-wide basis. It involved matched classrooms within schools, which were subjected to three different kinds of treatments, standard-sized classrooms, classrooms that were standard sized where a teacher was supplemented in her or her efforts by a teacher's aide and then reduced size classes. These were done in the early grades. It was begun in kindergarten and in the best cases the kids were continued in these various treatment conditions for four years, through first, second, and third grade. Thereafter they were returned to standard classrooms in their schools. (Biddle interview)

A follow on study, the Lasting Benefits study, conducted from 1989-1996 tracked achievement effects as students progressed into grades higher than those observed by the original

study (Nye et al., 1989, 1994-1999). This study yielded two primary findings: First, students in smaller class performed better in all academic areas studied. Second, students from the small classes expended more effort in the classroom, took greater initiative in learning activities, and displayed disruptive or inattentive behavior (U.S. Department of Education, 1998).

SAGE

We believe that state and local policies can and should ensure that teachers have the time and resources to get to know their students and parents and to give each child sufficient attention, ensuring academic achievement at the highest possible level. We believe that it is important to fully prepare children for the diverse civic and employment environment where they will live and work. And we believe it is time to free educators to educate, holding both teachers and *all other people* who hold a stake in the schools-administrators, board members, parents, and students-accountable for the results. (Molnar & Zmrazek, 1994)

The *Student Achievement Guarantee in Education*, SAGE, much like other Class Size Reduction reforms, was envisioned as a way to narrow the achievement gap for children living in poverty.

Class size reduction was only one aspect of a multifaceted program that also includes a family/school component, curriculum reform, and professional development. How these components are studied provides insight into how SAGE is being conceptualized and implemented.

Student outcomes were a central indicator of program efficacy. Students in SAGE classrooms outperformed the comparison group in reading, language arts and mathematics through third grade, with the most pronounced effects found for African-American students (P. Smith, Molnar, & Zahorik, 2003). Webb and colleagues reanalyzed the initial evaluation data and found similar patterns of advantage in the original data which did not sustain in other data sources in third and fourth grade (Webb, Meyer, Gamoran, & Fu, 2004). Hruz questioned whether the size of the effects were in fact worth the cost of the program and suggested that

implementation be limited to the subgroups that had proven effects (K and 1st grade and African American students) (Hruz & Wisconsin Policy Research Institute., 2000).

Bruce Biddle contrasted the design of SAGE with that of STAR:

It involved not a clean experimental design but rather, a massive intervention study in which class size reduction was a part of what was initially to be a multi-faceted intervention study but the other facets of the intervention essentially xxx out and much of the efforts are clearly associated with class size. This was a very good piece of research. . . . There were more controls built into it. It was focused specifically on schools where a lot of minority kids were present. (Biddle interview)

One advantage of the SAGE evaluation was its attempt to represent the complexity of the reform, with attention to teaching practices and professional development opportunities.

Through observations in math and reading instruction, Zahorik et al found that SAGE classrooms had fewer discipline problems, more time spent on instruction and more individualized, explicit instruction, the teachers had more knowledge of students and more satisfaction with teaching (John Zahorik, Halbach, Ehrle, & Molnar, 2003). Despite these enumerated changes, researchers suggested that teacher practice does not change in smaller classes (J. Zahorik, Molnar, & Smith, 2003). Examinations of the professional development components of SAGE found that they were only partially met (Molnar, 2002). The authors also explored how SAGE professional development guidelines aligned with both the Wisconsin Teacher Professional Standards (which went into effect July 1, 2004) as well as the mandates of No Child Left Behind. It is their belief that current professional development practices will not fully comply.

Kiger (2002) argued that smaller classes create a “coherence” of the all four SAGE components and that all components have a synergistic effect on one another. He goes on to suggest that research needs to explore why smaller class sizes seem to have a positive effect.

Missing from the research on SAGE is attention to English Language Learners, to administrative practices at the school and district level, to the effects the programs had on home-

school connections, to the specific implementation of the pupil teacher ratio in its varied forms, and systemic analysis of the politics and interconnections that facilitate or inhibit the successful implementation of this type of complex reform. While the designs of the SAGE evaluations did mirror the components that served to structure the policy, most attention was focused on the student outcomes, with the structural and process analyses examined at a diffuse level of many sites rather than at a level that provided in-depth knowledge. This attention to knowledge about many is important for population generalization but may be less informative for informing particular practice.

California

CSR in California is conceived in the literature as something that was born at a good time, with good intentions, but poor planning. Its enactment was motivated by a ten-year decline in achievement in California, tying for last place on NAEP in 1994. A snapshot of California classrooms at the time would have shown overcrowded classes, a shortage of qualified teachers, and minority students, ELLs and poor students falling more and more behind (Stecher, Bohrnstedt, Kirst, McRobbie & Williams, 2001). With a budget surplus and the impressive claims of the STAR study, the California government dove head first into CSR as a statewide program. This approach had its downside, outlined by David Grissmer:

We have the California situation – California has kind of taken a bum rap on some of this stuff especially when they had a huge budget surplus. They didn't have an option of let's do a little this year, a little this year, a little next year, that sort of not the way politics works. What this meant was that implementation was going to be swampy for a while. They are not going to have the supply of teachers they need, all of the classroom size is not going to be available, and so forth. They reduced from 28 to 20. they did it on all socioeconomic levels, there was no sort of distinction and what they ran into was that it created a lot of teaching positions and hiring in schools and inner city teachers moved to those. So it got implemented fairly shabbily in the beginning but in my sense it was still worthwhile. (Grissmer interview)

Researchers emphasize that CSR is only one part of a comprehensive plan to improve learning (McRobbie, 1996). It is essential to also change teaching and learning behaviors (McRobbie). This means providing teacher support and professional development—about individualized instruction, better use of methods and materials, improved organization, a better variety of activities, better assessment and a stronger curriculum (McRobbie). Researchers also point out that effective, specific strategies for teaching smaller classes have yet to be identified (McRobbie, year?). Moreover, teacher quality matters, because they feel high quality instruction is essential for CSR to succeed.

In California, the gains earned from CSR do not necessarily outweigh the costs of the policy, according to researchers. To be sure, there were gains: a slight increase in test scores after the second and third years, more time teaching and less disciplining, and more parent-teacher contact time (reported) (Stecher, et al., 2001). However, the costs of the CSR policy include, most prominently, that educational inequities were exacerbated, particularly in districts already considered disadvantaged. There was a 38% increase in the teaching force, accompanied by a serious decline in the qualifications of these teachers particularly in the highest poverty districts and schools (McRobbie, Finn & Harman, 1998; Stecher, et al.). It is also more expensive to implement the policy in these districts. Space for new classrooms was taken from special education, childcare, specials and library/computer facilities (McRobbie, Finn & Harman; Stecher, et al.). And, many teachers of English Language Learners (ELLs) and special education students switched to regular education to help fill the gaps (Stecher, et al.; Wexler, Izu, Carlos, Fuller, Hayward & Kirst, 1998).

From the California implementation researchers suggest that investments be made in ongoing professional development with a focus on literacy (McRobbie, year?; McRobbie, 1996)

and on instructional strategies for a diverse student population, i.e. language acquisition and communication strategies across languages and cultures. Further, a focus on prevention would help teachers identify and respond to student learning issues. Finally a systemic induction model would support new teachers, especially emergency credentialed teachers added to staff CSR classrooms (McRobbie, 1996). Good CSR practice broadens notions of defining policy success. While including measures of student achievement it also involves using resources creatively and establishing new district and community collaboratives to reach goals for student achievement (McRobbie, 1996; McRobbie, Finn & Harman, 1998).

California CSR researchers are skeptical supporters. They remind consumers of the literature that it is unclear how CSR affects student achievement, and that we know that CSR effects are coupled with teacher quality, instructional strategies, curriculum, student-teacher interactions, family background and economic situation (of students). Their discussions tend to focus on implementation critique. Given the differences between STAR and the California policy, they suggest that the Tennessee study was not a good model for California to try to duplicate. STAR was a limited controlled study and California implemented a statewide program for 1.8 million children. Availability of qualified teachers was not considered, the baseline class size in California was much larger than in Tennessee, and California's curriculum standards were not as fully developed. One of the most glaring differences, though, is the diversity in California, especially concerning English Language Learners (McRobbie, Finn & Harman, 1998; Stecher, et al., 2001). Joan McRobbie that not all of California's children gained:

It's fair to say that probably some kids have lost because the way the policy was setup, certain kids who all along have been more likely to get the least qualified, least experienced teachers, by virtue of class size reduction, are even more likely to get the least qualified, least experienced teachers because of the shortage you've created and the number of people brought in. . . At least initially, it probably made for a more negative situation for some kids (McRobbie interview).

The California experience suggests some guidelines for large scale implementation of CSR. Focus on equity. Develop policies that fully recognize the complexity of practice. Improve teaching strategies. Plan for sufficient staffing. Implementations that ignore these elements compromise the potential that CSR can have for student achievement. (McRobbie, year?; McRobbie, Finn & Harman, 1998).

British Implementation of CSR

In the 1990's the British government claimed that educational attainment was a function of teacher quality rather than class size. Teachers and the public resisted this conclusion, and through meetings and talks, it became clear that "a proper and thorough study of educational consequences of class size differences" (p. 4) was needed. It was in this context that Peter Blatchford aimed to bridge the gap between professional experience and research-based evidence. Rather than trying to settle the debate, he focused on 'why' and "provide a sustained inquiry" (Blatchford, 2003, p.1).

Blatchford believes that teachers interact with students in context, so that CSR is only one influence on the classroom experience. Moreover, the nested contexts of schools means that different sizes of classes can mean different dynamics and these might influence teachers and students. His research, then, aimed to 1) determine if CSR affects students' achievement; and 2) study the relationships between class size and classroom processes, which might illustrate any differences found in achievement. He constructed a three year longitudinal study that followed students from school entry, i.e. 4 years. . Both quantitative and qualitative methods were used including case studies, teachers reports, observations, teacher ratings of behavior, and teachers' estimates of time use.

In Reception Year (kinder), all ability groups had positive outcomes related to smaller classes, with larger effects in literacy for students with lower baseline. Long terms effects were maintained when students moved to small or similar sized classrooms after the first year but they were lost if students went from from a small class to a large class.

In larger classes, students were more likely to be off task, distracted, and to spend more time with their peers. In smaller classes, students had more interaction with their teachers—child initiations and answering the teacher. With the exception of Reception year students, peer relations were often *worse* in smaller classes because of aggression and rejection.. As class size increases, teachers perceive less time for individual reading support and rarely mention guided reading groups as an instructional option. They felt that the quality of interactions was better in smaller classes; with teachers knowing and supporting their students better. Teachers used more procedural talk in larger classes, and more social or personal interactions to manage the group in smaller classes.

In first and second year classes students read more frequently and for longer periods. There was less differentiated instruction with larger classes and questioning was used for controlling rather than to teach. The feedback in smaller classes was more immediate, but such feedback takes time from lessons and disrupts the flow of activities and instruction.

It was harder to extend reading strategies with larger classes and there were fewer individual reading opportunities, monitoring and checking for understanding. More time was used for teaching in smaller classes including more one-to-one teaching, more teaching in groups, more times when students were focus of teachers' attention, more student-teacher interactions, and more participatory interactions between students and teachers.

Syntheses

Focusing on class size alone is like trying to determine the optimal amount of butter in a recipe without knowing the nature of the other ingredients. (Berger, 1982)

The knowledge base on class size reduction is framed in relation to a pair of papers that presented a relatively new type of analysis of relationships among class size, student achievement, attitudes, and instruction. Utilizing meta-analysis, Glass & Smith and Smith & Glass used the power generated by aggregating results across studies to refine our understanding of the knowledge base. Based on 59 studies yielding hundreds of effect measures, the authors provided a large scale overview of how class size shapes educational outcomes. With studies that ranged across academic content, levels of schooling, everything from tutoring to large lectures, the meta-analyses indicated a generally positive effect for student achievement with larger effects in studies of secondary programs and in more well designed inquiry. The effects ranged from 30 percentile ranks when comparing a size of 40 and a size of 1 to 10 percentile ranks between a class of 10 and 20 (Glass & Smith, 1979). The effects for attitudes and instructional practices were even larger, with an average effect size of .49. To make this more concrete, Smith and Glass note that if we think of a class size of 30 as the 50th percentile, then reducing the class size to 15 would shift the percentile to 63rd and increasing it to 40 would make it drop to 45. Larger effects were found for teachers and for younger students (M. L. Smith & Glass, 1980). In contrast to the achievement analysis, less well designed studies had larger outcomes than more well designed work. The positive outlook presented on reducing class size in these studies was a major building block in the later large scale implementation. Several key issues could be noted however. The later work on class size reduction has focused almost entirely on the early primary grades, while the Glass and Smith analyses explore work through the secondary school.

Following the Glass & Smith meta-analyses, synthetic reviews of class size reduction or class size effects became more pervasive but also less inclusive. Rather than sampling the universe of studies on the topic, researchers made choices that narrowed the field, which in turn narrowed the dimensions of the problem considered. Hedges and Stock reanalyzed a subset of the Smith and Glass data using an alternate method of meta-analysis and found smaller, but still sustainable effects for class size reduction (Hedges & Stock, 1983). Concerned about the adequacy of study designs, Slavin completed a *best evidence synthesis* comprised of studies that met particular methodological requirements. With this constraint the number of studies considered in the original meta-analysis narrowed from 59 to 11 (Slavin, 1989). On the other hand, Robinson completed a related cluster analysis that grouped studies by like findings and similar characteristics. For example, Robinson did analysis of the studies by grade level, typically considering divisions such as K-3, 4-8, and 9-12. He did not however, consider the nature of the evidence provided (Robinson, 1990). Syntheses in the 1990's and later focused on each other and on the major class size reduction initiatives approaching the literature from a variety of perspectives. In the next section we present our readings of the major themes in this literature:

Class size reduction vs pupil teacher ratio

One of the major issues in this literature has been definitional – what exactly is class size reduction? Is it when you reduce the number of students in a particular classroom or is it when you reduce the number of students per teacher? While these two are related, they are not the same (Addonizio & Phelps, 2000; Biddle & Berliner, 2002; Ehrenberg, Brewer, Gamoran, & Willms, 2001; Jeremy D. Finn, 2002; J.D. Finn, Pannozzo, & Achilles, 2003; Odden, 1990). For example, many analyses of class size take the number of students related to the number of staff

to argue that class sizes have shrunk over the years (E. Hanushek & Rivkin, 1997). But that estimate often includes additional support staff who do not systematically provide instruction to most students. Further, because these analyses rely on data collected for other purposes, the estimates include effects not related to class size such as student ability, with lower ability students overrepresented in smaller classes (Addonizio & Phelps, 2000). Adam Gamoran made this point clearly in his interview:

Many of the correlational studies have been of pupil/teacher ratios rather than class sizes. Those are two very different things. For example, they basically mix in funds for special ed with class size so that a school with a lot of special ed teachers would appear to have a smaller expenditure. But while those special ed teachers are probably going to be good for the special ed kids, they probably are not going to raise average test scores in the school by any substantial amount. So that's a methodological problem in much of the correlational work. (Gamoran interview)

In contrast some implementations of CSR increase both group size and the number of teachers, reducing the pupil-teacher ratio. This is the case in many of the classrooms participating in the SAGE program (Grissmer, 1999). Charles Achilles points out that this is an administrative solution to an instructional problem:

Pupil/ teacher ratio is to assure equitable distribution of funds. To make sure that every building and every youngster, whatever the state formula is, gets his or her share of the money. It's not an organization for instruction. It's an administrative device to trace money and to deal with desegregation issues and things like that. This is the thing that amazes me, then to have people assume that those two are the same and then using them as the same. (Achilles interview)

One reason that the patterns in the literature appear to be unclear is that the treatment in question is muddled. What assumptions shape advocacy of PTR or group size? The focus on pupil teacher ratio has embedded in it a number of assumptions. The first is that quality in education should be tied to investment and when the number of professional staff goes up so should the quality. Hanushek is a primary advocate of that argument and he uses historical analysis of the staffing patterns to argue that despite significant influx of new staff in US

schools, student achievement has remained flat or decreased over time. Ehrenberg et al (2001) point out the problems with this approach, noting that indicators like drop out rates should be considered in school performance; that increasing numbers of students are coming to school with risk factors that compromise their potential achievement; teaching is attracting less able candidates; and changes in special education rules has increased the number of staff supporting students with disabilities. Further, Jeremy Finn disagrees with this rationale, arguing that:

Because pupil teacher ratios are usually computed for large, heterogeneous units (i.e., school districts, states, or countries), it is little surprise that they have a weak relationship with academic achievement. These levels of analysis may be appropriate for an economists work, but they are not useful for educators concerned with teaching and learning in individual classrooms. (Jeremy D. Finn, 2002)

The second assumption for using PTR is related to the mechanisms that change student achievement – from this perspective, researchers assume that reducing the ratio of students to teachers provides more access to teacher expertise, it provides more hands and eyes to support student behavior. Odden argues for an array of strategies that allow targeted reductions in specific content such as tutoring or the addition of a teacher for reading and mathematics instruction (Odden, 1990). Targeted addition of staff is based on the idea that smaller groups are effective only for specific content areas. In contrast, a number of implementations have chosen to add students and teachers within a classroom. We could find no researchers who advocated this strategy. In contrast, those that argue for group size suggest that the mechanism that is most potent in changing achievement is teacher knowledge of individual children and through building relationships among teachers, children and families. With large classes paired with more than one teacher, this might be less likely.

What we see in these contrasting assumptions is that researchers have used multiple theories of action related to class size reduction. These theories of action lead to valuing different criteria and the use of different measures. Even the same measures are interpreted

differently when they are informed by contrasting values. The effects of these theories of action will be seen in subsequent sections in which specific outcomes and mechanisms are described and analyzed.

Timing, intensity, and duration of treatment

Prevention scientists typically analyze treatments in terms of 3 key elements: timing, intensity and duration (Durlak, 1997). One way to assess the literature on class size reduction is to use like criteria. Important questions, from this perspective would be *when does class size reduction have the most effect?(timing)* *How small does the group need to be to optimize the advantage?(intensity)* *And How many years of small classes are necessary? Or conversely How few years of small classes will elicit a sustainable effect?(duration)*.

The literature on class size reduction mirrors many of the policies implemented in the past twenty years. In terms of the question of timing, researchers seem to agree that investments in class size reduction are most effective in the primary grades (Addonizio & Phelps, 2000; Biddle & Berliner, 2002; Jeremy D. Finn, 2002; J.D. Finn et al., 2003; Robinson, 1990). CSR is consistently suggested in the range of kindergarten to grade three. Bruce Biddle pointed to the clarity of that point in the research:

The evidence favoring the small classes in the early grades is just overwhelming, absolutely overwhelming. There is just no way to wiggle around it. Critics have tried to criticize individual studies, which you can always do, but that's not the issue. The issue is alright, what about the next study, and so forth. You can't brush it all aside. (Biddle interview)

He goes on to state however, that the findings related to timing might be an artifact of who has received the treatment. Despite strong support by teachers organizations for implementation across the board, most of the recent attention has focused on primary grades:

There is very little good research that has to do with class size reduction in the upper grades. Even though the teachers unions are very much in favor reducing class size and

teachers are very much in favor of reducing class size at all grade levels and I would prefer not to subject sophomores in high school to large lecture classes. It sounds ridiculous but the fact of the matter is that there is just very little really good evidence at that level. In other words, most of the stuff that has been done has been done at the early grades. (Biddle interview)

Others were not so convinced of the early is better argument. Alex Molnar felt that targeting an age group was a political decision to allocate resources rather than an empirical matter:

The argument over kindergarten and first grade is not an empirical argument. It's an argument about policy preferences and the relative cost of programming. That's what it's about. It can't be answered through social science, it's a political argument. So it's not in that sense an evaluation problem. People don't disagree about what the results say, they disagree about what the implications of the results are. (Molnar interview)

Theorizing the early is better effect has been broadly debated. For example Ehrenberg et al (2001) suggest that class size reduction might have a *latent* intervention effect—that the early experience in a smaller group provides the opportunity for children to develop skills and dispositions that allow them to take better advantage of the opportunities provided in later grades. Others broaden the potential explanations by noting that interventions have two different types of effects: they can change cognitive, psychological, or social development or they change the future context in which the individual grows and develops (Biddle & Berliner, 2002; Grissmer, 1999). Charles Achilles suggests that it is essentially socialization to the institution of schooling – start it early and it will be more successful:

Small classes have to start when the kids start school... the same reason by analogy that we do inductions for teachers or mentoring or we do apprenticeships or things like that, kids have to learn what school is, they don't know natively at age four and five what school is. (Achilles interview)

The issue of how small has been connected to the grade level at which class size reduction occurs (Addonizio & Phelps, 2000). Most researchers suggest that reductions of at least to 20 are necessary with larger gains seen for reductions below 20 (Addonizio & Phelps, 2000; Biddle & Berliner, 2002; Jeremy D. Finn, 2002). Researchers increasingly contrast

literature on class size reduction (or pupil teacher ratio) and analyses of more intense reductions of small group or one on one tutoring. Early analyses included these two types of research, which probably over-estimate the positive effects of class size reduction (Glass & Smith, 1979). Strong outcomes are asserted for tutoring or groups smaller than three (Odden, 1990; Slavin, 1989). While it certainly makes sense to compare class size reduction with tutoring when making policy decisions, it is in some ways an apples and oranges comparison. Both are strategies for allocating resources and for instruction but are quite different in intent and practice. Joan McRobbie argues that there isn't a specific size suggested by the literature – that this is a wonderful example of pragmatic policy implementation:

There is no magic number but the small classes in the STAR study were between 13 and 17 students. So then states look at that and say how much can we afford and come up with a number like California did, 20. There is nothing you can point to in research that says well 20 is the magic number. (McRobbie interview)

Grissmer's (1999) analysis of duration issues draws support from the early intervention literature that links length of intervention with longevity of benefit, suggesting that more years in small classes will produce larger and sustainable outcomes for students. Other researchers have agreed (Biddle & Berliner, 2002; Jeremy D. Finn, 2002)

Teaching practices

It's not just having small classes but the teachers have to change what they actually do. . . . Teachers have to do things differently with a small class, they have to give students more attention, that's theory but if they are going to get the effects from that, then they have to carry out the theory (Webb interview)

One of the most pervasively held beliefs among researchers is that class size reduction, in and of itself, is only one piece of a complex puzzle of schooling. Although the policy implies that teaching practices would shift in response to the smaller group, researchers have consistently found that teachers use the same strategies in smaller and larger groups (Ehrenberg et al., 2001;

J.D. Finn et al., 2003; Robinson, 1990; Slavin, 1989). While teachers suggest that they provide more individualized instruction, observations of teaching practice do not bear this out. Finn et al suggest that teacher practice changes in quantity (more or less of what they already do) rather than quality (doing things differently).

If we take the no change finding as true, it is important to theorize it. Most researchers suggest structural explanations for this lack of change, asserting that schools, as complex systems, are resistant to long standing practices that have served the system well.

Overall, the weight of the evidence tilts strongly toward a conclusion that reducing class size, by itself, does not typically affect the instructional activities that occur in classrooms. . . The finding that teaching practices do not vary with class size is consistent with recent work on school restructuring. Observers report that teaching methods are highly resistant to changes in school structure. . . This does not mean teachers *cannot* change practice along with class size reductions, but it may take time, and may require opportunities for teachers to learn about other approaches to teaching. (Ehrenberg et al., 2001)

Hanushek suggests another explanation – asserting that teachers do not seem to change their practice, he highlights the fact that teacher quality is more important than group size:

The fact that learning in a classroom is so dominated by the quality of the teacher that it overwhelms any of the small differences in class size we are normally talking about. And secondly, the fact that for the most part, the evidence seems to suggest that with changes in class size, teacher don't change much what they do. So there is no reaction. (Hanushek interview)

For Hanushek, we should be focusing on teacher quality rather than class size. The variation among teachers accounts for the variation in student performance and when class size reduction is implemented at the state level, it increases the number of teachers needed in what he would characterize as a shallow pool of talent. Employing poor quality teachers in more classrooms compounds the problem of low teacher quality (E. A. Hanushek, 1999).

This finding is contrasted with the assertion that small groups, regardless of the perceived teaching practices, differ in quality from larger groups (Biddle & Berliner, 2002). This was

specifically analyzed in Finn et al's review (2003) where they suggested that teachers in smaller classes get to know their students better because there are more opportunities for interaction and connection between life at home and school. Further, teachers tolerate a wider range of student behavior in smaller classes. These changes provide a context in which students become more engaged and therefore have more opportunity to learn.

Gamoran suggested a differential effect that was developmentally and institutionally fixed.

The kinds of activities that go on in first grade and kindergarten classes are ones that benefit from smaller classes. Whereas the kinds of activities that predominate in classes with older kids, doesn't make so much of a difference. By this I mean, by third, fourth, fifth, certainly middle school, teachers are basically lecturing and engaging in kind of question/answer recitation, and assigning seatwork. Well, those are activities that it doesn't matter if you have 27 or 17 kids, if you are lecturing and doing seatwork. . . . Whereas in kindergarten, first grade, . . . there is a lot more small group work, there might be more one-on-one-contact with teachers. Certainly there is more of a concern with establishing a nurturing environment. So those are the kinds of activities that might especially benefit from having small classes. (Gamoran interview)

One suggestion made by researchers in the literature and interviews related to teacher professional development was that any investment in a structural change needs to be accompanied by support for teacher change. Joan McRobbie made the case that:

You often hear people say what should accompany smaller classes is professional development of teachers so that they learn strategies for teaching in small classes but then no one can seem to answer the question what are those strategies, what do you do differently if you have a class of 35 versus a class of 20 or whatever. (McRobbie interview)

Echoing the findings on teacher practice, Odden (1990) pointed to work from the STAR study that noted that professional development related to teaching strategies for smaller groups had little effect on grade three teaching practice. Jeremy Finn alluded to the issue of teacher quality, noting that an influx of new teachers in a class size reduction context paired with the need for retraining current teachers makes a system of professional development especially important:

When inexperienced teachers are placed in classrooms without adequate preparation, the result can be a level of disorganization that requires weeks or even months to remedy. These dynamics can easily offset the academic benefits of small classes. Brian Stecher and his colleagues recommend that CSR initiatives be undertaken slowly and with careful planning. Well designed programs of professional support and development can also help. (Finn, 2002)

McRobbie did have some ideas at the level of professional community. She thought of professional development systemically, as something that was not a “make and take” approach provided in an isolated inservice but instead was based on and generated a stronger community of colleagues:

The key issue is figuring out how to make sure the teachers get the kinds of professional development that they need in order to be able to do the job, as opposed to the kind where you just go off to a hotel to a conference all day . . . , the kind that is sort of embedded in your daily work that has to do with mentoring and modeling and meeting with other teachers and sitting down and saying, “Here’s what we know the kids need to learn, how do we know if they are learning it and what do we do if they are not?” A staff working together, the staff who work all with the same students group sitting down. (McRobbie interview)

This is a very different image of professional development – instead of delivered to teachers by “experts” this collaborative approach suggests that cultural change and leadership are vitally important to the development enterprise.

Who benefits from class size reduction?

The question of who benefits can be addressed from a variety of perspectives. At the level of policy analysis the typical perspective is focused on who has the largest outcomes. When we compare the varied subgroups that have participated in CSR studies, researchers are typically united behind the idea that smaller classes are most beneficial for children living in poverty and for children of color. For this reason, most analysts suggest targeting the resources allocated for CSR to these groups.

Class size reduction in the early grades helps kids who are at risk primarily. In other words, it’s not an across the board thing. That doesn’t mean that middle class kids aren’t

helped, they are helped by class size reduction, but the evidence suggests that the amount of help is greater for kids from minority backgrounds and from impoverished homes (Biddle interview)

In our interview, David Grissmer distinguished middle class students from middle class parents, noting that it might be “sort of an insurance policy for high income parents,” because they will have to spend less time on homework and out of school support of student learning. A number of middle class parents found they wanted such an insurance policy in Wisconsin when the initial SAGE studies found positive outcomes for children living in poverty. The program was subsequently offered to all districts with reimbursement provided only for the low income students served.

Almost all researchers pointed to the incredible political popularity of class size reduction – that voters, particularly parents, found it an easy to understand, compelling program. This was surely the case in Florida, where voters mandated class size caps through a popular vote. Eric Hanushek described the political payoff in California

Well, my colleague Pete Wilson started this whole thing as governor of California, his popularity doubled within a two or three week period when he announced class size reduction policy for the state. And so, for that reason, 20 plus governors had announced their own class size reduction policies within a month and the President of the United States announced his class size reduction policy within a month and a half. (Hanushek interview)

Teachers were seen as the other beneficiaries of class size reduction. Increases in job satisfaction, reduction in stress and a perceived reduction in work load were the benefits suggested for teachers. Norman Webb noted SAGE’s popularity with teachers and suggested that its support from the state teachers union was instrumental in its strong position in the state educational program. Changes in California’s program have met with much resistance by teacher and parent groups according to McRobbie:

Every time it comes up to make some kind of change, groups like the CTA and PTA (the Teachers' Association and the Parent Teaching Association) slam the brakes on because they are afraid of losing it. They are afraid that it will get watered down and that that would be a first step toward losing it. (McRobbie interview)

The political power of teacher organizations to mobilize and leverage policy was often framed as pork barrel politics—on the one hand it diminished the notion that policy is implemented in an impartial manner relying on “science.” For researchers like Eric Hanushek, it was guild mentality – it made the job of teaching easier and increased the demand for new teachers. For others it biased teacher reports about their practice and therefore the results of many studies—they were likely to react positively to CSR because it was a good thing for them. This was the underlying image in Gamoran's caution about relying on teacher report data:

So it's really important to get beyond teachers' opinions. Their opinions are more favorable because the experience is better, because it's an easier job. It's really important to just not take what they say at face value and try to make an objective assessment of what they are experiencing (Gamoran interview).

On the other hand, the recognition that this reform had the backing of powerful political groups; that it was part of deals forged for both educational and political reasons during a period of relative budget affluence was surprisingly refreshing.

What research do we need?

In the most glaring example of how the literature does not talk, we found a cacophonous set of when we asked what kind of research was needed to settle the question of class size reduction. Responses ranged from Hanushek's design that addressed all the weaknesses in the STAR study, noted earlier in the paper. Adam Gamoran also favored another experiment, building on the use of high stakes tests for a measure of efficacy and paired with tightly designed observational studies nested within the experiment to detail instructional practices.

But I would like to see more quantitative work that would use the high stakes tests the states are always using, that are already using and would try to quantitatively or

statistically link instructional activities with student outcomes. That would be key.
(Gamoran interview)

Though he didn't call for a randomized experiment, Norman Webb focused on large scale comparative studies and multimethod studies that linked teacher practice and student outcomes. Variations on this theme came from those who thought we needed good cost-benefit analyses (Achilles, Webb, McRobbie) or longitudinal analyses that followed students into later schooling (Achilles, Biddle; Molnar). All of these researchers wanted to fine tune our knowledge of the effects of class size reduction through more detailed studies.

Experiments were much less important to others. For example, both Biddle and Achilles had a "been there done that" mentality – they felt that STAR had done the job of telling the story that could be told with an experiment:

I don't think we need another experiment. I think STAR answered that question. If you want to do one, if you don't believe that when they put hydrogen and oxygen together you get water, then you can go ahead and do it. We actually ran that experiment manipulative no variables but class size, despite what people say with a caveat that we must let schools run as they run. (Achilles interview)

Instead of seeing experiments as the gold standard for research (a pervasive concept in education research recently), David Grissmer came at the issue from a theory of experimentation—he knew the limits of what we can know from experimental studies

While more experimentation seems essential to making progress in educational research, experiments can never be depended on to solve all the complex and contextual effects. Educational research will probably never follow health research where trials are needed for every new intervention before implementation. (Grissmer, 1999, p. 239)

In its place, Grissmer suggests that we should be focusing on the *why* question at this point – that by asking why we get the outcomes we do in research on class size reduction, we can begin to develop what we called earlier a theory of action:

One of the key problems in the research community is that we are not oriented to answering the question "Why do things work?" We tend to measure class size effects and go home rather than say, "Well, why do class size effects work, why do smaller

classes work?” . . . [We need to] begin to develop a theory of what happens that improves achievement. That is, experiments only give us sort of a single measurement. What science tries to do is develop a theory to explain several measurements. Parents do this and teachers do this . . . But it’s a theory, the scientific consensus forms around a theory. (Grissmer interview).

A similar approach was suggested by Joan McRobbie, who thought we needed to be much more synthetic and derivative in understanding of class size reduction. Rather than focusing exclusively on new studies, McRobbie thought we needed a better understanding of the existing literature, with attention to distilling it for policymakers

If you don’t [make recommendations] then you have policy makers just sort of drifting and it seems like you really have to try to give them information about what’s been learned by other policy experiments. So I put emphasis on what you can take from all of the sort of state wide policies, or even local policies, that have been designed so far that are about class size reduction. What can you make of all of that in terms of the wisdom that has now been built about designing a policy based on the research that’s been done. (McRobbie interview)

Reading the literature and suggesting next steps for researchers is very much a situated activity. The range of responses to this question illustrates that what we know about a particular topic is shaped by how we know it – by the perspectives we bring to the literature, by the criteria we use for judging its adequacy, and the communities of practice to whom we speak. For some, the aggregation of data and findings and increasingly tight experimental designs are key to our knowledge. Through replication and better specification of findings we can assert with surety the outcomes of treatments. For others, longitudinal designs examining long-term effects are favored. And others look for policy relevant synthetic analyses that focus on the *how* rather than the *what*. From this perspective, treatments happen in contexts that make their effects contingent and often involve tradeoffs of other services and investments. While it would be tempting to suggest that perhaps all these researchers are right, that in a perfect world, we would have all these types of studies going on at the same time, we are not exactly in that perfect place. In

contrast to the context that prompted much of the work on the large state implementation, characterized by budget surplus and interest in multiple types of inquiry we are now in a time of budget retrenchment and increasingly narrow views of what counts as data or as science. In the final section we add our voices to this conversation, suggesting what we have taken away from our reading and interviews related to class size reduction.

Discussion

The large and varied literature related to class size reduction rests on early meta-analyses and re-analyses that variously find positive effects on student outcomes. More recently, three major state initiatives have given us a practiced based view of CSR. STAR provides a state-sampled experiment of various CSR treatments, SAGE provides a phased-in implementation of a targeted intervention, and the California implementation provided resources to the entire state for a midrange class size reduction of 20:1 in kindergarten to grade three. At this point, we know that none of the research is perfect – each had strengths and weaknesses that reflect the politics and compromise of the moment. But across these research conversations, we would have to say that we think we are getting *good enough* research – that taken as a group, we can begin to form a fuzzy picture of socially held knowledge of class size reduction.

The biggest take home message is that *class size reduction is not pupil teacher ratio*. As long as these two frameworks for allocating resources are conflated we will have continued disagreement about how CSR affects student outcomes. While PTR is the simplest indicator, it often does not reflect the day-to-day life in the classroom and is a good example of how a proxy is not always the best representation of a construct.

Considering CSR as a prevention strategy leads to consideration of how researchers have examined the intervention in terms of timing, intensity, and duration. While most researchers

agree that earlier is better, it has also been pointed out that the research base would be stronger if more studies were done at the intermediate and middle school level. The question of how small groups have to be generally is answered as at least 20 (California's strategy) but better at 15. One confound in these conversations was the inclusion of extremely small groupings – tutoring or pull-out programs. There is certainly a choice in terms of investment that is implied (do you reduce class size overall, do you reduce for particular content, or do you provide pull-out small groups) but including it within the literature seems to mix apples and oranges.

Researchers generally state that what teachers do is as important as how many students they have—that instructional practices must change if CSR is to be effective. Most researchers we read and talked with could not point to work that showed changed – in fact, they held that no change occurred. One question we find ourselves asking is what constitutes a change in teacher practice? What has to change for researchers to recognize it as something different? If the quantity of certain practices change, aren't things different? One way that we might better understand the outcomes of class size reduction is to have a better sense of the nature of instruction in the varied CSR instructional contexts.

If CSR is an intervention, should it be targeted to those most in need or should it be universally administered as a benefit to all? Whose benefit should be considered when such a program is implemented? Most researchers agreed that CSR had the greatest potential for students typically seen as at risk—children of poverty and of color. Others extend the notion of benefit to middle class families, suggesting that it is a kind of insurance policy for students who are likely to do well in school anyway. The immense popularity of CSR was also recognized – parents wanted the benefit for their children and teachers wanted it to make teaching easier. This

popularity made changes in the policy difficult to leverage as the electorate and lobbying groups worked hard to maintain an easy to understand policy alternative.

One of the first things we noticed when we got into the literature on class size reduction is that it is anything but dry and dispassionate. There is more sniping and namecalling than we ever thought there would be. One interview was practically unusable for this paper as the researcher spewed so much venom about state education officials he disagreed with and about other scholars who have researched the topic. Why do people get hot under the collar when talking about how many kids to put in a classroom? Our hunch, after much reflection, is that it turns on two key issues. The first is that the conversations about class size reduction are flashpoints for the issue of what counts as evidence in education research. When researchers don't see their value system operationalized in inquiry they raise flags about the work's adequacy. The rules of the game are not unitary, but are shaped to the perspectives of varied communities of researchers. No wonder policy makers have a hard time making sense of what we do – researchers are not a unified “we” but instead are more likely characterized as tribes of likeminded scholars. This played out specifically in terms of what researchers saw as the findings of the research and their suggestions for next steps.

The second point is that the policy of class size reduction turns on basic issues of equity. It forces us to think about the basic inequality of schooling, the mechanisms that might mediate those differences, and the investments that we are willing to make as a society to put success in reach of all children. The very idea that people would think that achievement would go up just by changing the number of students in a classroom is typical of the pie in the sky thinking that afflicts much of education. Class size reduction is not a silver bullet that can ameliorate the damage that poverty, violence, inadequate child care, etc set in the path of children coming to

school for the first time. It is enacted in the institutional context of high stakes testing, crumbling buildings, increasing numbers of children with high needs, and competition from other programs for teacher attention and effort. Adding resources for class size reduction is a nod toward the inequities that form the foundation of schooling but it is not enough. CSR is part of a system of reforms and problems that will be most fruitfully considered in a coordinated manner—in terms of both the practice and research of schooling. It necessitates implementation that connects the utilization of the resources for class size reduction with all curricular, administrative, and institutional efforts that shape teaching and learning. It will require many different kinds of inquiry – large scale studies that allow disaggregation of results as well as small scale studies of process and mapping of social relations and practice. Berger was right – focusing on class size reduction alone is like trying to decide the right amount of butter in a recipe. We have a better sense of the butter and now it is time to take a look at the other ingredients.

References

- Addonizio, M. F., & Phelps, J. L. (2000). Class Size and Student Performance: A Framework for Policy Analysis. *Journal of Education Finance*, 26(2), 135-156.
- Berger, M. A. (1982, April 17-20). *Class size is not the issue*. Paper presented at the National School Board Association, Atlanta, GA.
- Biddle, B. J., & Berliner, D. C. (2002). Small Class Size and Its Effects. *Educational Leadership*, 59(5), 12-23.
- Cullen, S. (2005, February 10, 2005). District says proposal would help with shortfall. *Wisconsin State Journal*, p. A5.
- Durlak, J. A. (1997). *Successful prevention programs for children and adolescents*. New York: Plenum.
- Ehrenberg, R. G., Brewer, D. J., Gamoran, A., & Willms, J. D. (2001). Class size and student achievement. *Psychological Science and the Public Interest*, 2(1), 1-30.
- Finn, J. D. (2002). Small Classes in American Schools: Research, Practice, and Politics. *Phi Delta Kappan*, 83(7), 551-560.
- Finn, J. D., & Achilles, C. M. (1990). Answers and Questions about Class Size: A Statewide Experiment. *American Educational Research Journal*, 27(3), 557-577.
- Finn, J. D., Pannozzo, G. M., & Achilles, C. (2003). The "whys" of class size: Student behavior in small classes. *Review of Educational Research*, 73(3), 321-368.
- Gilman, D. A., & Kiger, S. (2003). Should We Try To Keep Class Sizes Small? *Educational Leadership*, 60(7), 80-85.
- Glass, G., & Smith, M. L. (1979). Meta-analysis of research on class size and achievement. *Educational Evaluation and Policy Analysis*, 1(1), 2-16.
- Grant, C., & Graue, M. E. (1999). (Re)viewing the review: A case study of the Review of Educational Research. *Review of Educational Research*, 69(4), 384-396.
- Grissmer, D. (1999). Class Size Effects: Assessing the Evidence, Its Policy Implications, and Future Research Agenda. Conclusion. *Educational Evaluation and Policy Analysis*, 21(2), 231-248.
- Hanushek, E., & Rivkin, A. (1997). Understanding the twentieth-century growth in U.S. school spending. *Journal of Human Resources*, 31(1), 35-68.
- 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., & Stock, W. (1983). The effects of class size: An examination of rival hypotheses. *American Educational Research Journal*, 20(1), 63-85.
- Hruz, T., & Wisconsin Policy Research Institute. (2000). *The costs and benefits of smaller classes in Wisconsin : a further evaluation of the SAGE program*. Thiensville, WI: Wisconsin Policy Research Institute Inc.
- James, J., & Johnson, C. (2005, February 15, 2005). Bush yields on class sizes. *St. Petersburg Times*.

- Molnar, A. (2002). *School Reform Proposals: The Research Evidence*. Research in Educational Productivity Series (No. Vol-2). Connecticut.
- Molnar, A., & Zmrazek, J. (1994). *Improving the Achievement of Wisconsin's Students*. Urban Initiative Task Force Recommendations and Action Plan. Bulletin No. 95079. Wisconsin: Wisconsin State Dept. of Public Instruction, Madison. Bureau for Policy and Budget.
- Nye, B., Hedges, L. V., & Konstantopoulos, S. (2001). The Long-Term Effects of Small Classes in Early Grades: Lasting Benefits in Mathematics Achievement at Grade 9. *Journal of Experimental Education*, 69(3), 245-257.
- Odden, A. (1990). Class Size and Student Achievement: Research-Based Policy Alternatives. *Educational Evaluation and Policy Analysis*, 12(2), 213-227.
- Robinson, G. E. (1990). Synthesis of Research on the Effects of Class Size. *Educational Leadership*, 47(7), 80-90.
- Slavin, R. (1989). Achievement effects of substantial reductions to class size. In R. Slavin (Ed.), *School and classroom organization* (pp. 247-257). Hillsdale, NJ: Earlbaum.
- Smith, M. L., & Glass, G. V. (1980). Meta-analysis of research on class size and its relationship to attitudes and instruction. *American Educational Research Journal*, 17(4), 419-433.
- Smith, P., Molnar, A., & Zahorik, J. (2003). Class-Size Reduction A Fresh Look at the Data., *Educational Leadership* (Vol. 61, pp. 72-74): Association for Supervision & Curriculum Development.
- Webb, N., Meyer, R., Gamoran, A., & Fu, J. (2004). *Participation in the Student Achievement Guarantee in Education (SAGE) Program and Performance on State Assessments at Grade 3 and Grade 4 for Three Cohorts of Students—Grade 1 Students in 1996-97, 1997-98, and 1998-99*. Madison, WI: Wisconsin Center for Education Research.
- Zahorik, J., Halbach, A., Ehrle, K., & Molnar, A. (2003). Teaching Practices for Smaller Classes. *Educational Leadership*, 61(1), 75-77.
- Zahorik, J., Molnar, A., & Smith, P. (2003). *SAGE advice: Research on teaching in reduced-size classes*. Tempe: Education Policy Studies Laboratory.