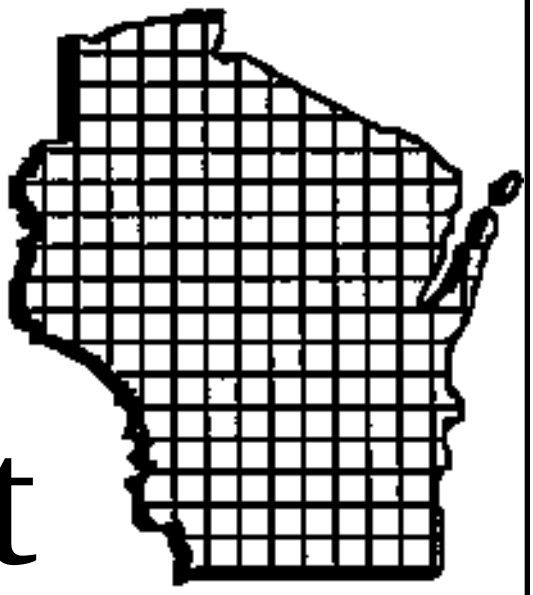
Wisconsin
Policy
Research
Institute
Report



September 2000 Volume 13, Number 6

THE COSTS AND BENEFITS OF SMALLER CLASSES IN WISCONSIN

A Further Evaluation Of The SAGE Program

REPORT FROM THE PRESIDENT:

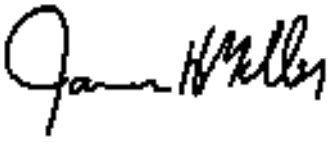
Class size. It is the current chic educational idea. In the past, chic reforms such as busing, open classrooms, whole language, and new math have produced no academic achievement. Smaller class size is an issue that most people in Wisconsin intuitively support. After all, there is little question that it is easier to teach fifteen children, rather that twenty or twenty-five — or even ten children rather than fifteen. The problem is that there has been very little research on the effectiveness of smaller class size, a program that potentially involves hundreds of millions of Wisconsin tax dollars.

We asked Thomas Hruz, a Resident Fellow at our Institute, to examine current research on class size, not only in Wisconsin, but also across the country. Hruz has a graduate degree from the La Follette School in Public Policy Analysis and Public Administration. He has worked as a researcher for the Consortium for Policy Research in Education at the University of Wisconsin. He is currently finishing a Law Degree at Marquette University and has years of experience in educational research.

In this study on smaller class sizes in Wisconsin, one reaches the conclusion that there is very little academic research showing significant gains for children — especially minority children — in smaller classes.

This research raises some interesting questions. Is bending to what the educational establishment wants the best use of tax dollars? We know that we will have to hire a lot more employees for smaller class sizes, but unfortunately there is no accountability for whether the hundreds of millions of tax dollars will improve academic skills. If we are going to spend more money on education, is it better spent on improving teacher quality? There has been little debate on the merits of the quality of the teacher in these small classes. If you have a bad teacher, it doesn't matter if they have fifteen, twenty or twenty-five students. They are still a bad teacher. There is growing academic research showing that the real key to improving the learning skills of all our children — and especially our minority children — is strong, quality teachers, and not necessarily class size.

This is the debate we should be having in Wisconsin. Do we continue to spend tax dollars on lowering class size, or should we concentrate on additional accountability for our educators, and provide the best quality teachers in the classroom with economic rewards for their expertise.



James H. Miller

WISCONSIN POLICY RESEARCH INSTITUTE, INC.
P.O. Box 487 • Thiensville, WI 53092
(262) 241-0514 • Fax: (262) 241-0774
E-mail: wpri@execpc.com • Internet: www.wpri.org

THE COSTS AND BENEFITS OF SMALLER CLASSES IN WISCONSIN

A Further Evaluation Of The SAGE Program

THOMAS HRUZ

	Page
EXECUTIVE SUMMARY	1
WISCONSIN'S CLASS SIZE REDUCTION PROGRAM	4
THE PROSPECTS FOR THE SAGE PROGRAM IN THE FUTURE	7
SUPPLEMENTAL CLASS SIZE REDUCTION EFFORTS IN WISCONSIN	8
TOTAL EXPENDITURES ON CLASS SIZE REDUCTION EFFORTS IN WISCONSIN	9
POLITICAL SUPPORT FOR SMALLER CLASS SIZES	11
IN SUPPORT OF SMALLER CLASS SIZES	12
CRITICAL ANALYSES OF SMALLER CLASSES	15
THE DEVIL IN THE DETAILS: (RE)EVALUATING WISCONSIN'S EXPERIMENT IN CLASS SIZE REDUCTION	17
SECOND AND THIRD GRADE RESULTS	23
DYNAMICS OF THE AFRICAN-AMERICAN VERSUS WHITE STUDENT ACHIEVEMENT GAP	29
EFFECT SIZES: JUST HOW BIG ARE THE SAGE GAINS?	31
SUBJECT AREAS	32
QUESTIONS LEFT UNANSWERED: LOW-INCOME STUDENTS AND MPS EFFECTS	33
SMALLER CLASS SIZE VERSUS OTHER EDUCATION POLICIES	34
CONCLUSION	37
APPENDIX	38
NOTES	40

EXECUTIVE SUMMARY

Wisconsin has joined prominently in the growing movement toward the implementation of smaller class sizes in public elementary schools. Based largely on the support of conventional wisdom, heavy lobbying by the teachers' unions and various political leaders, and a less-than-thorough presentation of the local and national evidence on class size, smaller classes are now becoming all the rage.

Wisconsin's Student Achievement Guarantee in Education (SAGE) program began in the 1996-97 school year as a pilot program in 30 schools serving predominantly low-income students. The program's primary feature is the reduction of class sizes to fifteen students per teacher in grades kindergarten through third. SAGE now boasts well over 500 elementary schools as participants, and continues to expand. In the 2000-01 school year, the State will spend approximately $58 million on SAGE, while also receiving $22 million in aid from the federal government's class size reduction program. In addition, all public schools are now able to participate in SAGE, regardless of their poverty rate.

So what will these changes mean for education in Wisconsin?

One critical lesson that can be drawn from both the national research on class size policies and the results of Wisconsin's own SAGE program is that smaller classes do not always provide identifiable achievement benefits. Moreover, when smaller classes do help raise student achievement, the greatest gains tend to occur only in certain grades and for particular populations of students. In addition, achieving these results necessitates an immense and continual cost to taxpayers.

As part of the SAGE program, annual evaluations have been conducted to determine the quantitative and qualitative effects of the program, particularly on student achievement. Unfortunately, a complete explanation of the statistical findings of these evaluations has yet to be conducted. Instead, only the positive effects of the program have been disseminated to the public, while the more ambiguous results revealing only minor effects from smaller classes have been suppressed.

When it comes to improving academic achievement, class size reductions achieved through the SAGE program have not been as significant as is commonly argued and assumed. Most notably:

- **Smaller classes in the second and third grades had a minimal impact — and in some cases had no additional impact — on student achievement**. The available data reveal that while greater gains are consistently made by students in smaller classes than students in regular-sized classes in the first grade, no such consistent advantage is found from being in smaller classes in the second and third grade. In other words, in many cases students in smaller classes achieve no better than their counterparts in regular-sized classes in these later grades.
- **African-American students in smaller second and third grade classes in particular did not gain relative to their gains made in the first grade or relative to African-American students in regular-sized classrooms**. Not only do smaller classes in the second and third grade have little discernable achievement effect for students generally, but they likewise have a minimal effect on African-American students specifically. This is remarkable, given that African-American students show by far the greatest achievement gains from being in smaller classes in the first grade.
- **Smaller classes appear to not have any effect on students who are not African-American, who constitute the majority of students in SAGE**. One of the most noted results of the SAGE program has been that smaller classes appear to reduce the achievement gap between African-American students and other students, most noticeably White students. Yet a closer inspection of the SAGE data suggests that the gap is being reduced only because African-American students are, on average, doing better in smaller classes, while all other students from all other ethnic groups seem to have a negligible achievement effect from being in smaller classes.
- **The actual magnitude of the gains experienced by students in SAGE are, on average, relatively meager**. While proponents of the SAGE program and smaller classes incessantly tout that the program shows "statistically" significant gains from students in SAGE over students in regular-sized classes, the actual magnitude of

these gains is small. On average, students in SAGE are scoring only about one and a half to five percent higher on tests, depending on grade level and subject.

Furthermore, many important questions related to class size reduction policy in this state remain largely unanswered. In particular, given what is known about the program's large effects on African-American students, and the much higher concentration of such students in Milwaukee, it may be useful to determine if the aggregate SAGE gains are being driven solely by students in the Milwaukee Public Schools (MPS).

Class size reduction policies should be made in an informed and efficient manner, such that the public investment in the policy results in a meaningful improvement in education. Unfortunately, the data results less favorable to the unabated expansion of SAGE have not been well disseminated to the public or to government officials who have greatly expanded the program in recent years. Such results suggest that more limited implementation of smaller classes, in only the first grade and in only high poverty schools, can produce nearly the same results presently experienced, but at a far lesser cost. Moreover, the funds expended to meet these class size reductions may be much more efficiently used for other programs that help the same students aided by smaller classes, or to improve such educational factors as teacher quality and experience, which have regularly been shown to have a greater impact on student achievement, whether in small or regular-sized classes.

INTRODUCTION: THE BOOM OF SMALLER CLASS SIZES

A growing centerpiece of current education reform, both nationally and within Wisconsin, is the movement toward smaller class sizes. The idea of reducing class size carries considerable appeal because it relies on presumably strong elements of common sense. It is believed, not unreasonably, that if a teacher has fewer students in a classroom, then each individual student will naturally receive more attention and individualized instruction, and will therefore learn better. Additional benefits are assumed from teachers being able to better maintain discipline, by improving teacher morale, and by increasing student participation — matters not directly related to student achievement, but which can have a derivative impact on the learning process.

Not unexpectedly, these perceived benefits have caused many educators, school administrators, policy makers, and much of the general public to applaud the concept of smaller classes. To a notable extent, members of both the major political parties are endorsing policy changes to this effect, although Democrats are generally the louder proponents. Smaller is better as far as classroom settings and instruction go, and since students deserve the best education, they deserve classes with fewer classmates competing for their teachers' attention.

Or so the story goes.

Unfortunately, when a public policy idea tends to become incredibly popular, many important details and relevant issues concerning that policy tend to get suppressed or, at a minimum, are obscured. This is the experience with class size reduction efforts in Wisconsin.

The contention set forth in this report is not that smaller classes do not help make teaching easier, nor that smaller classes do not have some degree of academic benefit, for some students, and in some educational contexts. The critique largely comes against the wholesale treatment of smaller class sizes as a simple, clear-cut reform measure that is unassailable. In Wisconsin, an unfortunate result of this uncritical support for smaller classes is that the state's class size reduction program — SAGE — is being expanded in ways that may prove to have little or no effect on improved student performance.

What can be drawn from the national research and the results of Wisconsin's own SAGE program is that smaller classes do not always provide identifiable achievement benefits, and when they do raise student achievement, the greatest results tend to occur only in certain grades and for particular populations of students. In addition, achieving these results necessitates an immense and continual cost to taxpayers.

HOW MANY STUDENTS PER TEACHER?

Class Size Versus Student-Teacher Ratios

In most of the estimations provided in the following text, the attempt has been to use a true "class size" measure, as opposed to simply a student-teacher ratio. Proponents of smaller class sizes often rebuke that student-teacher ratios are not the appropriate metric from which to determine if smaller classes create beneficial results. They note that teacher-student ratios are often strictly mathematical calculations of school district personnel records for licensed instructional staff, which also include special education, music, physical education, and teacher-aide personnel, all of whom do not perform their jobs in the traditional style of a classroom teacher. Therefore, these numbers do not reflect the reality of how many students a teacher must *actually* teach in any given classroom. In large part this point is a very salient one. The issue is what are the actual class sizes a teacher faces, for that will determine the pure effects of class size. Still, given that actual class sizes are often much larger than student-teacher ratios, one may reasonably wonder if smaller classes are achievable through a redistribution of teachers, either within or between districts and states, rather than from a mass recruitment of new teachers?

Estimations Of Standard Class Sizes In Wisconsin

An obvious, yet material assumption of any class size reduction policy is that current class sizes are too large, or at least too large to achieve the quality of education desired by educators and the public. We know from survey responses that teachers frequently claim that the number of students in their classrooms is "too high."[1] But is that really the case? How large are Wisconsin's classes in the absence of reduction policies?

Unfortunately, there seems be some confusion as to what exactly is the baseline class size from which reductions are to be made. A couple of sources are available to make this estimation. A recent estimate by the US Department of Education showed Wisconsin having 14.4 students per teacher statewide in school year 1998-99.[2] That ratio places Wisconsin 13th out of 50 states in the measure of more teachers per students, and below the national average of 16.5 students per teacher. The Department also has put out statistics showing the average class size in Wisconsin, which, in 1993-94, was at 23 students, just below the national average of 24.[3]

Another estimate has been made by the Wisconsin Education Association Council, the state's largest teachers union. According to its figures, the typical all-day kindergarten class in Wisconsin has 21 students, and elementary classrooms across the state average 22 students. Yet these numbers vary widely, depending largely on the wealth of a given district and by whether it is rural, suburban, or urban. For example, in the Milwaukee Public Schools the numbers are higher, at 25 in kindergarten and 27 in elementary schools.[4]

Yet the veracity of any of these numbers is debatable. The UW-Milwaukee SAGE evaluation, discussed in great detail later, uses "comparison" schools that are supposed to have class sizes reflective of those currently seen in schools across the state. In these comparison schools, the average class size for kindergarten classes was only 19 students, with first through third grade classes averaging around 20 students.[5] And of the sixteen schools that compromise this comparison group, six are from Milwaukee, which according to the WEAC estimates should have higher average class sizes.

Overall, there is still some uncertainty as to the actual number of students in each classroom in Wisconsin's public elementary schools. Yet it is important to establish a reliable baseline number for regular-sized classes in order to determine the marginal differences being experienced, and where those differences are great and where they are only minor. Nevertheless, the measures provided above give a rough sense of what is the baseline class size in the state, and gives us a sense of how far reductions in the various programs are coming.

Implementing and Evaluating Class Size Reduction Programs: A Growing Endeavor

Based on the perceived merit of class size reductions, states and localities across the country have increasingly implemented both experimental and more comprehensive programs aimed at class size reductions. Some of the most prominent of these programs have taken place in North Carolina, Indiana, Nevada, California, and Tennessee, to name but a few. In Tennessee, the STAR program, along with its correlative evaluation, is considered the benchmark of class size reduction efforts. The conclusions of the STAR evaluations and their potential flaws, along with the program's political impact, are discussed later in this report. Instead, the majority of the following analysis focuses solely on Wisconsin's own experiment in class size reduction — the SAGE program.

Wisconsin's Class Size Reduction Program

Wisconsin joined in the policy movement toward smaller class sizes largely through the development of its Student Achievement Guarantee in Education (SAGE) program. The centerpiece of this program is the reduction of class sizes to a maximum of 15 students per teacher in grades kindergarten through three.

The SAGE program was born out of recommendations from the 1994 Urban Initiative Task Force. This group was organized by the state Department of Public Instruction (DPI) and was composed of 34 members from around the state, representing various constituencies. The Task Force met to discuss issues involved with increasing student achievement in low-income school districts and to then develop potential strategies and recommendations to further improve educational outcomes for these students.

At the closure of the Task Force's meetings, a "Recommendation and Action Plan" was published that embodied the members views on how to best aid low-income students. The four recommendations the task force presented now embody the four elements of the SAGE program. As to class size reduction specifically, the Task Force recommended extensive reductions, such that class size reductions to 15 students per teacher would eventually occur in all kindergarten through fifth grade classrooms, and also in all the core subjects for grades 6 through 8. These recommended reductions are far greater than those that were eventually adopted by the legislature and enacted through the SAGE program. The Task Force also recommended that this action plan be implemented in every school in the state, although admitting a need to emphasize "schools serving the children most in need."[6]

The SAGE law codified the thrust of the Urban Initiatives Task Force recommendations when it was officially created by the state legislature in the 1995-97 biennial budget. SAGE is currently governed under Wisconsin Statute 118.43.[7]

The SAGE Contract: The Mandates On A SAGE School

Under the current law, all public schools in the state serving grades K-3 can participate in the SAGE program. If more districts apply than there is funding available, the DPI will award the funds based on two factors: (1) the number of low-income pupils enrolled in grades K-1 in the schools being considered; and (2) the balance of rural and urban schools already participating in the program.[8]

SAGE is administered through a system of five-year grants to specific elementary schools selected to participate in the program by their district. Each grant is tied to a statutorily constructed, five-year contract that is created between the district housing a SAGE school and the DPI. In exchange for $2,000 per low-income student enrolled in the grades for which class size reductions are required,[9] districts in the program are required to implement four program elements in schools receiving SAGE funding:

1. Reduce class size to 15 pupils in the grades for which grants are awarded — these grades will eventually include kindergarten through the third-grade.
2. Keep the school open every day for extended hours and collaborate with the community in educational, recreational, and social service activities. This element is often referred to as the "Lighted Schoolhouse."
3. Provide a rigorous academic curriculum designed to improve academic achievement and which preferably focuses on issues of cultural diversity.
4. Create professional development programs for teachers and require performance evaluations.

Of these four program elements, it is clear that class size reduction is the primary feature. For better or for worse, the Wisconsin SAGE program has become synonymous with class size reduction, while the other three potentially important education reforms are relatively ignored. In many cases, implementation of the three non-class size components has been minimal and is considered achieved by changes not much beyond the status quo. Moreover, the establishment of smaller classes dominates nearly all political discussions of SAGE. As a result of these realities, this report focuses almost exclusively on the class size element of SAGE and reasonably assumes this element is what is driving any academic effects witnessed in SAGE schools.

SAGE schools are given options on how to achieve the class size reduction mandate. As to be expected, a regular classroom with one teacher and 15 students was the primary method of achieving the 15:1 ratio. Of the 356 SAGE classrooms in 1998-99, 264 (72%) were regular classrooms with one teacher, while 69 (19%) were conducted with a two-teacher team teaching up to 30 students, and the remaining 23 (6%) employed various other strategies for achieving the 15-to-1 ratio, including the use of a floating teacher during instruction of specific subjects.[10] The SAGE evaluations have examined if any of these settings were better or worse at improving student achievement than the others. With only a few exceptions, the results show that classrooms with team teachers teaching up to 30 students achieved just as well as classrooms with the regular 15-to-1 student-teacher ratio.[11]

The SAGE Program In Its Infancy

The SAGE program began as a pilot program and was originally implemented on a limited scale. When SAGE began in the 1996-97 school year, 30 schools in 21 districts from across the state participated in the program by reducing class sizes to 15 or less students in kindergarten and first grade classes.[12] The program then required implementation of the same class size reductions for two additional grade levels (grade 2 in 1997-98, and grade 3 in 1998-99). By the time all grades K-3 had reduced class sizes in these 30 schools, the annual cost of the program to these schools stood at approximately $8.6 million.

But SAGE did not remain limited to its original 30 schools. Instead, schools were added to the SAGE program when the state legislature allowed for the creation of a second round of five-year SAGE contracts in the 1997-99 state budget.[13] These contracts began in the 1998-99 school year and are scheduled to run through June 2003. The total number of schools entering under this round was limited to the funds available from state appropriations for the program, after the original SAGE schools had been fully funded for any given year.

Under this expansion of SAGE, a total of 48 more schools from 25 districts were added to the program by 1999, bringing the total number of SAGE schools to 78 in 46 districts.[14] Individual SAGE schools and their funding amounts for years 1998-99 and 1999-2000 are listed in the Appendix. Funding ranged from a low of $24,000 (Boyd Elementary in the Stanley-Boyd Area school district) to a high of $768,000 (Forest Home Avenue Elementary in the MPS), with the average amount being $220,511 and the median amount, $168,000. In 1999-2000, 8,935 low-income K-3 students were in the program, with all schools receiving $2,000 per low-income student who was enrolled in SAGE grades. But since SAGE schools also have a significant number of students classified as non-poor, the total number of students in smaller classes in grades K-3 was larger, totaling 13,635.

To accommodate the inclusion of more grade-levels (in the original 30 SAGE schools) and the addition of more schools (in the second round of contracts), it was necessary to increase the program's funding. As a result, state funding for the SAGE program more than tripled during its first three years of existence, from $4.2 million in 1996-97 to slightly under $15 million in 1998-99. Approximately $8.6 million of that amount went to funding the original 30 SAGE schools, with the remainder going to the second round of SAGE schools. By the 1999-2000 school year, estimated aid for SAGE totaled $17.4 million, with $7.7 million allocated to the original SAGE schools.

Expanding And Reconstituting SAGE In The 1999-2001 Budget

Only a few years into its life as a pilot program, SAGE became targeted for a dramatic expansion that would redefine and essentially change the basis of the program's existence. These changes were promoted, often with great vigor, by certain state legislators, public education officials and union leaders, and other interested parties. During this time, a state-mandated program evaluation of SAGE had been conducted by a team at the University of Wisconsin-Milwaukee, which reported very favorably as to the program's beneficial impact. Armed with this apparently overwhelming positive evidence, the state legislature decided to pass dramatic changes to the SAGE program.

Changes were made both in the magnitude of the program and with regards to its substance. These expansions were embodied in the 1999-2001 biennial state budget act, which increased state support for SAGE by an additional $46.2 million over the next biennium and which allowed all school districts to apply for grants, regardless of their percentage of low-income students. The budget also installed a measure creating a new $3 million categorical aid program to reimburse school districts, with the exception of MPS, for twenty percent of debt service costs associated with SAGE building projects.[15]

In terms of substantive changes, starting in the 2000-01 school year, *every* school district that provides instruction in grades K-3 may enter into a SAGE contract on behalf of one or more of its schools if: a) the school is not receiving a grant under Wisconsin's P-5 program, which is designed to serve disadvantaged children in those elementary grades,[16] and b) the school board had not declined to participate in the first and second round of contracts, if participation had been offered. No longer do schools need a significant number of low-income students to be eligible.

To accommodate this new influx of schools, a third round of five-year grants and contracts was created. Although all schools are now able to participate in the program, actual participation is limited based on the amount of funds available. In apportioning the available SAGE funds for 2000-01, DPI must first pay $2,000 per pupil on continuing first- and second-round contracts and then, to the extent possible with remaining funds (approximately $37.8 million in 2000-01), allocate $2,000 per pupil under the third-round contracts to newly participating schools.

As before, schools must apply to DPI to be part of this third round of contracts. According to the SAGE law, DPI awards grants to newly applying schools by giving priority to those with the highest concentration of low-income pupils. The DPI determines the low-income enrollment for each school by using subsidized lunch eligibility data collected by the Department. According to the DPI, based on 1998-99 data, if all eligible schools in Wisconsin were to apply for third round contracts, the aid available under the 1999-01 budget is enough to fund every school with a poverty rate of 27% or more. This total includes roughly 600 of the approximately 1,100 public elementary schools in the state.

TABLE 1 GROWTH OF THE SAGE PROGRAM

SAGE Contract Round	Initial Academic Year	Number of Schools
First Round	1996-97	30
Second Round	1998-99	48
Third Round	2000-01	500-525

Table 1 shows the cumulative implementation schedule

for the three rounds of SAGE contracts, and the corresponding number of schools that have or will enter into the program in each round.

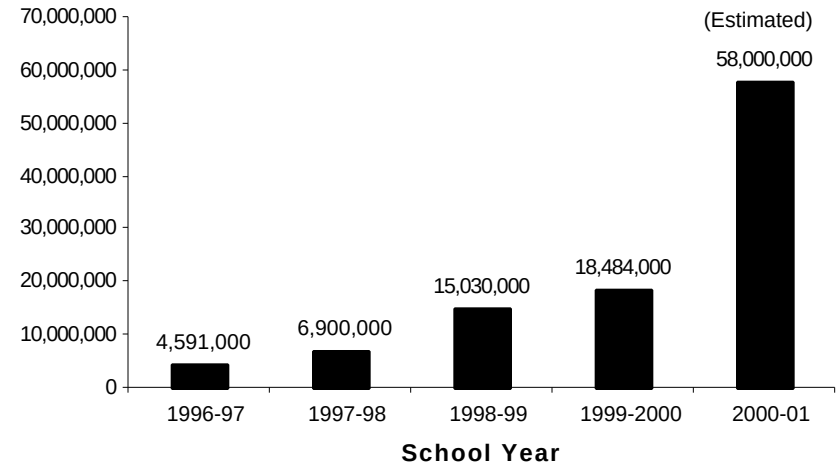
The Impact Of These New Changes To SAGE

Instead of waiting until the beginning of the 2001-03 budget cycle, which would have nearly coincided with the end of the original five-year SAGE contracts, the state has gone ahead with making SAGE a full-fledged state education program. The magnitude of these changes, both in terms of the funding amounts and program characteristics, is difficult to overstate. These new changes greatly expand and essentially redefine the SAGE program.

Figure 1 shows graphically the SAGE funding increases made since the program's inception. The funding increase required for the latest additions stands at 229.2% over base SAGE funding for the previous biennium.[17] To give a sense of the relative size of these increases, at $58.8 million in 2000-01, SAGE is now the second-most-expensive, categorical-aid program in education, behind only special education.[18] While it is dangerous to make some comparisons across school-based programs, it is clear that SAGE is quickly taking its place as one of the most prominent, and expensive, education programs in the state.

Yet perhaps more important than strictly the funding increases, as large as they are, is how the program has lost it primary focus and, arguably, its initial justification. We are witnessing the beginning stages of unveiling the guise of ensuring that the SAGE program only targets low-income, at-risk students. Then again, the SAGE program never actually targeted low-income students, as much as it financed schools that were deemed to have enough low-income students. Still, the program was initially sold a means to reduce the impact of *poverty* on student learning.

FIGURE 1 TOTAL SAGE EXPENDITURES



Source: Wisconsin Department of Public Instruction

Now the SAGE program has become a class size reduction program that is no longer restricted to high-poverty schools. Representative John Gard, co-chairman of the Joint Finance Committee, appropriately remarked that these changes take away SAGE's emphasis on benefiting low-income students.[19] Although distribution of available funds is still based on consideration of the number of low-income students in schools and districts, efforts at targeting high poverty students are diminishing. The full consequence of dropping the low-income requirements will be even more apparent, when it is explained later in this report how both national and local evidence on the effects of class size reductions demonstrate that smaller classes have little or no effect on middle- and high-income students.

THE PROSPECTS FOR THE SAGE PROGRAM IN THE FUTURE

The process is already in motion to determine how SAGE will be modified in the upcoming budget cycle. It is becoming apparent that when the original SAGE schools reach their "sunset date," these schools will experience a

swift renewal of their contracts. In the Wisconsin State Senate's recent "mini-budget," there was a provision explicitly authorizing the DPI to renew contracts with SAGE schools who see their five-year term end. Likewise, Democratic members of the Wisconsin Senate are already in discussion over eliminating, in one form or another, the contractual nature of the program. All these developments indicate that no schools will be dropped from the program, only added. Given the public endorsement for these smaller classes, and the well-publicized (yet uncritically inspected) findings from the UWM program evaluations, it is very likely that the sun will never set on any of these five-year contracts, and the program will eventually morph into a standard, state-aided, education program.

Funding for SAGE will also continue to increase, barring some dramatic turn of events. Funding for SAGE rises for two primary reasons: (1) the number of grade levels being exposed to class size reduction requirements increases, or (2) more schools enter into SAGE contracts.

According to estimations made by DPI officials, the potential breakdown of program costs over the next few years can be calculated. In the 2000-01 school year, approximately $58 million will be spent on SAGE. This amount includes all the funding needed to reduce class sizes to 15 in grades K-3 (all of the grades for which reductions are required under the SAGE contract) at schools in both the first and second round of SAGE contracts. It also includes funds going to the approximately 525 schools participating in the new third round of SAGE contracts, but for class size reductions in these schools for grades K-1 only. The additional aid required to expand SAGE in these new schools to the second grade in 2001-02 and third grade in 2002-03 will be roughly $18 or $19 million per year. But, of course, this process is cumulative, so that by 2002-03 the total additional cost of the SAGE program would be approximately $37 million, for an annual total of $95 million for all schools in the program and all participating grades in those schools.

The DPI has also made preliminary estimations of how much it would cost to expand the SAGE program to every school containing grade levels of at least K-3.[20] After the third round of SAGE contracts, approximately 500 more schools would have to be added to cover all of Wisconsin's approximately 1,100 public elementary schools under SAGE. The costs of adding these schools would be approximately $15 to $18 million in the first year of SAGE implementation (grades K-1) and between $8 and $9 million per year in the following two years, as grades 2 and 3 are incorporated into the smaller classes.[21]

In total, given the preceding estimations and current SAGE costs, the state of Wisconsin will have to spend somewhere between $126 and $131 million a year to achieve the SAGE class size reductions in grades K-3 in all of the state's public elementary schools. These funds are independent of any aid received under the federal government's class size reduction program, which in 2000-01 awarded Wisconsin around $21 million in aid for class size reductions.

At that point, if every elementary school in the state were admitted into the program, the annual amount of funds needed should hit a ceiling. This, of course, assumes a number of things. *First*, it assumes that more grades will not be exposed to the smaller class size requirements, which otherwise would cause a proportionate increase in costs related to hiring new teachers and constructing more classroom space. *Second*, it also assumes a static number of elementary schools in the state. *Third*, the total assumes that the preceding DPI estimates are accurate, while only actual implementation will confirm those funding predictions. A *final*, yet very important assumption is that the aid per low-income pupil remain at $2,000. In its proposal for the 1999-2001 budget, the DPI originally put $2,000 at the low end of its $2,000 to $3,000 estimate of how much was to be needed each year, per-pupil, to expand the program to grades 2 and 3. Therefore, it seems there is an inclination to raise the per-pupil aid amount, which would cause all costs to increase proportionally.

SUPPLEMENTAL CLASS SIZE REDUCTION EFFORTS IN WISCONSIN

As alluded to earlier, in addition to the SAGE program that emanates from the state government, local school districts in Wisconsin have other options to fund and implement class size reductions in their schools. These options primarily include (1) federal government grants under the Federal Class Size Reduction program; and (2) programs created directly by local districts in Wisconsin. In many cases, these non-SAGE efforts are being used by districts in conjunction with their limited participation in the SAGE program, filling in the holes that SAGE does not cover.

The Federal Government's Helping Hand

Given the widespread popularity of class size reduction policies, it is not surprising that the Clinton administration has joined prominently in the movement. As part of the 1999 federal budget agreement, the US Congress autho-

rized the spending of $1.2 billion for states to hire 30,000 new teachers under the US Department of Education's new Class Size Reduction Program (CSR). This action was precipitated by a threat from President Clinton to veto the entire federal budget if these funds were omitted. The amount was considered a "down payment" on an effort to reduce class size in early grades to 18 or fewer students per class. The process continued with an additional $1.4 billion in fiscal year 2000 (school year 2000-01).

In Wisconsin, these federal aids are playing a significant role. Under the CSR program, the state of Wisconsin was allocated approximately $21.8 million in fiscal year 2000 to be spent by local districts on class size reduction. In the previous year — the first year of the federal program — Wisconsin received $20.1 million in federal CSR funds, of which 30% went to MPS. Original estimates of aid to some other areas include: Dane County ($925,000), Racine County ($741,000), Brown County ($641,000), Rock County ($601,000), Kenosha County ($564,000), and Waukesha County ($542,000).[22]

Participation In The Federal CSR Program

To receive federal CSR dollars, school districts apply directly through their state for portions of the state's total grant as part of the district's Title VI applications.[23] The formula for allocating grants to school districts works to target high-poverty communities, as they are distributed based 80 percent on poverty rates and 20 percent on general school enrollment numbers. As it is, nearly all districts in Wisconsin receive awards of some amount.

Each local school district may use the funds received to either (1) recruit, hire, and train fully qualified teachers who can then be used to reduce class sizes in grades K-3 to 18 or fewer students, or (2) to engage in professional development and testing of new teachers. Districts that have already reduced class sizes in the early grades may use these federal funds to (1) make further reductions in grades K-3; (2) to reduce class sizes in other grades; or (3) to carry out activities to improve teacher quality and professional development. The law restricts districts from using these funds to augment salaries and compensation of teachers who are not hired under the program.

Although primarily emphasizing class size reduction, the federal program puts an additional emphasis on improving teacher quality. The law specifies at various points that any teachers used to reduce class sizes must be "fully qualified teachers who are certified within the State." In addition, up to 25 percent of funds each district receives may be used for professional development of teachers, either in the form of testing the academic content knowledge of new teachers, or for more general activities that "meet the goal of ensuring that all instructional staff have subject matter knowledge, teaching knowledge, and teaching skills necessary to teach effectively in the content area or areas in which they provide instruction." Finally, as part of the Education Flexibility Partnership Act, qualifying districts can apply for a waiver to spend more than 25 percent of their state CSR funds on professional development.

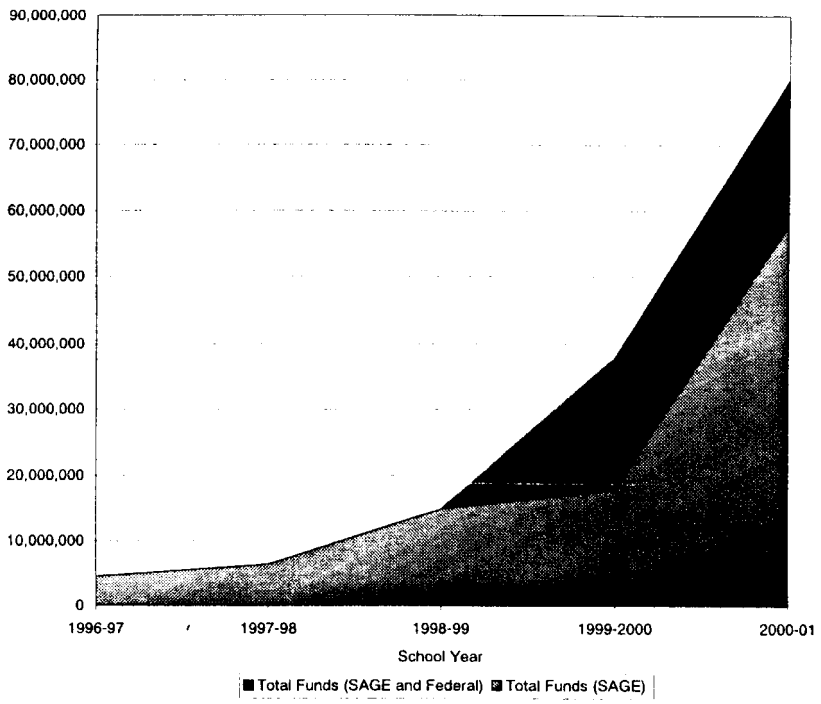
TOTAL EXPENDITURES ON CLASS SIZE REDUCTION EFFORTS IN WISCONSIN

Class size reduction is clearly a policy whose time has come in the minds of many. At least that is the inference to be drawn by the continuing expansion of SAGE and other class size reduction policies in Wisconsin, and the concomitant expansion of public expenditures to met these reduction goals. In many ways, the mathematics of class size reduction and public funding is simple. When either more grades or more schools are engaged in reducing class size to 15 students per teacher, the costs of this endeavor will rise proportionally to the number of new teachers who will have to be hired to meet this demand.

Figure 2 shows the dramatic expansion in government spending on class size reduction efforts in Wisconsin. Public schools in the state will be spending a total of approximately $80 million dollars during the 2000-01 school year to reduce class sizes, mostly in the early primary grades. Of that amount, roughly $58 million is being generated from state-appropriated SAGE funds, with the remainder coming from the federal government. These totals do not include funds spent by local districts for class size reductions that come from their own discretionary funds. Only three years earlier, when the concept of SAGE as only a pilot program remained, and before the federal CSR program, total annual funding on class size reduction in Wisconsin was only $6.9 million. Assuming federal funding remains at its current level, if SAGE funding continues to expand to all public elementary schools in the state, the likely costs of class size reduction policies in Wisconsin, both state and federal, will be between $147 and $152 million per year.

In addition, the SAGE program, as with the federal CSR program, has moved further away from being aimed at aiding low-incomes students in the early primary grades, to now being simply a means of reducing class sizes across

FIGURE 2 CLASS SIZE REDUCTION SPENDING IN WISCONSIN



the board — in more grades and for students who are not poor. The trouble is that it is precisely these types of changes that research, even that conducted by proponents of further class size reductions, has consistently shown has little or no academic effect. Currently, smaller class sizes are so popular that the cautious manner with which the programs were initially advertised has been completely abandoned and the legitimacy of calling for dramatic class size reductions on a wholesale level has begun.

These costs appear staggering. But it is up to the public and their elected representatives to determine if this continual amount of funding is worthwhile. To make that calculus, though, the general public and elected officials need accurate information as to what precisely are the results that smaller classes bring about. At this time, the public support for smaller classes is not based on a full and fair representation of the evidence on class size and its effect on student achievement. Therefore, the remainder of this report explores the effects of smaller classes.

DETERMINING THE BENEFITS OF INVESTING IN CLASS SIZE REDUCTION: MOVING TOWARDS A MORE OBJECTIVE APPROACH TO THE EVIDENCE ON SMALLER CLASS SIZES

It is evident that class size reduction policies are proliferating and will seemingly continue to do so in the near future. To this point, this report has covered the costs associated with these movements, especially as they relate to Wisconsin. The natural questions to now ask are in what way and to how great of an extent do smaller classes create beneficial outcomes, of which increasing student achievement is the primary goal? More specifically, is reducing class size a cost-effective way of increasing student achievement?

Class size reduction is an educational policy option that competes with other education reforms, both in terms of political attention and public funding. Whatever funds are utilized for class size reduction are funds that may not be used for other educational initiatives and processes. To better allocate our educational resources, therefore, requires a thorough and honest look at the evidence available on if, and how, smaller classes improve learning.

Three forms of evidence on the effect of smaller classes are presented below. First is a summary of some of the existing research on class size from across the United States. Both the favorable and unfavorable results on the apparent wisdom of smaller classes will be explained. Together, these competing views will show that a much larger amount of disagreement exists over the merits of smaller classes than is commonly presented.

Second, there will be a close inspection of the findings of the SAGE program evaluations, which have been conducted in conjunction with the program's first few years of existence. Some of the results of these evaluations are well known and have been cited to support the dramatic expansions of SAGE described earlier. But many important findings have been less well disseminated or have been outright suppressed. A new attention to these findings will cast doubt on the wisdom of indiscriminate class size reductions in Wisconsin.

Finally, the benefit of smaller class sizes will be compared to other available programs and uses of educational resources, particularly that of improving teacher quality, which will help determine whether funds used in programs like SAGE could be more efficiently spent on other aspects of public education.

But first, a brief discussion will be made regarding the very significant role political interests have played in the debate over class size. This understanding is important for two reasons. First, it helps explain why those who disagree with the seemingly inherent benefit of smaller classes, or even those who simply express doubts, have a difficult time being heard over the reverberations of support coming from all the levels of public education and government.

Second, an understanding of the political dynamics of class size advocacy shows precisely how important questions never get asked, much less answered, about the effectiveness of smaller classes in specific contexts and about the reform's position relative to other interventions aimed at increasing student achievement.

POLITICAL SUPPORT FOR SMALLER CLASS SIZES

To question the efficiency or desirability of smaller class sizes, particularly in Wisconsin, is seemingly an exercise in futility. It is difficult to dent the political weight found in support of reduction initiatives. Not only is the issue fought for vigorously by national, state, and local teacher unions, it is also well-promoted by government agencies, such as the US Department of Education and the Wisconsin Department of Public Instruction.

At the national level, the US Department of Education has picked up the banner of class size reduction in a less-than critical fashion. According to a 1998 report, *Reducing Class Size: What Do We Know*, the department concludes, "Overall ...the pattern of research points more and more clearly toward the beneficial effects of reducing class size."[24] This statement is made despite the report's acknowledgment that many well-recognized scholars in the field of education policy and finance recommend that a high degree of caution must be applied when considering class size reduction policies.

Closer to home, the Wisconsin Education Association Council, Wisconsin's state-level teachers union, has lobbied diligently to advance class size reduction efforts in Wisconsin, largely through the SAGE program. The WEAC website dedicates a whole section to puffing the SAGE program.[25] In the organization's latest resolutions, the union states that it "believes that excellence in the classroom can best be attained by small class size," and that the Council believes in "an optimum class size of fifteen students in regular programs." But WEAC's resolution is simply a verbatim regurgitation of the National Education Association's resolution on class size, and offers little by way of explanation.

This past year, as part of WEAC's new television promotions, the group has sponsored commercials that praise the class size reduction movement in the state, stating that smaller classes are helping all students learn more. The commercial omits the fact that smaller class are only being implemented in the early primary grades, leaving a subtle impression that smaller classes are, or at least should be, implemented in all grades. Such a view comports with its non-specific resolution on class sizes of 15 students and the WEAC mantra that "SAGE schools are great schools." Speaking in such generalities is a safe-haven for class size reduction advocates, and, so far, the public has been uncritical of these vague comments.

Yet by failing to separate blanket advocacy from critical analysis, these parties fail to inform the public about the intricacies of class size, both as a singular issue, and also its place in the education policy matrix. Nowhere is this last point more evident than the nearly complete avoidance of looking at class size policy through the metric of cost-effectiveness. Then again, smaller classes very likely make teaching easier, and it *is* the duty of their union to argue for whatever makes the jobs of teachers easier, regardless of any actual achievement effects and public costs.

In this vein, WEAC has remained highly incredulous of the evidence opposed to class size reduction. In fact, the union admits to a prejudice over the merit of class size reduction. According to its 1997 legislative talking points, "SAGE presents an excellent opportunity *to prove* that lower class sizes *will* increase pupil achievement." (emphasis added). The fact that this comment came out *before* the first UW-Milwaukee evaluation of SAGE's impact had even been completed is revealing. Evidently there never was or is any question for WEAC if class size reduction is beneficial. It simply must be.

WEAC's lobbying efforts have been highly effective. Attempts to limit the funding of SAGE expansion in the 1997-99 budget were quickly discarded due to significant pressure by various parties within the state's public education system. Indeed, legislative support for smaller class sizes has been very deferential to the rhetoric supplied in support of SAGE. There is no significant segment opposing the unabated expansion of class size reductions, which could counter the oft-exaggerated claims coming out of WEAC and other advocates of SAGE.

Furthermore, politicians and school board members who doubt the impact of smaller classes have yet to equip themselves with a tenable argument against smaller class policies. The same cannot be said of their counterparts who laud class size reductions. Smaller classes give a tangible form to the more common argument for increased spending on public education. As such, it is a golden egg to teachers' unions and other organized interest groups, who have been presenting this general argument for increased funding for decades. By having a readily identifiable, and popular, use to the funds, the argument carries much greater force than in the past. This reality explains why class size reduction policies are being vigorously advocated, and will continue to be so treated.

Overall, supporters of smaller classes, particularly at the organized level of teachers' unions and the DPI, have successfully, but erroneously, painted a picture that smaller classes obviously help all students to a significant extent, and that to be against smaller classes is to be against the educational needs of children. Therefore, one can add public emotion to conventional wisdom as another hurdle confronting anyone who desires greater inspection and intellectual rigor being applied to the issue. In doing so, these interests succeed in preventing serious discussions of the costs as well as the real benefits of smaller classes. Acquiescing to this mentality is shortsighted on the part of elected officials, since they are being encouraged to spend more funds than they may have to in order to achieve gains in student achievement.

IN SUPPORT OF SMALLER CLASSES: HOW SMALLER CLASSES ARE CLAIMED TO HELP

The notion of reducing class sizes is not a novel one. For a long time there have been questions as to what effect the number of students a teacher faces has on enabling student learning.

Going along with this sentiment, student-teacher ratios and class sizes in the United States have, in general, decreased steadily during the last century. Some of the decline in student-teacher ratios is the result of the increasing number of teaching positions in non-classroom instructors, such as in special education. In any event, according to one estimate, between 1955 and 1990, the average class size (*not* simply pupil-teacher ratio) declined from 30 to 20 students,[26] while another figure puts the drop from 30 students per class in 1961 to 23 in 1998.[27]

Many laud this trend and believe it should continue. They point to the various direct and indirect benefits that accrue from educational settings characterized by smaller classes. The following are some of the commonly offered benefits said to arise from implementing smaller classes, usually of a size between 15 and 19 students.

- ***The creation of more opportunities for teachers to focus on the needs of individual students***, particularly with attention to struggling students. The notion is that some students get "lost" in larger classes and, to use another euphemism, "fall through the cracks." John Zahorik, one of the UWM SAGE evaluators, delved specifically into the issue of how smaller classes lead to individualized instruction. He found that in smaller classes teachers appear to individualize with students as soon as the school year begins; that all students — able and problem learners, and typical students — received comparable amounts of attention; and that more content is covered during the year under this style.[28] The nostrum of individualization is probably the aspect of smaller classes that receives the greatest rhetoric.
- ***The development of more opportunities for students to actively participate***. Students are said to be less intimidated when surrounded by fewer classmates and are, therefore, more willing to be involved in class. Increased student engagement in the classroom is then assumed to lead to increased student learning.
- ***The occurrence of fewer problems with classroom management, including discipline troubles***. Smaller classes allow teachers to oversee and eliminate misbehavior more effectively, and gives students less opportunities (in the form of fellow classmates) to be mischievous. This reality then allows for more time spent on instruction and less time enforcing discipline.
- ***The development of higher morale among teachers and students***. This increased morale is part and parcel of what is frequently described as a more family-like atmosphere in the classroom. It also relates to secondary effects on improving teacher retention and recruitment, as the classroom environment is perceived to be more desirable and less stressful.
- ***The allowance for teachers to use innovative practices that they would be less likely to use in larger classes***. Since classroom-management stresses are decreased, the notion is that teachers have fewer risks facing them if they were to attempt teaching styles that deviate from traditional lecture styles.

Overall, the positive effects are generally said to be that teachers have a greater knowledge of their students and are able to spend less time on classroom management and discipline, and more time on instruction. According to Wisconsin State Superintendent John Benson, there are four revealed characteristics to smaller classes: (1) high levels of classroom efficiency; (2) a positive classroom atmosphere; (3) expanded learning opportunities; and (4) enthusiasm and achievement among both students and teachers.[29] While these outcomes may be important, the extent to which these characteristics of smaller classes translate into improved student performance remains the preeminent question.

The Tennessee STAR Experiment

The class size debate was intensified by the implementation of Tennessee's Project STAR (Student-Teacher Achievement Ratio). This program, considered the seminal experiment in class size reduction, was a four-year, longitudinal study of kindergarten through third grade classrooms in the state that began in 1985. STAR compared classes of 13-17 students with classes of 22-26 students, both with and without an additional instructional aide in the larger classes. Students who were randomly assigned to smaller classes stayed in them from grades kindergarten through third, were tested each year of the program, and had their performance tracked after they returned to larger classes.

The Project STAR studies reported that students in the smaller classes scored better on standardized tests than their counterparts in larger classes, and that advantage diminished or marginally improved in the years following their return to regular classes. Gains were also much larger for inner-city students (of which 97 percent were minority students) than for suburban, rural, or other urban students. The findings also showed that children in smaller classes were less likely to be retained in grade.[30]

The Tennessee STAR researchers have also released findings showing that students placed in smaller classes in the early grades continue to perform better academically through high school. These findings, located in the "Lasting Benefits Study," showed that students in smaller classes:

- Had higher test scores in math, reading, and science in grades four, six, and eight.
- Were less likely to drop out of school.
- Were more likely to graduate in the top 25% of their class.
- Were more likely to take college entrance exams (although not more likely to do better on those exams).[31]

It is nearly impossible to discuss the effects of class size reduction without referencing the STAR results and the conclusions generated from them. At least that is certainly the case for reduction promoters. STAR is the absolute boon to advocates of smaller classes. The American Federation of Teachers calls the STAR study "the "gold standard" of class size studies.[32] According to Frederick Mosteller, a Harvard statistics professor and one of the primary researchers of the program, "[STAR] definitively answers the question of whether reduction from this size to that size does make a difference, and clearly it does."

Yet STAR is simply not the definitive answer that it is always put forth as by reduction advocates. *The Peabody Journal of Education*, published out of Vanderbilt University, devoted an entire issue to the analysis of the STAR program, and its primary conclusions (summarized in Figure 3) are not as flattering as those commonly offered. Moreover, other analyses of the STAR findings have echoed the *Peabody* conclusions.

One of the main concerns with STAR is that achievement differentials between students in small and regular classes mostly occurred for students in only the first year, with the gap *not growing* during subsequent years of exposure to smaller classes.[33] This fact suggests just a one-time impact from smaller classes during the first year of a child's formal schooling. The STAR findings also offer no insight into the effects of reductions to a lesser degree, somewhere less than the one-third reduction down to 15 students done in STAR. Appropriately, Erik Hanushek concludes that "[t]his policy interpretation is quite different from that commonly attributed to the STAR analysis, which many cite when they wish to justify any sort of reduction in class size at any grade level."[34]

Additionally, there seems to be a lamentable level of exaggeration used while reporting the magnitude of the STAR achievement results. Repeatedly heard are claims that students in the small STAR classes outperformed students in regular-sized class by "significant" margins. But differences that are statistically "significant" to the mathematician can easily be insignificant in the common understanding of the term. This is the case with the STAR data, where the actual size of the effects ranges from minimal to small, depending on the test and subject.[35] On top of all this, what was known about the STAR results for many years was only based on that which had been disseminated by the project's own researchers, as the data were not made available to most other researchers for more critical analyses.

FIGURE 3 Lessons from Project STAR

1. The maximum effect of reducing class size is in kindergarten and first grade. The effect on achievement levels off and declines in second and third grade even when students remain in smaller classes.
2. The achievement advantage of small class students dropped about 50% the first year after they were back in regular-sized classes (21-28 students) in the fourth grade.
3. Class size reduction appears to be very expensive. The cost of reducing class size is proportional to the size of the reduction: i.e. a one-third reduction in class size will increase per pupil costs about one-third.
4. The high costs of substantial reduction in class size and the modest achievement gains that can be expected, even in kindergarten and first grade, suggest that less expensive targeted reductions should be tried.
5. The most important lesson may be that just changing class size without changing what is taught or how it is taught will probably have modest results, because the various factors all influence achievement.

Source: Folger, John. "Lessons from Class Size Policy and Research" Peabody Journal of Education, *Vol. 67, No. 1, Fall 1989.*

Despite these and other important reservations about the STAR results,[36] the program continues its role as the primary exemplar of smaller classes and as a source for defending the benefit of smaller classes. And based on the favorable presentation of the STAR evidence, other states have used the program as validation and a blueprint for their own programs, including Wisconsin's SAGE program.

Additional Research Supporting Smaller Class Sizes

Although the STAR evaluations have undoubtedly been the main source of evidence supporting smaller class size policies, a few other sources of evidence are put forth by reduction advocates. The most significant new addition is that of the SAGE program evaluation, which is discussed in much greater detail below. Some other studies exist that present conclusions favorable to the continued implementation and expansion of class size reduction policies. Unfortunately, many of these studies simply reiterate the same findings (usually from STAR), without the addition of any further independent research.

Before STAR and SAGE, supporters of smaller classes looking for empirical support stayed afloat largely on one study from 1978. This study was an analysis by Gene Glass and Mary Smith, which suggested that class size reductions produced greater achievement gains than regular-sized classes.[37] This study is now well-disregarded for a variety of reasons, mostly because the studies used in the meta-analysis were flawed and not representative of what would be deemed the standard conception of what are "smaller classes." For example, one of the "classes" used by Glass and Smith was a gym class testing the ability of students to bounce tennis balls off a wall. Moreover, the "larger reductions" they referred to are what would commonly be called tutoring — classes of five students. These are not the types of class sizes imagined by current discussions on class size policy.

The Policy Information Center, a division of the Educational Testing Service, published a study in 1997 titled *When Money Matters*, which covered 10,000 fourth graders in 203 school districts, and 10,000 eighth graders in 182 districts from across the country, and looked at the effects of class size.[38] The report defined small classes as those of less than 20 students, and large classes as those with any greater number of students. The research looked at the impact of class sizes on National Assessment of Educational Progress (NAEP) mathematics scores, controlling for other factors, such as socio-economic status, educational expenditures, and regional cost of living. The results showed that students in smaller classes did better, but that gains were larger for fourth-graders than for eighth-graders. Fourth-graders in smaller classes were said to progress 33 percent more quickly than in larger classes; while eighth-graders were expected to progress 12 percent quicker. Gains were also much greater for inner-city students than any other group. This concurs with other studies which argue that smaller classes are an effective strategy for reducing the white-minority achievement gap.[39]

When it comes to evidence on the positive effects of smaller classes, the preceding studies give an accurate picture of about all one will find. Frequently, reduction advocates exaggerate or playfully word the findings from these studies to make the gains seem higher and broader than they actually are. Many dance around the specifics and use language suggesting that smaller classes have much greater positive effects in all settings and for all students. In any event, based on but a few studies, and overstatements as to the effects of STAR and SAGE, advocates of smaller class sizes now claim that smaller classes unequivocally mean higher levels of student achievement.

CRITICAL ANALYSES OF SMALLER CLASSES: THE UNHEARD STORY

Given the preceding volume of support for smaller classes, one may naturally question how there could be any disagreement over the policy. This is especially true when conventional wisdom sides with the apparent merit of smaller classes. Yet a vigorous debate over the effectiveness and merit of class size reduction in K-12 schools has been carried on for most of the past century.

It is clear that until the STAR and SAGE studies, the academic research on the effects of class size and teacher-pupil ratios was at best ambivalent, and at worst, suggestive that the funds needed to achieve smaller classes could be better spent on other programs aimed at increasing student achievement, especially for low-income students. We have already seen some of the contentions with the STAR findings. A look at the research outlined below will further show both the greater depth of this issue, and the legitimate claims that remain undiscovered by a simple deference to the conventional wisdom of smaller classes.

A sampling of some noteworthy studies shows:

- The National Conference of State Legislatures, a non-partisan organization that provides information to all 50 state legislatures, concluded that "[a]lthough over 1,100 studies examine the relationship between class size and student achievement, no definitive conclusions have been reached. While positive results have been demonstrated in Tennessee and Wisconsin, other research finds little connection between student-teacher ratios and student performance, especially when measured against other types of educational reforms."[40]
- Looking at trends involving student performance, econometric evidence, international comparisons, and analysis of state-level data, Eric Hanushek, Professor of Economics and Public Policy at the University of Rochester, concluded: "Existing evidence indicates that achievement for the typical student will be unaffected by instituting the types of class size reductions that have been recently proposed or undertaken. The most noticeable feature of policies to reduce overall class sizes will be a dramatic increase in the costs of schooling, an increase unaccompanied by achievement gains."[41]
- In another report, Hanushek reviewed 277 studies that examined the impact of student-teacher ratios on learning and found that: (1) only 15 percent of the studies showed that a lower ratio caused a significantly positive impact on performance; (2) 13 percent of the studies actually showed a negative effect; and (3) the remaining 72 percent yielded no conclusive results. In addition, many Asian countries often have vastly larger class sizes of often 40 to 50 students per teacher, and yet these students frequently outperform US students on international achievement assessments.[42] While it is always prudent to be weary of international comparisons, especially in the area of education, this fact may suggest that there are underlying factors in our education system, besides class sizes, which are of greater importance.
- Robert Slavin, Director of the Center for Research on the Education of Students Placed At-Risk and Professor at Johns Hopkins University, explains that when only valid studies on the effects of class size are analyzed, the evidence shows that "substantial reductions in class size do generally have a positive effect on student achievement, but the effects tend to be small." The reductions he analyzed were from an average of 27 students down to 16 — a 40% reduction — yet still the gains were meager. He also stated that teachers may change their behavior while teaching in a smaller class, but it is usually only in subtle and insignificant ways.[43]
- Glen Robinson, former President and Director of Research at the Educational Research Service, a non-profit organization that provides objective research and information on education issues, performed a similar meta-analysis of the class size research and stated that "research does not support the expectation that smaller classes will of themselves result in greater academic gains for students."[44]
- Allan Odden, Professor of Educational Administration at the University of Wisconsin-Madison, has stated that smaller classes should be used "sparingly and strategically," and that there are more cost-effective means available to achieve the results of smaller classes, without requiring large amounts of new funds.[45] Odden reviewed data on programs in Tennessee and Indiana, and concludes that these studies show "that new and costly state programs that reduce class size to under 20 students do not produce very large gains in student performance."[46]
- Researchers at the RAND Corporation recently estimated the costs of different types of national class size reduction policies and made comparisons of these costs to other education programs. The researchers concluded, in part, that "the high monetary costs and probable implementation problems associated with a

national CSR program suggest a reconsideration of its likely benefits," and that reducing class size to 15 students costs twice as much as reductions down to 18.[47]

- Dr. Kirk Johnson, an analyst for The Heritage Center for Policy Analysis, a Washington DC-based think tank, performed a statistical comparison of performance in 1998 NAEPreading achievement scores between smaller classes (less than twenty students) and larger classes (greater than 30 students). After controlling for income, family background and other demographics, he found that fourth- and eighth-grade students in the smaller classes did no better than students in larger classes.[48]

Overall, these examples draw attention to the questionable assumptions supporting smaller classes, and, if heeded, could further a more constructive dialogue on the topic. They point to legitimate concerns about the extraordinarily high costs of creating smaller classes and the subsequent meager level of improvement they generate. Certainly, there are other studies that have concluded in favor of the positive effects of smaller classes, as were discussed earlier in this report. But even the results of these studies must be tempered by a disciplined look at what they actually do show, as opposed to what they just purported to conclude.

Explaining The Doubts Over The Effectiveness Of Smaller Classes

What is driving these ambiguous conclusions on smaller classes, especially when intuition tends to dictate that smaller classes should be very beneficial? The answer lies at a couple of levels.

First, and most generally, one recognizes that making classes smaller does not *by itself* improve student learning. Even proponents of reductions explain the process in terms of how the smaller classes create conditions that enable students and teachers to interact more productively, *thereby* improving the students'learning experiences and achievement. The main pedagogical issue therefore is how class sizes, on the margin, truly affect relevant learning activities within a classroom.

The evidence is clear that interactions of size of around 1-to-1 or 1-to-3 of teachers to students — what normally is referred to as tutoring — works wondrously in helping students learn. But such a "class size" is a logistical and fiscal impossibility and is not what is at issue in the current conversations of class size policy. What is being talked about is a reduction from roughly 23-25 students (27-30 in MPS) down to 15 students. But what does *that* reduction mean for student learning?

The answer is apparently not that much, for the following reasons:

- ***Most teachers do not teach very differently in class sizes of around 15 students as opposed to larger classes.***

  A common concern expressed among researchers who have looked at the pedagogical impact of smaller classes, is whether teachers really change their manner of teaching. Slavin's research suggested that most teachers in smaller classes rarely change their behavior. According to Zahorik's research, while small classes often have teachers who individualize instruction, they do not individualize the content of what is taught. He concludes that the individualization that occurs in small classes is more procedural than substantive, and that direct instruction remained the dominant mode of teaching.[49] The SAGE evaluation itself states that teachers in SAGE classrooms still use direct instruction as the dominant mode of interacting with students. They simply perform the same style of instruction, only with the added benefit of having fewer students to oversee, and fewer tests and homework assignments to assess.

- ***Sometimes lower-quality teachers are used to fill the need of smaller classes.***

  A crucial assumption of any class size reduction program is that the teachers hired to achieve the new reductions will be of high quality, or at least an adequate level of ability. This assumption was dramatically proven wrong in California during that state's implementation of its Class Size Reduction program (CSR). California's program requires all elementary schools to reduce class sizes to 19 or less in grades K-3. The demands this mandate put on finding new teachers was substantial. One of the most distressing side-effects of the program was that many of the higher-quality teachers left at-risk schools to fill the new openings in suburban schools, leaving the low-achieving urban districts with teachers who were often the least experienced and least qualified. According to a California Legislative Analyst's Office report, 24 percent of teachers hired for CSR did not have a teaching credential, and an additional three percent of new CSR teachers were hired on waivers that require even *less* in terms of qualifications.[50] The impact of this class size reduction policy was to expose more students to lower quality instruction, which simply offset any gains from the smaller classes.

- ***Based on the STAR and SAGE evaluation data, there is persistent evidence that smaller classes only have an meaningful impact in a student's first year of exposure to formal schooling (kindergarten), and pos - sibly the second (first grade).***

  Simply put, the effects of smaller classes in later grades is minimal, while the costs remain large. For advocates of class size reductions, that smaller classes are ineffective in later grades should be inherently odd. If smaller classes are good in and of themselves, they should be equally effective at any grade level, with possibly some minor variations across settings.

  Yet there is a reason why this finding could make intuitive sense, if one alters his perception of how it is precisely that smaller classes help students. In kindergarten or the first grade, a child is first becoming acclimated to the process of formal schooling and what that experience entails. In a classroom setting with fewer students, the ability of a student to make this adjustment is heightened, as any negative effects of being surrounded by more students are diminished, particularly if these other students, who are themselves becoming acclimated to schooling, are not making the adjustment well. In a sense, what occurs is a class- room environment peer effect.

  The perspective of students may, therefore, better explain why the impact of smaller classes in subse- quent grades has no effect. Once children are able to learn what it means to be a student, and what this whole "school stuff" is all about, any marginal changes of class size from 15 to 20 to 25 students in later grades makes less of a difference to the students' ability to learn. Therefore, smaller classes aid in a student's accli- mation to formal school in their first year, but after that the pedagogical effect is insignificant.

- ***The non-achievement effects witnessed by teachers and students in smaller classes are not translated into significant improvement in academic achievement.***

  Not to minimize the impact of smaller classes on improving teacher morale, student discipline or feel- ings of individualized treatment, but some researchers have questioned the actual connection between these results and increasing student achievement. The issue is not just making teaching easier, but also having a genuine impact on student learning.

In addition to these pedagogical reasons for smaller classes not having a significant impact on student learning, there are other concerns with smaller class size policies that relate to the resources needed to achieve smaller class- es. Two of the most prominent of these concerns are:

- ***Besides personnel investments for new teachers, school facility capacities will eventually have to be expanded, which can be very costly.***

  In general, smaller classes require more classrooms, and building issues in public schools are already a highly contentious matter in this state, as evidenced by the inability of many districts to pass building ref- erendums. This problem will, of course, be more or less acute depending on the existing building capacities in school districts, and on the rate of increase or decrease in the number of students within a district.

- ***Costs used to achieve smaller class sizes are better spent on improving teacher skills and competencies, which are shown to have a greater impact.***

  As noted earlier, researchers frequently point out that the achievement gains realized from smaller classes are much smaller than other educational intervention programs, yet at the same time cost much more than these programs. This matter is discussed in greater detail near the end of this report, using matters of improving teacher skills as a specific example.

The main point is that many factors besides simply making class sizes smaller work to determine whether small- er class sizes will have a large impact. There remain strong reasons to be skeptical about the impact of smaller class- es, of a magnitude between 15 and 19 students, in all grades. These reasons help explain why the assumed positive impact of smaller classes on student learning is not as large as is generally anticipated.

THE DEVIL IN THE DETAILS: (RE)EVALUATING WISCONSIN'S EXPERIMENT IN CLASS SIZE REDUCTION

Although simple deference to popular belief dominates most discussions on SAGE, there nonetheless remains the important task of verifying whether class size reductions actually achieve the goal of improving student learning.

Like most education policies, class size reduction programs exist to improve student learning, and as such, stu- dent performance is the primary measure against which the program's success should be evaluated. This is not to

impugn the other benefits that may result from smaller classes, such as greater discipline, increased student participation, and more individualized attention. These outcomes carry merit in and of themselves, whether or not student achievement may increase. Still, even these measures are often understood as conditions that aid in learning, which may eventually cause higher student achievement.

In any event, increased student performance is one of the primary results that is advertised and expected to occur from SAGE class size reductions. And this outcome measure is even more important given the enormous costs necessary to realize and maintain class size reductions. Spending $100 million or more a year to enable classrooms to have more of a family-like atmosphere, without any significant increase in student achievement, will simply not pass the test.

The Importance Of Evaluating SAGE

As originally a pilot program, it was understood from the beginning that the evaluation of SAGE in its initial schools would be central to whether the program would continue or expand. If the program could show beneficial results, the argument for expansion would gain more resonance, while ambivalent results or, more specifically, results showing no differences between smaller and regular-sized classes, would argue against the continuance of class size reductions, at least on the basis of improved student achievement.

Wisconsin's SAGE program has two primary means of program evaluation. The first entails an internal review by state government officials to ensure that schools receiving SAGE funds are complying with the program's contractual mandates. The second evaluation is meant as an external review of the complete program and its comparative gains. This analysis is completed by an evaluation team at the University of Wisconsin-Milwaukee School of Education. The UWM evaluation is significant because it supposedly answers in a quantitative and qualitative fashion just how the SAGE program affects student learning. The UWM evaluation, and its resulting reports and interpretations, will receive the greatest level of attention in this report's review. This emphasis is made primarily because the UWM evaluation is currently the bulwark that protects the SAGE program from any substantive critique.

In discussing each of the evaluation processes, I will outline the general contours of what the SAGE evaluation entails by looking at what the results purport to show and, more importantly, by also examining findings that have been less well disseminated to the public. It is these less publicized issues with the SAGE program that require much greater attention as SAGE continues to move further into becoming a broad-based and continual class size reduction program.

Internal Review

School districts participating in the SAGE program must pass an annual review to continue receiving funding. According to the SAGE law, if a school is deemed to have failed to fully implement any of the four contract requirements, the DPI may terminate the contract.[51]

Despite the establishment of this internal review process, not a single school that has entered into a SAGE contract has found its funding terminated for non-compliance. Likewise, no district has even been approached by DPI with concerns over the level of effort a school has exhibited regarding contract performance. At the start of the 2000-01 school year, thirty schools will be in their fifth year of SAGE implantation, with 48 more entering their third year. The absence of any sanctions or even threat of sanctions for non-compliance during this time implies one of two things: (1) all schools have performed well enough to be in compliance, or (2) non-compliance has been dealt with by means less drastic than actual contract termination.

If the situation is truly of the first nature — that all schools are adequately complying with SAGE program requirements — then the low frequency of non-compliance cases is a fact to revel in and continue. According to the DPI, this is the reason for the lack of any sanctions being needed.[52]

Yet it also appears that the threat of sanction is not serious, as some evidence suggests that full compliance has not been achieved. For example, in the first year of SAGE implementation, some of the requirements of the contract were not even close to being complied with. According to the 1996-97 UWM evaluation, "during the first year of the SAGE program the primary focus of participating schools has clearly been on implementing the reduced student-teacher ratio. The other SAGE interventions…have been attended to by SAGE schools in varying degrees."[53] In fact, after the first year, approximately 60 percent of teachers in SAGE schools claimed they had no "personal, formal, written development plan" to satisfy the professional development requirement of the SAGE contract.[54] Similar apathy was prevalent towards satisfying the lighted schoolhouse and rigorous curriculum requirements, with most school

personnel claiming that they were simply carrying on previously existing strategies with regards to those areas. These trends have continued into the more recent years of the SAGE program.

Moreover, determinations of compliance are based largely on self-reporting by teachers and principals in SAGE schools and involve the answering of generalized questions. Unlike the class size reduction requirement, which by its very nature is clear and assessable, there is no semblance of objective criteria from which to measure success in implementing the three other contract requirements. Certainly, the "sufficient progress" standard from which the committee is to determine compliance necessitates a great deal of discretion. But that standard, or simply the effectual implementation of that standard, may not be requiring that much.

Even the class size reduction requirement has not been strictly followed. In 1998-99, 24 SAGE classrooms had more than 16 students, with a number of others having exactly 16 students, which in both cases are greater than allowed by the contract.[55] Numbers on specific SAGE classroom sizes were not presented in the previous two evaluations, and, therefore, it is uncertain as to whether this noncompliance also occurred in earlier years.

All in all, the DPI's annual review of the SAGE program, and the progress of SAGE schools in meeting all four program requirements, can be best described as passive and characteristic of great deference to the self-reporting of SAGE schools.

External Review: The UWM School Of Education SAGE Evaluation

Since its inception, the SAGE law has mandated that the program be annually evaluated to determine the extent to which implementation is proceeding, in what manner schools are going about achieving their class size reduction requirements, and, most importantly, to determine the effectiveness of the program in improving academic achievement.

This evaluation process is currently being conducted by a team of researchers at the University of Wisconsin-Milwaukee. In the program's first few years, the evaluation was housed within UWM's Center for Urban Initiatives and Research, but it was subsequently moved to the School of Education, largely to improve the evaluation's administration and accounting. The evaluation team officially works under contract with the Wisconsin Department of Public Instruction, and DPI staff worked hand-in-hand with the evaluation team in the initial development of the SAGE evaluation design and continue to be intimately involved.

The first evaluation report was released in December 1997, with two subsequent evaluations released in December of each of the following years. The fourth evaluation will be released sometime around December 2000. Each report summarizes the findings on SAGE results for the previous academic calendar year, so, for example, the first annual evaluation was based on SAGE implementation in the 1996-97 school year. Each report has focused primarily on identifying what differences, if any, exist in student achievement between SAGE schools and comparison schools.

Evaluation Design

On a general level, the SAGE evaluation design mirrors the earlier STAR evaluation design in that there exists a "treatment" group (those classes with smaller student-teacher ratios) and a comparison or "control" group (those classes in which the program and its class size reduction provision are not implemented — larger, regular-sized classes in other words).

The primary benefit of this design is that any differential outcomes between the two groups can be interpreted as solely due to exposure to the intervention, which in this case is the SAGE program, generally, and smaller classes more specifically. If all other relevant variables can be shown to randomly apply equally to both groups, with the exception of involvement in SAGE, any differential results are to be inferred as caused by SAGE involvement.

The tests used to evaluate student achievement between SAGE and comparison students are the Comprehensive Test of Basic Skills (CTBS) Complete Battery, Terra Nova edition. The CTBS is a standard test used for achievement measurement in early primary grades. The main benefit to using this test series is in its longitudinal nature, which allows for comparisons over time. This feature is particularly important to measuring the continual effects of the SAGE program from year to year. In addition, according to the evaluation team, the CTBS is beneficial because "it is very user friendly, and it calls for responses from students that are similar to those asked in classroom instruction."[56] Moreover, the test is said to be "one of the few instruments that attempts to minimize items biased against minorities and educationally disadvantaged students."[57]

For purposes of the evaluation, first-grade students are given the Level 10 version in October, and then the Level 11 version in May of the same school year.[58] The Level 10 test serves as a baseline measurement of student academic ability at the start of the first grade. Second grade students are given the Level 12 version in May of their second grade school year, while third grade students are examined with the Level 13 version. Although kindergarten students are also in smaller classes under SAGE, they are not tested.

The first SAGE evaluation analyzed program results from the 1996-97 school year, and during that year SAGE classrooms (those with 15:1 student to teacher ratio) were established in both kindergarten and the first-grade. It was the data on SAGE students and comparison students May 1997 first grade "post-test" results that were compared to the "pre-test" of those same students from October 1996.

For the second evaluation, these same students were followed up by their performance on the next level test at the end of the second grade. Only those students who took both the first grade pre-test and post-test, as well as the second grade post-test, were used in the 1997-98 second grade analysis. Also, a new batch of first graders were tested; these first graders had been in smaller classes as kindergartners the year before.

Finally, the third and most recent evaluation looked at the 1998-99 school year and tested the original cohort of students, who were now in third grade, along with second and first grade students, the latter of which were being tested for the first time by the evaluation team.

The students tested in all of the annual evaluations were only those who were in the original 30 SAGE schools; any students in schools that began participation in the SAGE program after the first year (1996-97) were not tested or used in the evaluation. Seventeen comparison schools were also tested each year, although the composition of this group varied across years.

Potential Problems Faced In Evaluating SAGE

Although the SAGE evaluation is for the most part well constructed, and employs a standard experimental design, there are some possible issues with the evaluation that arise from the program's nature.

First, the SAGE evaluation may fall prey to two related and common problems in social science experiments of this type — the Hawthorne effect and selection bias. The Hawthorne effect is when people act different precisely because they are part of an experiment. Of course many social programs explicitly have the purpose of prompting people to change behavior. The controlling question is, therefore, whether it is truly the program that changed the behavior (as designed to) or whether it is simply *knowledge of* the program, and in this case its political meaning, that causes behavior changes.

The SAGE program is obviously meant to cause teachers (and students) to act differently. But the attention that the program generates additionally makes teachers aware that they are part of a prominent educational reform, and, therefore, they will possibly be inclined to work harder to ensure it succeeds. This concern is accentuated by the highly political nature of SAGE and questions surrounding its continued funding. If SAGE becomes the standard as opposed to the exception, and the program loses its status as "experimental," these teachers need not carry the same level of diligence to ensure its continued existence.

Conversely, teachers in larger classes certainly do not have any added motivation. In fact, the less well they do, the more credence is given to the need for smaller classes — something that most teachers have already indicated they desire. Certainly, teachers in comparison schools have not tried to have their students do worse to "sabotage" the results. That is preposterous. The point is simply that they had a significantly smaller incentive to help their students score well than did their SAGE peers.

Moreover, there is evidence of a degree of disdain from schools selected to serve as SAGE comparisons. After the first year, the DPI and evaluation team were met with resistance from some comparison schools that wanted out of the hassles tied to their involvement. These schools had received nothing but a hearty word of thanks for their participation, even though the testing process required a great deal work by the schools. The DPI did not resist and allowed comparison schools to discontinue their status.

As a result, since the SAGE program began in 1996-97, seven of the original seventeen comparison schools have ended their role as comparison schools. Moreover, five of those seven applied and became SAGE schools themselves. The DPI could have easily stood its ground and told district leaders that if they wanted to have one or more SAGE

schools within their district, then they must offer one or more of the remaining elementary schools as comparison schools. Such a requirement would have been more than fair and would have maintained the integrity of the comparison group's composition.

This problem with the departure of comparison schools relates more generally to concerns of selection bias, which occurs when tested members of a study volunteer to participate because they have a vested interest in the program. The problem is particularly acute in SAGE, where schools apply to participate (and are publicly eager to do so) and where comparisons are made against schools that are forced, seemingly with contempt, into participating.

Another potential concern is that six of the current 16 comparison schools are in the MPS. Therefore, 38% of comparison schools are in the MPS, while only 23% (seven out of 30) are MPS schools from the original batch of SAGE schools, which are the only schools tested for the basis of the annual evaluation. Granted, each year demographic data are provided to show a relative equivalence between the student populations in SAGE versus comparison schools. Nevertheless, to the extent that MPS schools have a special, systematic effect on performance, which is statistically uncontrollable by simple demographic characteristics, that effect can be distorting the SAGE results.

A final concern with the SAGE design, at least from the perspective of a statistical research priority, is why the comparison is made between SAGE and non-SAGE *schools* and not between SAGE and non-SAGE *classrooms.* In other words, if a school has two or more first-grade classrooms, then one of them should have been selected to experience the reductions, while the other could be used for purposes of comparison. Such a procedure works to diminish the extent to which between-school factors affect the statistical results. This procedure was followed in the STAR evaluation, and was one of the most lauded design features of that study. Why a similar design was not followed by the UWM evaluation is uncertain.[59]

Despite these potential concerns with selection bias, Hawthorne effects, the assignment of comparison schools, and inter-district disparities, it is fair to conclude that the SAGE evaluation design, on a whole, is sound and deserving of the quantitative findings generated from its execution.

SAGE Results As Supplied To The Public

Arriving at sound quantitative results is only half the battle. The next step is to ensure a full and thorough explanation of all the findings, not just those that are favorable to reducing class size and continuing the program.

The UWM SAGE evaluations have been disseminated in a very selective manner to the general public and to state legislators who have voted to expand the program. The positive findings have been explained repeatedly and with vigor by the evaluation team and other education interest groups. The more ambiguous results that cast doubts on the continued, unabated expansion of class size reduction policies have been largely dismissed or marginalized. But it is the content of these dismissed findings that offer some of the best insights into how class size reduction efforts can be done more effectively and at a much lower cost, without any significant reduction in educational benefit to those populations who benefit from SAGE exposure.

To be fair, there is much to the SAGE program that appears highly beneficial. The summary of the "Major Findings" of the latest UWM evaluation, as they pertain to student achievement, are presented in Figure 4. On first glance, the results reported in the evaluation mirror the positive findings suggested by the STAR program, including higher achievement for students in smaller classes, particularly for minority students.

As with the first two evaluations, the results indicate that the program has a positive effect on student achievement, as shown by higher academic achievement in all core subjects over the comparison group of non-SAGE schools. These differences were found in both the raw scores and within the statistical models that controlled for various other factors. The gains were especially pronounced for African-American students prompting the following conclusion: "The data on the average performance of students in SAGE classrooms suggest that the lower student-teacher ratio in SAGE classrooms mitigates the negative achievement consequences of poverty." Such a finding directly matches the goal articulated in the original Urban Initiatives Task Force agenda.

FIGURE 4

Major Findings on the Achievement Effect of Class Size Reduction, 1998-99 UWM Evaluation

First Grade

- As was found in 1997-98, test scores of SAGE and Comparison schools show statistically higher performance of SAGE students in language arts, math and total scores on the post-test in 1998-99.
- In 1997-98 and again in 1998-99, African-American SAGE students made significantly larger gains than Comparison school students from pre-test to post-test, surpassing African-American Comparison school students on the post-test.
- In both 1997-98 and 1998-99 African-American SAGE students achieved greater relative gains on the total scale score than White SAGE students from pre- to post-test, closing the achievement gap. At the same time, African-Americans in Comparison schools achieved lesser gains, and the achievement gap with their White Comparison school classmates widened.

Second Grade

- African-American SAGE students scored significantly higher than African-American Comparison school students on every sub-test and total scores in 1998-99.
- When looking at gains made in 1998-99 from the first grade post-test to the second grade test, SAGE African-American students made the same significant gains that the White SAGE students did, and did close the achievement gap between African-American and White SAGE students although the relative gain was not significant.

Third Grade

- SAGE students scored significantly higher in reading, language arts, math, and total score than Comparison students on the third grade test in 1998-99.
- In 1998-99, test results suggest that statistically significant positive effects of SAGE occurred in first grade, were maintained in second and third grade.
- In 1998-99, African-American SAGE students performed significantly higher on every sub-test and total score over African-American Comparison students on the third grade test.
- When second grade is used for a baseline score, African-American SAGE students outperform African-American Comparison students in reading, math, and total in 1998-99.
- African-American students gained significantly more than SAGE White students in third grade, closing the achievement gap. In Comparison schools the gap between the performance of White and African-American students widened.

Additional Analyses

- Although they are not statistically significant, there are positive relationships between years of SAGE experience and student performance. This suggests that years of participation in SAGE may have a positive influence on achievement, although further research is necessary in this area.
- In all cases, classrooms with more affluent children outperformed classrooms with children from poorer families.
- Classrooms with 30:2 student-teacher ratios achieved just as well as classrooms with 15:1 student-teacher ratios with the exception of language arts and mathematics sub-tests in second grade.
- In 1998-99 at the class level of analysis, smaller classrooms tended to score significantly higher in language arts, mathematics and reading, as well as total score after adjusting for individual pre-test results, socio-economic status and attendance. In other words, classrooms with fewer students are more likely to have higher class average achievement scores.

Source: 1998-99 Results of the Student Guarantee in Education (SAGE) Program Evaluation, Executive Summary.

And Now, The Rest Of The Story: Results Not Presented To The Public

Unfortunately, the complete story on the effects of the SAGE program has yet to be told.

Certainly, the UWM evaluators did not make false representations of the SAGE data. The data results, so far as can be recognized by their replication in the SAGE evaluation text, are sound. But throughout the evaluations' reported analyses there remains a high reoccurrence of manipulation of the data's presentation. This manipulation is manifested in how the media reports of the program are disseminated to the public, which, as the findings below will show, simply does not tell the whole story.

Main Points

- Smaller classes in the second and third grades had a minimal impact and in some instances had no additional impact on student achievement.
- African-American students in smaller second grade and third grade classes in particular did not gain relative to their gains made in the first-grade *or* relative to African-American students in regular-sized classrooms.
- Smaller classes appear to not have any effect on students who are not African-American, who constitute the majority of students.
- The actual magnitude of the gains experienced by SAGE are, on average, relatively meager.
- Smaller classes have different effects on achievement depending on subject areas.
- The data do not separate out findings for MPS and non-MPS schools, nor did they look directly at the effects of smaller classes by income-level.

One of the distressing problems with uncritically accepting the assumed benefit of smaller classes is that many questions relevant to ensuring the most productive implementation of the program remain unanswered. These questions include whether smaller classes have an equal impact in all contexts, such as grade levels, school system size, and student characteristics, and if differential results occur based on other identifiable, or even unidentifiable, factors.

Most of the qualms with the UWM SAGE evaluation discussed below involve answering the preceding questions in some manner. Up until now, these questions have not been systematically answered. This is true in how the statistical findings were reported in the actual evaluation document (if they were reported at all) and even more so in how the findings of the three annual evaluations were presented to the public. Therefore, this report does not discuss all the findings of the evaluations, as those positive to smaller classes are already well discussed. The focus here is simply on the results that, given the manner in which the program is currently expanding, have not been well-argued, and which, if given greater attention, would create a more accurate picture of the effects of SAGE on Wisconsin education.

The following analyses are based solely on the SAGE data as presented within the three publicly available UWM SAGE evaluations for the school years 1996-97, 1997-98, and 1998-1999.[60] The data presented below are drawn directly from these evaluations, although frequently data tables have been reconstructed by this author for both the sake of clarity and emphasis. In all cases though, fidelity to the data as presented in the evaluations is maintained.

Second And Third Grade Results

The original SAGE schools (from which all the annual evaluations are based) were required by law to implement class size reductions in a staggered process. In the 1996-97 school year, the program's first year, grades kindergarten and first were required to implement the 15 to 1 ratio, and funding was allocated for that purpose. In the following school year, 1997-98, students in the second grade in the SAGE schools also experienced smaller classes. These students were, for the most part, those students who had also been in smaller classes as first graders the preceding year. Finally, in 1998-99, the initial SAGE schools were required to implement the smaller classes in grade three, which completed implementation in the grades for which the program applies.

Essentially, two main results can be discerned from the achievement scores of students in the second and third grades:

- First, it can be determined whether SAGE students do better academically than students in comparison schools *within those specific grade levels*.
- Second, we can see the achievement differentials between the two classroom types after two or three years of being in smaller classes, looking back at the *cumulative results* from all the grades.

It is to the second question that the SAGE evaluations frequently direct their findings. Yet answering the first question is of greater importance, because it explains whether placing students in smaller classes in later grades has any *independent* impact on improving student learning. According to the data presented in the SAGE evaluations, the achievement differences between SAGE students and comparison schools were either not increased by exposure to smaller classes in the second and third grade, or the increases were minimal and occurred only in some subject areas. According to the evaluation's own admission, in measuring the relative achievement gains in the second grade, "when the first-grade post-test is used as the baseline score, no significant results are found. This suggests that the statistically significant positive effects of SAGE occurred in the first grade. These positive effects were maintained, but did not significantly increase in second grade."[61]

Data from the 1997-98 and 1998-99 evaluations exemplify this point. In Table 2, results from the 1997-98 evaluation are replicated, and they show a minute and statistically insignificant gain for SAGE students from the end of the first grade to the end of the second grade. In other words, students in smaller classes in the second grade, who had also been in smaller classes in the first grade, performed *no better* on the second grade test than students who had been in larger classes in both the first and second grade.

TABLE 2: SAGE AND COMPARISON GAIN 1997-98, SECOND GRADE

	From First-grade Pre-test to Second Grade Post-test			From First-grade Post-test to Second Grade Post-test		
SCALE SCORE	***SAGE Gain***	***Comparison Gain***	***Gain Difference***	***SAGE Gain***	***Comparison Gain***	***Gain Difference***
Language Arts	77.07	71.74	5.33*	25.67	22.76	2.91
Reading	72.78	69.62	3.16	22.33	22.01	0.32
Mathematics	77.54	70.07	7.47*	22.86	21.97	0.89
Total	75.90	70.80	5.10*	23.67	22.36	1.31

*Significance at .05 level

Source: 1997-98 Evaluation of SAGE Program, Table 34.

It is interesting to note the reporting of the "from first-grade pre-test to second grade post-test" difference. This value spins the findings to better support the program's implementation in second grade, even though the better measure (that coming from the first-grade post-test) shows otherwise. The second measure is preferable because it separates out the gains experienced in the first grade from any second grade effects. Saying that significant gains were made by students in SAGE schools between the start of first grade and the end of second grade only makes it appear that independent gains were being made in both years. Yet the statistics in this case are clear that any remaining advantage between SAGE and comparison schools is simply a carry-over from the first year.

To those who support smaller class sizes, though, this result is disconcerting and unexplainable. The notion that smaller classes makes teaching and student learning more effective should apply equally in any classroom setting or grade level. Therefore, independent achievement gains should be expected by students exposed to smaller classes in the second and third grades, beyond merely the maintenance of any gains experienced from being in smaller classes in the first grade. But the evidence does not reveal these anticipated additional gains from continued exposure to smaller classes.

In the following year's evaluation (1998-99), results of between-grade gains (shown in Table 3) were more positive to future exposure to SAGE in the second and third grades, but the magnitude of the difference in gains decreased in each subsequent year, suggesting, at best, decreasing marginal returns from smaller classes in subsequent grades. The fact that SAGE students did statistically better than comparison students in the second grade in

TABLE 3 SAGE AND COMPARISON GAINS 1998-99, SECOND AND THIRD GRADE

	From First-grade Post-test to Second Grade Post-test			From Second-grade Post-test to Third Grade Test Post-test		
SCALE SCORE	***SAGE Gain***	***Comparison Gain***	***Gain Difference***	***SAGE Gain***	***Comparison Gain***	***Gain Difference***
Language Arts	23.88	16.98	6.90*	19.19	22.31	-3.11
Reading	26.90	26.32	0.59	26.25	21.85	4.40*
Mathematics	32.60	23.45	9.15*	36.40	29.42	6.98*
Total	27.87	22.20	5.68*	27.33	24.33	3.00*

*Significance at .05 level

Source: 1998-99 Evaluation Results of the SAGE Program, from Tables 24 and 35.

three of four scale scores in 1998-99, while SAGE students in the second grade a year before did no better in any of the four scale scores, only shows that the findings are simply not robust.

Overall so far, we have data on four, scale-score increase (student achievement) measures during the second grade in two years, and the third grade for one year. Therefore, there exist a total of 12 achievement gain differentials between SAGE and comparison students. In only half (6 of 12) did SAGE students perform statistically better, while in the other six there was no statistically significant difference between SAGE and comparison gains. Moreover, many of the statistically significant differences in gains barely reached the level of statistical significance employed by the evaluation team, and were relatively meager.[62] In other words, these are ambivalent results concerning the benefit of implementing smaller classes in grades after the first grade.[63]

African-American Students In The Second And Third Grades

Not only do smaller classes in SAGE seem to have little or no effect in later grades for students generally, but it is also unclear whether they have any positive impact on the group of students who otherwise appear to overwhelmingly benefit from being in smaller classes in the first grade — African-American students.

According to the first grade results from each annual evaluation, African-American students in SAGE statistically outperformed their counterparts in comparison schools by a fairly large margin (certainly much larger then the rest of the student population). The UWM evaluations announce this point prominently and repeatedly, as well they should.

What is not acknowledged to nearly the same extent is what happened to these same African-American students with an additional year of exposure to smaller classes in the second grade and then later in the third grade. This omission is unfortunate because it conceals the fact that these students did not consistently supersede comparison students in later grades, but instead actually had *smaller gains* in some cases. Even though these differences were not statistically significant, we do know for certain that African-American students did not gain more than their peers in larger classes during the second grade.

Tables 4 and 5 show the relative gains between African-American students in both SAGE and comparison schools in the second grade during the two years for which data are available. In all but one of the eight measures of student achievement, there was no difference between African-American students, whether they were in SAGE or comparison schools. In five of the eight measures, African-American SAGE students actually did *worse*, although the differences were not statically significant (at the .05 level). Reading scores in particular show consistently smaller gains for African-American SAGE students, relative to their counterparts in comparison schools.

In the most recent evaluation, there is a statistically significant positive gain difference of 8.38 points on the mathematics test for African-American students in the second grade, relative to comparison schools. This result occurred despite the fact that similarly situated students in SAGE schools average 4.12 points *worse* than students in comparison schools in the previous year. What explains this discrepancy from one year to the next is uncertain, but again, at best, it means the results are ambiguous.

The third grade results from the 1998-99 evaluation are more favorable and do show some statistically significant greater achievement gains for African-American SAGE students versus comparison students. These gains are shown in

TABLE 4 MEAN CHANGE FROM FIRST GRADE POST-TEST TO SECOND GRADE POST-TEST, 1997-98

SCORE	SAGE	COMPARISON	DIFFERENCE
Language Arts	19.41	20.62	-1.21
Reading	19.30	20.91	-1.61
Mathematics	16.46	20.58	-4.12
Total	17.99	21.14	-3.15

*Significant at .05 level

Source: 1997-98 Evaluation, from Table 43.

TABLE 5 MEAN CHANGE FROM FIRST GRADE POST-TEST TO SECOND GRADE POST-TEST, 1998-99

SCORE	SAGE	COMPARISON	DIFFERENCE
Language Arts	21.09	18.75	2.34
Reading	20.88	27.19	-6.31
Mathematics	28.79	20.41	8.38*
Total	24.04	22.05	1.99

*Significant at .05 level

Source: 1998-99 Evaluation, from Table 27.

TABLE 6 MEAN CHANGE FROM SECOND GRADE POST-TEST TO THIRD GRADE TEST, 1998-99

SCORE	SAGE	COMPARISON	DIFFERENCE
Language Arts	28.73	24.47	4.26
Reading	29.87	18.94	10.93*
Mathematics	48.20	29.59	18.61*
Total	35.87	24.06	11.81*

*Significant at .05 level

Source: 1998-99 Evaluation, from Table 39.

Table 6 and indicate that African-American SAGE students made greater gains than African-American students in comparison schools in the reading and math sub-tests and the total score, but they did not gain statistically greater in the language arts sub-test. It is interesting to note, though, that this cohort is the same as the one tested in the second grade in 1997-98, but which then produced no gain differences, as reported in Table 4 and discussed above.

Again, as was done with the total student populations, for just African-Americans, we find that in only one-third of the achievement gain measures (4 of 12) did SAGE students perform statistically better, while in the other two-thirds of measures there was no statistically significant difference between SAGE and comparison gains. In other words, the ambivalent results concerning the benefit of implementing smaller classes in grades after the first grade also exists for African-American students.

The UWM evaluations talk frequently about how exposure to SAGE has helped reduce the achievement gap between White and African-American students. This result is important and did occur, and will be discussed more below. But according to a more thorough reading of the data, nearly all of the reduction in the gap has occurred during the first grade and has not been significantly reduced in later grades, from repeated exposure to smaller classes. Arguments for expanding class size reductions to more grades, and to more second and third graders, are not supported by these, much less-publicized, results

The Evaluation's Attempt To Reconcile The Second Grade Results

Some explanations for these ambivalent results in later grades have been offered by the SAGE evaluation team. In the second year evaluation (1997-98), the evaluation team attempted to fend off the problems with the second grade results by incorporating three defenses.[64] For the most part these defenses are red herrings, pointing out concerns that, while possibly true, do not disallow the preceding interpretive effect of the second grade findings. The pertinent sections read:

1. *"During 1996-97 a considerable number of SAGE first graders achieved perfect scores on the spring post-test. This had the potential effect of placing a "ceiling" on the gains reported for SAGE first grade students. Conversely, what was a "ceiling" in 1996-97 became a "floor" for the scores of this group of SAGE students in second grade. It is not possible to know to what extent this phenomenon had an impact on the 1997-98 SAGE second-grade achieve - ment results."*[65]

The fact that the first grade post-test did have a significant number of students hit the ceiling is certainly a problem. But the nature of the problem would be to create an *overestimation* of the SAGE impact during the second grade, since the "floor" mentioned would be lower than it should. In other words, those students who achieved perfect scores in the first grade test will be treated as if they scored lower when the comparisons are made to their second grade scores. Therefore, gains made in the first grade will erroneously be applied to the second grade.

Regardless, the first-year evaluation results suffered from this same problem, yet it was not employed as a "caveat" in the same way (because the results were favorable). The first year study also had students who achieved perfect scores on the first grade post-test. To address this important statistical fact, the analyses of comparisons between pre- and post-tests in the first grade were first run with all the students, and then separately with the top quartile omitted. The rationale was to see if the differences between these pools of students (those who scored perfectly on the post-test versus. those who did not) were significantly different. They were not.

What this all means is one of two things. First, since the ceiling/floor effect was not found to be important when measuring comparative differences of performance between the pre-test and post-test in the first grade, we should infer it would also not be a problem with comparisons between the first grade post-test and the second grade post-test. Alternatively, we are left to ponder why the evaluation team considered this problem and set up a technique to address it in the first grade results, but then did not use the same method with the second grade? In any event, this "perfect score" caveat is not valid, at least to an extent that it would disallow the apparent lack of gain differentials between SAGE and comparison students in the second grade.

2. *"A second factor that may have influenced the second-grade results reported for SAGE students was that, because of uncertainties over funding for the second grade, nine of the thirty SAGE schools did not implement the program in second grade until after the start of the school year. In some cases implementation was delayed until January 1998."*

This is a most ingenious defense. Not only does it attempt to establish a quantitative concern, but it does so by placing blame on the political opponents of SAGE who caused the program's extension to be put in doubt. Of the

nine schools that implemented SAGE classrooms after the start of the school year, only two did so for classrooms in January, while the others mostly did so in October and November. Nevertheless, such a problem could very well taint the results.

Yet one obvious method to analyze this potential problem would be to run data analyses with these "late" schools omitted, and see if those results differ significantly from when *all* the second grades were included. In other words, take the SAGE classrooms in the remaining 21 schools that began at the year's onset and run the models with only them included. There certainly would still be a large enough sample size, and any other design concerns could be addressed statistically. The point is that this concern could have been answered and determined whether it truly had an impact on the findings. Not doing so, like in the example above, simply begs the question of why potential problems with the positive first grade findings were not addressed with the same level of effort as the potential problems with the unflattering second grade results.

3. *"Finally, the impact on class achievement scores of non-SAGE students entering the SAGE program for the first time in second grade is unknown."*

This is probably the most feasible defense. The students from which comparisons were made in the second grade were only those who had been in SAGE for both the first and second grades. Nevertheless, due to student migration, some students moved into SAGE classrooms for the first time when they were in the second grade. What is contended here is that while students who were first exposed to SAGE and smaller classes in the second grade were not tested, they may have had a negative classroom effect on the students who were tested, and who were in smaller classes for both years.

New And Refined Interpretations Of The Results Of SAGE In The Second And Third Grade

While the second and third grade results of SAGE are just as important as the first grade results, they have not been as well presented to the public. When these findings are discussed, it is frequently through language stating that gains for SAGE students from the first grade were "maintained" but did not increase in subsequent grades from being in the SAGE program. But education is a cumulative process and, therefore, one should expect that if smaller classes have an impact on learning in one grade, they should have an additional impact in other grades. To say that the gains from smaller classes in the first grade were "maintained" is just a sly way of saying that students in smaller classes did no worse than students in larger second and third grade classrooms. But they did no better either.

The main point is that the effects of smaller classes in later grades on students generally and on African-American students specifically is ambivalent.[66] This finding is important for a couple of reasons.

- *First* and foremost, this result supports the notion that smaller classes only have an academic effect in the first year of a child's schooling, while subsequent exposure to smaller classes is largely irrelevant to improving student achievement. Smaller classes are simply an effective means to enable students less-disposed to formal education to adjust to this setting in their first year or two of formal schooling. Once they have adjusted, further exposure to smaller classes seems unnecessary.
- *Second*, it reflects the findings from the Tennessee STAR study. STAR also showed that while a small but statistically significant difference in achievement was found for students in smaller classes after their *first year* of exposure to smaller classes, no gains were made relative to comparison students in larger classes from subsequent exposure to smaller classes. Since both SAGE and STAR, the two primary experimental studies on class size effects, had limited effects on students beyond the first-grade, the veracity of these findings is strong.
- *Third*, this finding suggests that smaller classes in later grades (such as grades two and three) are not required. Such an interpretation should be sobering to those who want to expand SAGE to even more grades and to all second and third grade classrooms in the state. Moreover, since kindergarten students were not assessed, we are uncertain what gains, if any, were made for SAGE students during that year, and if those gains increased or stayed the same in the following year, during the first grade. At minimum, much greater attention to the question of SAGE results in later grades is warranted.

DYNAMICS OF THE AFRICAN-AMERICAN VERSUS WHITE STUDENT ACHIEVEMENT GAP

In the previous section, it was established that African-American students in the SAGE program often did no better in the later grades than African-American students who were in larger sized classes. Still, the UWM evaluations have also reported that African-American students in SAGE classrooms closed the achievement gap with White students in the first grade, while that gap either widened or stayed the same in comparison schools with larger classes.

The question is if these results are because smaller classes have a greater achievement effect on African-American students, while White students also do better in smaller classes, *only to a lesser extent*? Or is it simply due to African-American students doing better in smaller classes in the first grade, while White and other students *do no better* at all when in smaller classes? If the case is the former, that would greatly support the merit of implementing smaller classes, since all students gain from being in smaller classes, only with African-American students — students who have traditionally achieved less — simply gaining more. That is a win-win situation for any education program. But if the reality is the second scenario, it helps to shatter the myth that smaller classes benefit all students, since only this one demographic group is essentially driving all the gains witnessed by SAGE and its smaller classes.

The available data appear to reveal that it is largely the second scenario that is taking place. In other words, there appears to be another major untold story from the SAGE program. Put simply, smaller classes, on average, do not appear to benefit non-African-American students, and when the performance of these students are analyzed separately, the effects of SAGE appear to be insignificant.

Unfortunately, the data from the evaluations skirt around this issue and do not provide direct analysis on this matter. Nevertheless, some statistical information is available from which such a conclusion could be inferred. For example, results reported from the 1996-97 evaluation stated that first graders in SAGE schools scored higher than did students in comparison schools. According to the evaluation's primary researcher, the difference averaged between the subjects was about 12-14% for all students.[67] The largest improvement was for African-American males, whose scores improved by 40% over similar students in regular classes. The results for female African-American students were just slightly lower.[68] We also know that African-Americans comprised 25% of the student population in that study. Therefore, simple mathematics reveals that the non-African-American students (75 percent of all students) only averaged approximately a four percent gain in SAGE versus the comparison schools.[69] This effect is not very large, especially when compared to other intervention options aimed at increasing academic achievement for low-income or African-American students.

Furthermore, results reported in the last two annual SAGE evaluations support the conclusion that gains made from African-American students are driving nearly all the gains witnessed by the program. Table 7 reveals a much smaller difference between SAGE and comparison schools for White students than for African-American students during the first grade in 1996-97 and 1997-98.

TABLE 7: CHANGE FROM PRE-TEST TO POST-TEST FOR FIRST-GRADE STUDENTS

	1996-97	1997-98
African-American		
SAGE Schools	54.4	52.15
Comparison Schools	42.0	32.78
Difference	***12.4***	***19.37***
White Students		
SAGE Schools	49.0	45.99
Comparison Schools	48.0	41.14
Difference	***1.0***	***4.85***

Source: 1996-97 Evaluation, Table 34, and 1997-98 Evaluation, Table 25.

Table 8 shows the gain differences for African-American and White students in comparison versus SAGE schools in 1998-99, a year later. The table reveals gains from smaller classes for White students, although the second and third grade gains were fairly small. Once again, these new findings suggest that the beneficial effects of smaller classes for White students are at best ambiguous.

That the SAGE program has positive effects for African-American students (in the first grade at least) is important in and of itself. Yet why the evaluations do not also explain the comparative lack of gains from being in small-

TABLE 8 CHANGES ON TOTAL SCALE SCORES, 1998-99

	From First Grade Pre-Test to First Grade Post-Test	From First Grade Post-test to Second Grade Test	From Second Grade Test to Third Grade Test
African-American			
SAGE Schools	48.46	24.04	35.87
Comparison Schools	34.62	22.05	24.66
Difference	***13.84***	***1.99***	***11.21***
White Students			
SAGE Schools	44.73	27.69	25.88
Comparison Schools	37.05	23.57	24.06
Difference	***7.68***	***4.12***	***1.82***

Source: 1998-99 Evaluation, from Tables 17, 28, and 40.

er classes for students in other racial groups is unclear. The possible counterpoint is that SAGE is helping to bridge the gap between these two racial groups and even if smaller classes are not helping White students, the cumulative result is beneficial. Perhaps. But if this is the case, then it is reasonable to ask if there exist other, much less expensive and more direct means of helping African-American students who are served well by smaller classes. Available research suggests there are such means, especially in the form of before- or after-school tutoring, among other methods. Likewise, students in demographics not performing much better in smaller classes may be better served having their per-pupil SAGE allocations used to enhance other educational factors.

TABLE 9 CHANGE FROM FIRST GRADE POST-TEST TO SECOND GRADE TEST

SAGE	1997-98	1998-99
African-American	17.99	24.04
White	24.50	27.69
Differences	***-8.46****	***-3.65***
Comparison		
African-American	21.14	22.05
White	23.80	23.57
Differences	***-.93***	***-1.52***

*Significance at .05 level

Source: 1997-98 Evaluation, from Table 45; 1998-99 Evaluation from Table 28.

It should also be noted that the SAGE evaluation only makes race-based comparisons between African-Americans and White students. In the three years of evaluations, African-Americans averaged 24.5% of the students in SAGE schools and 25.8% of comparison students; for White students, the average percentage in SAGE over the three years was 45.6% and for comparison was 51.5 percent.[70] What this means is that a significant portion of students tested (30% in SAGE schools and 23% in comparison schools) are not used in this race-based analysis. This emphasis is probably based on the tendency in educational research to analyze the "Black-White achievement gap," while ignoring other races of students.

One final comment on the Black-White achievement gap. The reductions of this gap did not occur during all grades in the analyses. In the second grade in 1997-98 and 1998-99, White students in SAGE actually outperformed African-American students in SAGE to *a larger extent* than in comparison schools. These results are shown in Table 9. Granted, in both of those years and in 1996-97, the first grade African-American students in SAGE did close the gap with White students in SAGE, while White students gained more than African-American students in comparison schools. Likewise, in the only year data were available for the third grade, results showed a collapsing of the achievement gap in SAGE schools, while African-American and White students in comparison schools gained the same amount between second and third grade. Regardless, the results exhibited in Table 9 show another area of ambivalence in the SAGE results, suggesting that other factors besides class size are at play.

A New And Refined Interpretation Of The Effect Of SAGE On All Students

Supporters of smaller class sizes are fond of claiming that all students in reduced-size classes are benefiting. But the preceding analysis, given the available data, does not appear to support that claim. The data seem to strongly suggest that the achievement gains witnessed by African-American students are driving the total aggregate gains witnessed by SAGE, while the program is having almost no meaningful effect on all other students. Once again, by uncritically accepting the assumed benefit of smaller classes, the fact that discrete populations of students benefit considerably more from smaller classes than others remains largely unexamined.

EFFECT SIZES: JUST HOW BIG ARE THE SAGE GAINS?

Another important aspect of the SAGE results is the practical size of the academic gains made by students in smaller classes over students in regular-sized classes. Until now, this issue has not been looked at in a very thorough and honest fashion. The SAGE results are frequently presented in language pronouncing that certain gains made by students in SAGE are statistically significant. Occasionally, even the word "statistically" is omitted and SAGE students are said have done "significantly" better than students in larger classes.

There is an inherent danger in such a presentation of the findings. The interpretation of statistical significance is clear, but unfortunately its relevance can be easily presented to the public in a disingenuous manner. Statistical significance simply means that some difference, *no matter how small in magnitude*, can be expected to actually exist. Or, alternatively, it means that we are certain to a high degree of probability that an apparent difference did not result simply by chance. Yet statistical significance is purely a mathematical construct, which requires a deeper level of analysis. Give a statistician a large enough sample size, and he can make nearly any difference mathematically come out statistically "significant." This is the case with the SAGE data.

Instead, the deeper issue is to what degree a statistical difference is *meaningful* to the public and to those involved in the program. In other words, the salient question is what is the actual size of the effects on student achievement caused by a program like SAGE? How much better in practical terms do students in SAGE perform than their counterparts in regular-sized classes? To answer that question we must look to a measure called "effect size."

While an effect size is also a statistical construct, its interpretation and relevance is more readily available to the average person. In layman's terms, effect sizes measure just what the name suggests — the size of the program's effect. But it does so in a statistical fashion that creates a figure readily comparable to other educational programs, even if the other programs have the same outcome (in this case student performance) being measured by different metrics.

The UWM evaluation ignored the salience of effect sizes in the first two of its three evaluations. Fortunately, an effect size calculation was presented in the most recent evaluation for the 1998-99 school year. The effect sizes for all grades and subjects in 1998-99 are presented in Table 10. Although the calculations were presented, the evaluation text contains not a single effort at interpreting the practical significance of the effect size figures provided. This is despite admitting that these computations "should aid in the evaluation of the practical significance of the class-size advantage."[71]

TABLE 10 ADJUSTED EFFECTS SIZES, SAGE 1998-99[72]

Test	First Grade	Second Grade	Third Grade
Mathematics	.147	.115	.193
Reading	.161	.277	.136
Language Arts	.308	.308	.350
Total Score	.236	.258	.213

Source: 1998-99 Evaluation, Table 44.

This omission is unfortunate, but not unexpected, given that the readily ascertainable interpretations of these effects sizes do not bode well for proponents of class size reductions. Essentially, what the effect size estimates reveal is that the test score gains experienced by SAGE students are relatively modest — only about five percent better than comparison students, on average.

In the case of the SAGE evaluation, looking to effect size measures provides very important revelations for two reasons. First, these numbers reveal the relative impact of class size reduction initiatives as compared to other available public policy programs aimed at achieving the same goal of increased student achievement. For the most part, the effect size values for SAGE are about the same, or slightly below that of other educational programs aimed at

similarly situated students as those found within SAGE.[73] The effect size values also comport with the values found for the Tennessee STAR data.

Second, effect sizes allow us to calculate the average differences between SAGE and non-SAGE students. Avoiding statistical verbiage, the measures of effect size allow us to estimate just how many points greater students in SAGE classes average over students in comparison classes.[74] That amount is calculated by multiplying the effect size by the pooled standard deviation from the average scores for both SAGE and comparison students. Table 11 shows the results of these computations and reveals the average number of points students in SAGE did better than students in comparison schools across the various tests. Unfortunately, the lowest and highest possible scores for each test vary by grade level.[75] Therefore, to give a sense of the relative magnitude of these gains, the total number of variable points between the lowest and highest obtainable scores are given for each subject and grade level.[76]

TABLE 11 AVERAGE GREATER TEST POINTS FOR SAGE OVER COMPARISON, 1998-99

Test	First Grade (Variable Points)	Second Grade	Third Grade
Mathematics	5.96 (356)	4.58 (373)	7.23 (355)
Reading	6.54 (294)	10.05 (299)	5.25 (323)
Language Arts	13.47 (280)	11.99 (282)	12.19 (275)
Total Score	8.53 (163)	8.62 (160)	7.12 (166)

What this exercise shows is the practical exaggeration of the effects of SAGE. Instead, what develops is an interpretation that moves beyond the statistician's refuge of saying that some difference is "statistically" significant, no matter how small that difference is in magnitude and by common understanding. SAGE students score just six points higher (less than two percent better) in mathematics in the first grade, and just five points better (again less than two percent better) on reading in the third grade. Even the largest average percentage increase is only 4.8 percent better for SAGE students on the Language Arts sub-test in the first grade. On the total scores, SAGE students averaged only about five percent better than comparison students. Is it worth spending millions of dollars a year on class size reductions to help students achieve only marginally better? Common sense would cast some doubt as to answering that question in the affirmative.

Appreciating the actual magnitude of the impact of class size reductions is another element of the SAGE program that is rarely discussed. This is unfortunate to those striving for a more intellectually rigorous look at the results of the SAGE program. By uncovering the ruse of hiding behind "statistical" significance, the SAGE gains are shown to be what they really are — relatively small and unimportant. Although achievement differences between SAGE and comparison students, in general, exist, the size of these differences is relatively insignificant.

SUBJECT AREAS

Having answered questions as to whether specific students and specific grades benefit from smaller classes, the next inquiry is whether all subjects are better learned by students in smaller classes, relative to their counterparts in regular-sized classrooms. Again, the data suggest different effects from smaller classes between subject areas.

In nearly all grades during the SAGE program's first three years, the performance of SAGE students versus comparison students on reading test scores was usually not that large, even when the difference was statistically significant. This fact did not stop State Superintendent Benson from claiming that "we know that smaller class sizes . . .increase student performance in critical areas like reading."[77] Reading is the subject area *least* improved by the SAGE program. SAGE gains in mathematics, on the other hand, were much larger, relative to the other subjects. The language arts score differentials usually fell in-between.

What is the practical significance of this finding? Obviously students in early primary grades are almost universally taught by one teacher in one classroom; we do not send first graders to different teachers depending on the subject matter, as is the common practice in high schools and many middle schools. Therefore, even if gains are appearing for some subjects and not others, it is seemingly unrealistic to expect a teaching style change based on this information.

TABLE 12 DIFFERENCE OF MEANS ON CTBS TESTS FOR SAGE OVER COMPARISON

Grade/Year	1st 1998-99	1st 1997-98	1st 1996-97	2nd 1998-99	2nd 1997-98	3rd 1998-99
Reading	1.99*	9.53	3.67	5.29	3.54*	6.31
Language Arts	4.32	12.04	5.78	11.06	8.21	6.56
Math	8.04	13.49	7.29	16.07	7.75	12.24
Total	4.54	11.76	5.40	10.94	5.89	8.13

*Not significant at .05 level

Source: 1996-97 Evaluation, tables 15, 17, 19, 21; 1997-98 Evaluation, tables 14-17 and 33; 1998-99 Evaluation, tables 13, 23, 34.

An answer to this legitimate inquiry is found in the various ways in which achieving the 15:1 student-teacher ratio occurs. One option in the SAGE program is to have floating teachers who enter classes of greater than 15 students at certain times of the day to aid in the teaching of core subjects. Therefore, knowledge that a smaller student-to-teacher ratio does not cause much greater student performance in, for example, reading instruction, would allow for a more efficient distribution of teachers overall. Regular class sizes used in this setting would be justified. Only during times that subjects are taught in which a smaller number of students per teacher is shown to have an effect would a roaming teacher be needed. Therefore, since the findings appear robust that learning in mathematics is greatly improved for students in smaller SAGE classes, this is where a roaming math teacher could improve outcomes.

The main point is that options exist that may better target the use of smaller classes during the teaching of subjects that have shown the greatest achievement benefit from that setting. Moreover, following such a strategy would generate cost savings.

QUESTIONS LEFT UNANSWERED: LOW-INCOME STUDENTS AND MPS EFFECTS

So far, the issues brought out above are based solely on the data as presented by the UWM evaluations. Still, some other important questions remain unanswered by the available data. It is certainly possible that other analyses could have been conducted and presented, which would have answered even more questions with regard to the relative impact of SAGE.

Some of these analyses were mentioned above, such as splitting the student populations by income class, instead of merely by race. A much more interesting and meaningful measurement would be emphasizing the relationship between poverty levels and the effect of smaller classes. To the extent that race (at least African-American heritage) is a reliable proxy for being low-income we are uncertain. It is always a dangerous sociological and statistical approach to equate poverty with race variables, which seemingly is the method of the UWM evaluations. In any event, data were collected on SAGE and comparison students' eligibility for free and reduced-price lunch, which is seemingly a more accurate measure of poverty. When a program is meant (at least originally) to target low-income students, it would be helpful to directly measure the program's effects on students by income level, instead of merely using race as a proxy for that measurement.

Another seemingly pertinent analysis not undertaken would involve separating the tested populations by those in MPS and those of out-state students. The SAGE evaluation does not separate MPS SAGE schools from non-MPS SAGE schools in the data presentations. This is done in part because, according to the DPI, "the SAGE evaluation is not designed to rank order the performance of SAGE schools." But the DPI goes on to say, "[the SAGE evaluation] is intended to provide policy makers with information they will need to understand the impact of the SAGE interventions."[78] Yet how are policy makers to be better informed about the real effects of SAGE in specific contexts, without this type of performance sorting, especially if there exist district-based effects. Moreover, another reason why the SAGE gains are seemingly the largest in Milwaukee may be because, based on current estimations of class sizes, classrooms in MPS had the farthest to go to get down to a fifteen-to-one student-teacher ratio.

Juxtaposing the previous points, it would be useful to know how many of the African-American students in SAGE are in MPS versus the rest of the state, given the emphasis on reductions between the Black-White achievement gap and the large percentage of African-American students in MPS. We know that the vast majority of low-income students in Wisconsin are concentrated in MPS,[79] and it would be useful to examine if the aggregate SAGE gains are really just being driven by results in MPS schools, as could be inferred.

These separate analyses would be highly useful in better tailoring the SAGE program to target those students who benefit the most from its implementation, and to further see if all students are improving their achievement level in SAGE classrooms, or just students within particular demographics. Further independent analysis of the SAGE results, which could look at these and other issues, must be continued, so that more of the important questions about the real effects of SAGE are answered.

Summary Comments On The SAGE Evaluation Results

Given the dramatic expansion of the SAGE program currently under way in Wisconsin, and the budgetary increases that correlate to this expansion, it is of great importance that the results of the program on which these expansions are based are presented accurately and completely. In one of the UWM evaluations, the statement is made that "studies such as STAR and SAGE can provide crucial information to policymakers."[80] This assertion is certainly true, provided that policymakers are supplied with the whole story and with undistorted interpretations about the magnitude and importance of the program's results.

Many of the comments made with regard to the SAGE evaluation must be taken in the light of what is feasible within this type of research. The great majority of the SAGE evaluation series is characteristic of a well-constructed design and generates enlightening findings. These findings are crucial to an informed understanding of how smaller classes are implemented and what their effects are, both on classroom instruction generally and, more importantly, on student achievement. The problem with the previous SAGE evaluations is not that they do not provide accurate and useful results; it is that they do not provide enough of the available information. In particular is the repeated suppression of the negative or ambiguous findings from the SAGE program.

A more detailed and open explanation of the SAGE results is important because many people still feel that smaller classes are good in and of themselves, and that they greatly benefit all students and in all contexts. Given the preceding analysis, instead of saying generally that students in SAGE's smaller classes do better than students in larger classes, we should say that "only African-American students in SAGE do better than African-American students in larger classes, and they do so only in the first grade, and only in certain subjects and classrooms. Furthermore, average SAGE gains are relatively meager."

This conclusion is a much narrower, but more accurate, portrayal of what the SAGE results reveal.

SMALLER CLASS SIZE VERSUS OTHER EDUCATION POLICIES: SMALL GAINS AT WHAT COST?

One point that nearly all the research on class size has shown is that class size reduction efforts are very expensive, especially if done in the manner advocated by the NEA, WEAC, and the Clinton administration — which is to simply fill the schools with more teachers. And for every percentage increase in the number of teachers, there is a proportional increase in the cost of a school districts' largest expense — that of teacher salaries and benefits.[81]

The relevant question to policy makers is not simply if class size reductions — of the nature currently proposed — increase student achievement. Regardless of the answer to this question, there remains the often overlooked, corollary issue of whether these increases are acceptable relative to the costs needed to achieve them. In addition, there is a definite need to see if other educational programs can better aid the students who seemingly benefit from smaller classes, and if these other programs do so at a substantial cost savings to the government and the taxpayers who underwrite public education in Wisconsin.

In this regard, much more attention must be drawn toward the costs associated with the SAGE initiative, relative to its impact. In an earlier section of this report, the large aggregate costs associated with the SAGE program and its expansion were explained. Another helpful way to look at SAGE costs is by focusing more on the school and stu-

dent level. The SAGE program gives a participating school $2,000 per at-risk student on top of the state-average, regular instruction costs per-pupil of approximately $3,600.[82] This represents a 56 percent cost increase for what appears to be only a five percent average increase in student achievement. This is not a very large effect, given the costs.

Teacher Quality And Experience

A number of education policy experts suggest that the gains realized by class size reductions are not nearly proportional to the costs required, especially on a comprehensive scale. To better highlight how the public funds spent to achieve smaller classes may be used more efficiently, we can look to one of the most prominent needs in education: that of increasing teacher quality and experience.

According to the National Commission on Teaching and America's Future, available research strongly shows that the impact of teacher qualifications is much greater than any other single factor, including class size.[83] Another recent study has shown that when measuring the effect of a $500 investment per student, spending on teacher education had the greatest impact on student achievement, while lowering the student-teacher ratio was found to have a smaller effect than increasing teacher education, teacher experience, and teacher salaries.[84] Finally, other studies have presented evidence showing that factors related to variations in teacher quality dwarf the effects of class size, and that "improvements in teacher quality … may be much less expensive than the large-scale reductions in class size needed to produce similar improvement."[85] This reality leads one education scholar to conclude that "perhaps children would be better served if the quality of instruction was viewed as more important than the number of children instructed."[86]

Another way to look at this tradeoff is to think of how the funds being used to reduce class sizes could alternatively be used to increase teacher salaries, which, theoretically at least, would draw more qualified people into the profession. While analyzing the costs of a national class size reduction program, a team of researchers explained that the costs needed to reduce class size to just 18 students in grades 1-3 would be enough to raise the salaries for every teacher in those grades, in every public school in the nation, by $10,000 per year.[87] A similar, but more accurate calculation could be made strictly for Wisconsin. Given that Wisconsin would need to spend approximately $128 million a year to implement SAGE in all K-3 public school classrooms, and that there are roughly 16,800 teachers in the state teaching those grades in public schools, that money could instead raise the salary of every one of those teachers by approximately $7,600 per year.

Chester E. Finn Jr., former Assistant Secretary to the US Department of Education, suggests that the $12 billion in funds being promoted by the Clinton Administration for smaller classes could be better used to offer scholarships to entice education students to graduate with majors in mathematics and sciences. According to the Department of Education, 36 percent of public school teachers of academic subjects have neither majored or minored in their main teaching field, and this problem is particularly acute in mathematics and sciences. Finn suggests that the aforementioned $12 billion could provide $4,500 college tuition grants for every one of the nation's 2.7 million teachers to rectify this serious competency problem.[88] While the intricacies of such a recommendation are substantial, the basic point is strong.

Yet one need only look to a deeper analysis of Wisconsin's own SAGE data to discover how much more important teacher characteristics are to improving student learning, independent of class size. Two researchers at UW-Milwaukee, one of whom is a member of the SAGE evaluation team, performed a very insightful analysis by looking into what may be driving the small, aggregate gains witnessed for SAGE.[89]

The analysis revealed two important findings. First, the researchers looked at the distribution of SAGE versus comparison school classrooms according to test score gains in the first grade in 1997-98. What is not seen is a discernable cluster of SAGE classrooms on the high-achieving end and a separate cluster of comparison classrooms on the low end, as may be expected. Indeed, the vast majority of both SAGE and comparison classrooms are clustered together in the middle, and there are a comparable number of SAGE classrooms to comparison classrooms in the bottom level. This result would suggest that something beyond just class size is driving achievement levels in these classrooms.

Second, they analyzed various classroom characteristics (including class size) to see what differences were pronounced between those classrooms in the top ten percent and bottom ten percent in terms of first grade achievement

gains. In general, the best scoring classrooms have no better attendance, similar socio-economic status, a slightly lower proportion of minorities, and average 16 students per class versus 20 students per class in the lower achieving classes. By far, the greatest differences are found with regard to parental contacts — high-achieving classes average many more annual parental contacts (34 vs. 8) — and teacher experience. High-achieving classes (whether SAGE or comparison), have teachers who average many more years of teaching experience in their local district. Both the average and median years of experience are about 17 years in high-achieving classrooms and only six years in the low-achieving classrooms.

These results compel the researchers to conclude that "small classes seem to set the stage, but do not guarantee better achievement," and that any achievement advantage from smaller classes is not shared by all classrooms. Instead, the results suggest that teacher experience plays a much greater role in generating student achievement gains than simply smaller classes.

The truth of the matter may very well be that smaller classes make it easier for teachers to do their job, as shown by the consistently positive effects of smaller classes in lowering disciplinary problems and increasing teacher morale. Yet at the same time, these benefits will not be directly transformed into improved student performance unless teachers are able to adapt their teaching styles to better suit the smaller classes. Given the preceding evidence, it seems that experienced or high-quality teachers are the only teachers who are incorporating the pedagogical benefits of smaller classes in a manner that results in greater academic achievement by their students.

Cost-Conscious Approaches To Smaller Classes

Even if smaller classes are desired, there are both efficient and inefficient ways to achieve the effects of smaller classes. For example, one of the greatest expenses needed for smaller classes is that of capital expenditures on building construction. Smaller classes usually mean more classrooms, and therefore bigger schools. Yet according to the UWM evaluations, students in classrooms of between fifteen and thirty students, but with two teachers, scored just well as students in classrooms of fifteen or less students and one teacher. These results suggest the ability to achieve the results of smaller classes in a manner that avoids the exorbitant costs of constructing more classrooms or enlarging school buildings.

In addition, this report has detailed how the gains from smaller classes are more prominent in certain settings (kindergarten and first grade, but not later grades) and for certain students (mostly African-American and low-income). Focusing class size reductions on schools with these populations, and only implementing the program in the first year of schooling, would have a dramatic effect on cost savings. According to one cost-benefit analysis on the impact of class size reductions, the public investment for reducing class sizes by one-third in only kindergarten and the first grade would produce an approximate rate of return on students' future earnings of around fifty percent. Implementing class size reductions by that amount in all grades K-6 would produce a rate of return of only seven percent.[90] Therefore, if SAGE is implemented only in kindergarten and the first grade, where the effects are the greatest, then SAGE expenditures could be cut roughly in half, without any appreciable decline in the recognized effects of SAGE. Alternatively, if SAGE funding were to be kept at its current level, but only K-1 classes were required to meet the 15 to 1 student-teacher ratio, then nearly twice as many schools could participate in SAGE — again, at no additional cost.

Also, if, as is suggested by the data, the positive effects generated from SAGE are driven by students in the Milwaukee Public Schools, and not statewide, it may be wise to target the class size reduction to just the MPS. Such a change may not be a politically viable option, given the desire of representatives statewide to secure SAGE funds to school districts they represent, but it would be markedly more cost-effective. As smaller class sizes become increasingly available to all students, regardless of economic status, the more diluted the impact of such a program will become.

It seems we are faced with two general ways to approach class size reduction in this state. The first would be to continue wholesale expansions of class size reduction efforts through the hiring of more teachers, and justify the investment through faith in the magic of these reductions. The second approach is to use the available research to see which schools, students, and subjects most directly benefit from smaller classes, determine what additional elements truly make smaller classes work, and target those areas with class size reductions.

Conclusion: A More Complete Picture Of The Costs And Benefits Of Class Size Reduction In Wisconsin

The impulse of moving toward smaller class sizes has been growing largely unabated for the past five years, as state, federal, and local governments across the nation have picked up the mantra of smaller classes. Yet the move to smaller class sizes necessitates a closer examination of these policies. Such an examination generates concerns as to the effectiveness of these smaller classes, relative to their costs.

Wisconsin's SAGE program began in the 1996-97 school year in 30 schools, and only two years later boasted 78 schools as participants. With the massive program expansion authorized in the 1999-2001 budget, as many as 600 schools throughout the state will now be participating, at a cost of roughly $60 million a year. What was once a pilot program aimed at determining if smaller classes improve student learning, has now become a full-blown policy that is on the fast track to becoming a permanent fixture of public education in Wisconsin. The SAGE program has apparently taken off on automatic pilot.

Unfortunately, too much of the current attractiveness of smaller class sizes is dictated by results purported to show conclusively that smaller classes have a significant effect on student learning. In particular, the recent SAGE expansions have been fueled primarily by the results of an evaluation series emphasizing the beneficial results of the program, while largely suppressing findings that question the efficacy of the very expansions the SAGE program is now experiencing.

As was demonstrated throughout this report, the research being presented in Wisconsin on this issue has some dubious results that cast doubt on the conclusions being used to perpetuate smaller classes. A closer examination of the available SAGE data reveals that, at best, smaller classes in the range of about 15 students per teacher help a small population of students: namely those in kindergarten and first grade and African-American students. Implementing class size reductions beyond these parameters is not justifiable based on the available evidence concerning the achievement effects of smaller classes in Wisconsin, and elsewhere.

Yet this type of expansion is precisely what is occurring within the SAGE program. The program was originally advertised and implemented as a means of increasing low-income student achievement by reducing class sizes in the early primary grades in schools with a large proportion of low-income students. Now the SAGE program contains little fidelity to targeting only low-income schools, as the 1999 budget revisions have completely eliminated the income requirements of students for schools to participate. Given the evidence that smaller classes only have an impact on low-income students, the wisdom of the recent decision by the State to eliminate any poverty requirement is dubious.

It appears that SAGE, as with the national policy discussion about smaller classes, is phrasing the debate less and less in terms of smaller classes in the earlier grades for low-income students, and more and more as simply the need for smaller classes across the board, for all students and in all grades. This distinction has dramatic policy and cost implications, with the former option being more limited and effective, while the latter much more costly.

The question is, and always has been, what effect is there, *on the margin*, from changes in class sizes? We already know that class size reductions cause a roughly proportional shift in cost increases. Unfortunately, the marginal achievement gains from these shifts do not appear to be very large, especially after the first grade. Therefore, class size reductions should not have to be implemented on a comprehensive scale with little or no reference the programs' relative academic gains.

It is time for Wisconsin to reexamine its impetuous rush to implement smaller classes, particularly in later grades and in schools that do not predominantly serve low-income students. This recommendation is made with full understanding that if state policy makers were to recommend limits in the SAGE program, they would likely be met with great resistance, as has occurred before. Smaller classes provide a tangible example for those who have always been arguing in a more general manner for greater spending on public education, and smaller classes are currently a very popular political idea. Nonetheless, issues of class size reduction must be looked at objectively and with an eye to matters of cost-effectiveness, if they are to be understood accurately and class size reduction programs are to be implemented effectively.

APPENDIX

District	School	1999-00 Funds	1998-99 Funds
Adams-Friendship	Adams-Friendship Elementary*	$336,000	$282,000
Adams-Friendship	Castle Rock Elementary	$58,000	$42,000
Antigo	Mattoon Elementary	$46,000	$36,000
Antigo	North Elementary	$120,000	$68,000
Appleton Area	Foster Elementary	$56,000	$39,600
Appleton Area	Jefferson Elementary	$191,200	$81,510
Ashland	Lake Superior Elementary	$416,000	$156,000
Ashland	Marengo Valley Elementary	$90,000	$78,000
Baraboo	South Elementary	$93,000	$60,000
Bayfield	Bayfield Elementary	$192,000	$114,000
Beloit	Robinson Elementary*	$108,000	$132,000
Bowler	Bowler Elementary	$72,000	$100,000
Boyceville Community	Tiffany Creek Elementary	$158,000	$126,000
Bruce	Exeland Elementary	$0	$18,000
Bruce	Bruce Elementary	$160,000	$86,000
Clayton	Clayton Elementary	$92,000	$66,000
Cudahy	Kosciuszko Elementary	$62,000	$47,000
Cudahy	Parkview Elementary*	$174,000	$171,194
Eau Claire	Longfellow Elementary	$242,000	$119,000
Fond du Lac	Chegwin Elementary	$154,000	$60,000
Gilman	Gilman Elementary*	$128,000	$198,000
Glidden	Glidden Elementary*	$72,400	$79,200
Green Bay	Fort Howard Elementary	$214,000	$152,000
Green Bay	Jefferson Elementary*	$202,000	$220,000
Janesville	Wilson Elementary*	$338,000	$310,000
Kenosha	Durkee Elementary*	$154,000	$112,727
Kenosha	McKinley Elementary	$144,000	$66,000
Kickapoo Area	Viola Elementary	$86,000	$52,000
La Crosse	Franklin Elementary*	$166,000	$187,547
La Crosse	Hamilton Elementary*	$260,000	$242,000
La Crosse	Jefferson Elementary	$106,000	$62,000
La Crosse	North Woods Elementary	$196,000	$140,000
Lac du Flambeau	Lac du Flambeau Elementary*	$300,000	$406,000
Ladysmith-Hawkins	Ladysmith Elementary	$232,000	$0
Laona	Robinson Elementary*	$120,000	$104,000
Madison	Glendale Elementary*	$200,000	$240,541
Madison	Mendota Elementary	$154,000	$148,000
Madison	Midvale Elementary	$484,000	$318,000
Menomonie Indian	Keshena Elementary*	$512,000	$590,000
Menomoniee Area	River Heights Elementary*	$330,000	$382,800
Milwaukee	Allen-Field Elementary	$714,000	$476,000
Milwaukee	Browning Elementary	$372,000	$0
Milwaukee	Carelton Elementary*	$429,000	$502,000

Milwaukee	Fairview Elementary*	$362,000	$498,000
Milwaukee	Forest Home Avenue Elementary	$768,000	$529,000
Milwaukee	Longfellow Elementary*	$634,000	$656,000
Milwaukee	Maple Tree Elementary*	$522,000	$564,000
Milwaukee	Maryland Avenue Elementary*	$228,000	$320,000
Milwaukee	Sherman Elementary*	$581,000	$619,000
Milwaukee	Story Elementary	$274,000	$0
Milwaukee	Thirty-Eighth Street Elementary	$444,000	$0
Milwaukee	Twenty-First Street Elementary	$464,000	$326,000
Milwaukee	Wisconsin Avenue Elementary	$634,000	$0
Milwaukee	Wisconsin Conservatory of Lifelong Learning*	$346,000	$404,000
Necedah	Necedah Elementary	$160,000	$102,000
Northwood	Minong Elementary	$0	$58,000
Northwood	Northwood Elementary	$102,400	$0
Oshkosh Area	Webster Stanley Elementary	$148,000	$78,000
Prentice	Ogema Elementary*	$56,000	$84,627
Prentice	Tripoli Elementary*	$60,000	$89,247
Racine	Giese Elementary	$156,000	$101,000
Sheboygan Area	Jefferson Elementary	$166,000	$142,000
Sheboygan Area	Washington Elementary	$236,000	$126,000
Siren	Siren Elementary*	$193,000	$186,000
South Shore	Oulu Elementary	$0	$28,000
South Shore	South Shore Elementary*	$68,000	$28,000
Stanley-Boyd Area	Boyd Elementary	$24,000	$29,000
Stanley-Boyd Area	Stanley Elementary	$156,400	$91,000
Stevens Point Area	Jefferson Elementary	$174,000	$91,000
Superior	Blaine Elementary*	$356,000	$457,580
Superior	Cooper Elementary*	$264,000	$329,974
Superior	Lake Superior Elementary	$116,000	$86,000
Superior	Pattison Elementary	$96,000	$98,000
Suring	Mountain Elementary*	$30,000	$38,000
Suring	Suring Elementary	$82,000	$48,000
Waukesha	White Rock Elementary	$194,000	$115,000
Wausau	Hawthorne Hills Elementary	$126,000	$0
Wausaukee	Wausaukee Elementary	$168,000	$116,000
Wauzeka-Steuben	Wauzeka Elementary	$56,000	$46,000
Webster	Webster Elementary*	$130,000	$212,000
Winter	Winter Elementary	$112,000	$64,000
Total		**$17,420,400**	**$13,631,547**
Average		***$220,511***	***$184,210***

*=Schools from the original 1995-97 budget cycle

Source DPI Website

NOTES

1. A 1997 *Education Week* survey found that 83 percent of teachers and 60 percent of principals believe class size should not exceed 17 students. According to a Milwaukee Teacher's Education Association survey conducted in the spring 1998, 90% of responding teachers say they have too many students in their class to meet the needs of individual students.

2. McDowell, Lena M. (June 2000). "Early Estimates of Public Elementary and Secondary Education Statistics: School Year 1999-2000," U.S. Department of Education, National Center for Education Statistics, Table 6.

3. *Digest of Education Statistics, 1999*, Elementary and Secondary Education, Table 69. National Center for Education Statistics. Washington, DC: Department of Education.

4. Allen, Russ. (1998). Class Sizes in Milwaukee Public Schools. Madison WI: Wisconsin Education Association Council. See also, http://www.weac.org/sage/research/gsclasssize.htm.

5. The exact class size averages for these grades in comparison schools in 1998-99 were: Kindergarten = 19.04; First Grade = 19.63; Second Grade = 20.04; and Third Grade = 20.51. These numbers infer that class sizes became marginally larger in each successive grade, but at a very small rate on average. Also, since these are averages, the numbers may be significantly affected by outliers.

6. Molnar, Alex & Janice Zmrazek. (September 1994). *Urban Initiatives Task Force: Recommendations and Action Plan* . Wisconsin Department of Public Instruction: Madison WI, p. 5.

7. SAGE funding is appropriated under Wis. Stats. s.20.255(2)(cu) (general fund); s.20.255(2)(cv) (supplemental fund), and s.20.255(2)(cs) (debt service fund).

8. To be eligible for SAGE prior to the program's 1999 revisions, a school district needed to have at least one school with an enrollment of at least 50% or more low-income pupils in the previous year. "Low income" means the measure of low income that is used by the school district under 20 USC 2723. Wis. Stats. 118.43(1)(b). The district could then enter into a contract with DPI on behalf of any of its schools that had a poverty rate of at least 30%, including the one exceeding the 50 percent poverty rate.

9. The SAGE program is considered categorical aid and, therefore, for every additional dollar of state funding under the SAGE program, local districts can see their amount of equalization aid from the state reduced by 33.3 cents. In other words, SAGE funds do come at some expense to local districts in the form of foregone state funds for other programs.

10. Teacher aides may be present in SAGE classrooms, but they may not be used to meet the SAGE class size requirement. The fifteen-to-one ratio must be achieved through the use of regular classroom teachers. DPI defines a "regular classroom teacher to be one with a license to teach in the grades covered by the SAGE program whose regular assignment is in one of those grades." *DPI SAGE Program Guidelines: Class Size*.

11. Class sizes were actually reduced in most SAGE classes to less than 15 students per teacher. According to the most recent year of available data, the average SAGE classroom size was 13.5 students per teacher. Forty-nine percent of SAGE classes had between 7 and 13 students per teacher, 44.8 percent had 14-16 students, while 6.7 actually had over 16 students per teacher. Molnar, A., Smith, P., Zahorik, J. et. al. (1999). 1998-99 Evaluation Results of the Student Achievement Guarantee in Education (SAGE) Evaluation. Wisconsin Department of Public Instruction, pp. 14-15.

12. In that first year, there were a total of 3,674 students in SAGE classrooms. Student numbers are based on Full-Time Equivalent (FTE) students. This means that kindergarten students that attend school for only half a day are considered as "one-half student" for calculation purposes.

13. Wisconsin 1997 Act 27.

14. Originally 50 new schools were added in the 1998-99 school year, but in the following two years two of those schools were consolidated out of existence.

15. School boards are required to pass a referendum and obtain DPI approval prior to June 30, 2001, to be eligible for this funding. The referendum must specify the amount of bonding attributable to increased classroom space needs as a result of participation in the SAGE program. MPS was excluded primarily because of the financial diversion already at work in that district with the implementation of neighborhood schools and their construction, which began in 1999 with the ending of forced busing.

16. A school that is currently in the P-5 program may enter into a SAGE contract if it removes itself from that program during its time in the SAGE program.

17. 1999 Wisconsin Administrative Code, Public Instruction — General School Aids, Table 3, p. 1069

18. In 2000-01, Act 9 funding for special education aids totaled $315,681,400.

19. *The Milwaukee Journal Sentinel*, June 19, 1999

20. Information provided by Janice Zmrazek, Wisconsin Department of Public Instruction.

21. The reason the annual increases required to meet the SAGE class size reduction requirements in the remaining 500 schools will be less than half the cost it will take to fund the current third round of contracts is likely due to the fact that the remaining schools will be those with the smallest proportion of low-income students. Therefore, the $2,000 per low-income student figure, from which all SAGE aid amounts are calculated, would apply to fewer students, and as a result the costs will be less.

22. *The Milwaukee Journal Sentinel*, October 22, 1998.

23. In requesting funds under this section, each district is required to include in the application a description of the district's program to reduce class size by hiring additional qualified teachers.

24. *Reducing Class Size: What Do We Know?* Washington DC: US Department of Education, May 1998.

25. To view WEAC's lobbying efforts, opinions and writings dealing with SAGE and class size, visit their internet site at: http://www.weac.org/capitol/sagepage.htm.

26. *See* Tommy Tomlinson. (1990). "Class Size and Public Policy: The Plot Thickens." *Contemporary Education*. Vol. 62, No 1, pp. 17-23.

27. Lartigue, Casey J. Jr. "Politicizing Class Size." *Education Week*. September 29, 1999.

28. Zahorik, John A. (September 1999). "Reducing Class Size Leads to Individualized Instruction." *Educational Leadership* Vol. 57, No. 1 pp. 50-53.

29. State Superintendent radio broadcast, February 1, 2000.

30. *See* Word, Elizabeth, et. al. (1990). "Student/Teacher Achievement Ratio (STAR) Final report 1985-1990," Tennessee Department of Education; *and* Finn, Jeremy D. & Chuck M. Achilles. "Tennessee's Class Size Study: Findings, Implications, Misconceptions." *Educational Evaluation and Policy Analysis* Vol. 21, No. 2, pp. 97-109.

31. *See* Nye, Barbara, et. al. (1995). "The Lasting Benefits Study," Nashville: Tennessee State University; *along with*, Nye, Barbara, Larry V. Hedges, and Spyros Konstantopoulos. (1999). "The Long-Term Effects of Small Classes: A Five Year Follow-Up of the Tennessee Class Size Experiment." *Educational Evaluation and Policy Analysis* Vol. 21, No. 2, pp. 127-142.

32. http://www.aft.org//vouchers/report/trump.htm

33. According to one recent analysis of STAR, performance on standardized tests increased, on average, by four percentile points the first year students attend smaller classes, while expanded by only one percentile point in subsequent years in smaller classes. Krueger, Alan. (1999). "Experimental Estimates of Education Production Functions," *The Quarterly Journal of Economics* Vol. 114 No. 2, May 1999, pp. 497-532.

34. Hanushek, Eric A. (1998). *The Evidence on Class Size*. Rochester, NY: University of Rochester, W. Allen Wallis Institute of Political Economy, p. 30 (also available in *Educational Evaluation and Policy Analysis* Vol. 21, No. 2, pp. 143-163).

35. For a more detailed discussion of this element of the statistical treatment of the STAR data turn to: Flake, Jeffery, Eric VonDohlen, & Mary Gifford. (1995). "Class Size and Student Achievement: Is There a Link?" Arizona Issue Analysis Report #135. Goldwater Institute: Phoenix, Arizona.

36. Other noted concerns over the STAR study involve the possibility of biased results due to the so-called Hawthorne effect and from self-selection. In terms of the Hawthorne Effect, there are concerns whether the participating teachers and schools, knowing that they were part of this highly publicized study, altered their behavior in ways beyond that which would have occurred if the program changes had occurred without this motivating force. Similarly, the second concern involves the method with which schools could be selected to participate in the program - namely, schools volunteered. Schools did need to have at least one of each type of classrooms (small, large and large with aid), which did enhance the experimental design, but the self-selection process could bias the results from the use of "eager" schools as opposed to all schools.

37. Glass, G. and Smith, M. (1978). Meta-Analysis of Research on Relationship of Class-Size and Achievement. *Educational Evaluation and Policy Analysis* Vol. 1 No. 1, pp: 2-16.

38. Welingsly, Harold. (April 1997). *When Money Matters*, Princeton NJ: Educational Testing Service.

39. *See* Bingham, S. C. (1993). "White-minority achievement gap reduction and small class size: A research and literature overview." Nashville, TN: Center of Excellence for Research and Policy on Basic Skills.

40. National Conference of State Legislatures. *Class Size Reduction*. www.ncsl.org/programs/educ/class.htm.

41. Hanushek, Eric A. (1998). *The Evidence on Class Size.*

42. Hanushek, Eric A. (1997). "Assessing the effects of school resources on student performance: An update." *Educational Evaluation and Policy Analysis* 19, no. 2 (Summer 1997), pp. 141-64.

43. Slavin, Robert E. (1989). "Class Size and Student Achievement: Small Effects of Small Classes." *Educational Psychologist*, Vol. 24; Slavin, Robert E. (1990). "Class Size and Student Achievement: Is Smaller better?" *Contemporary Education*, Vol. 62, No. 1.

44. Glen E. Robinson. (1990). "Synthesis of Research on the Effects of Class Size." *Educational Leadership*, Vol. 47, no. 7, pp. 80-90.

45. Some of these options include: peer tutoring (in the context of early primary students this would likely involve students in later primary grades assisting kindergarten through third graders); reading and language arts recovery programs for low-achieving students; and schedule shifting to allow for 30 minutes of the day for specialized instruction from smaller classes (half the students show up thirty minutes early and half stay thirty minutes late).

46. Odden, Allan. (1990). "Class Size and Student Achievement: Research-Based Policy Alternatives." *Educational Evaluation and Policy Analysis*, Vol. 12, No. 2.

47. Brewer, Dominic, et. al. (1999). "Estimating the Cost of National Class Size Reductions Under Different Policy Alternatives." *Educational Evaluation and Policy Analysis* Vol. 21, No. 2, pp. 179-190.

48. Johnson, Kirk A. (2000). "Do Small Classes Influence Academic Achievement? What the National Assessment of Educational Progress Shows." The Heritage Foundation, Washington: DC.

49. Zahorik. (1999).

50. For a complete discussion of this impact in California, *See*: *Class Size Reduction Program: Policy Brief*. California Legislative Analyst's Office. February, 1997.

51. Wis. Stats. 118.45(5).

52. Interview, Janice Zmrazek, April 13, 2000.

53. Maier, Peter, Alex Molnar, Stephen Percy, Phillip Smith, and John Zahorik. (1997). "First Year Results of the Student Achievement Guarantee in Education Program." Milwaukee, WI: University of Wisconsin-Milwaukee, p. 49.

54. *Ibid*, p. 51.

55. 1998-99 SAGE Evaluation, Table 6.

56. SAGE evaluation web site. http://www.uwm.edu/Dept/CERAI/sage.html.

57. Molnar, et. al. 1998-99 SAGE Evaluation, p. 15.

58. In the first evaluation, first grade students were given the level 10 test for both the pre- and post-test. This created a problem where a significant number of students (particularly SAGE students) achieved a perfect score on the post-test.

59. The evaluation team may counter that there is no difference in these constructs, and if there is, it would be accounted for in the statistical formulas and controls that were used, particularly in the Hierarchical Linear Model data analysis. But that statistical fact does not negate the reality that districts and their school *volunteered* to become a SAGE school.

60. Copies of these evaluation reports are available on-line at: http://www.uwm.edu/Dept/CERAI/sage.html

61. Molnar, A., Smith, P., Zahorik, J., et. al. (1998) 1997-98 Evaluation Results of the SAGE Program, Milwaukee, WI: University of Wisconsin-Milwaukee, p. 37.

62. Throughout the evaluations, many data results are reported as "significant at .05 level" but fail to specify the precise t score. For a difference to be significant at the .05 level, a t-score of at least 1.96 is required. T-scores on the differences between SAGE and comparison students, of those given, mostly hover around 2.00 - 3.00. But these values are only for the t-scores given. Nevertheless, given the size of the gains and the known standard deviations, it appears as if many of the "significant" values that were listed without t-scores also barely reached the value to be significant at the .05 level.

63. A note on the significant gains achieved by SAGE students in the first grade. As mentioned previously, the statistical gains for SAGE students were greater than students in regular-sized classrooms. Yet some of the explanation for the larger gains, at least in the 1998-99 analysis, can be attributable to the fact that SAGE students had greater room to improve since they scored statistically significantly lower than comparison students on the pre-test, even before any exposure to smaller classes. In other words, in the first grade, SAGE students' gains can be much larger without the actual within-grade changes being that large. This happened for first graders in 1998-99, while in the other two years no statistical difference existed between SAGE and comparison students on the pre-test.

64. Please note that in 1997-98 no results were yet available for SAGE students in the third grade since the program had yet to expand to those grades in the original 30 SAGE schools.

65. 1997-98 Evaluation of the SAGE Program. University of Wisconsin-Milwaukee, December 1998, p. 84.

66. Unfortunately, when the "major findings" of the evaluation are presented to the public, only the instances where positive gains are found for SAGE are those results mentioned. When no gains arise for SAGE students, the evaluation either remains silent or employs euphemistic language to say achievement gains are "maintained in later grades." Moreover, the evaluations' major findings include the reporting of statistically *insignificant* gains when SAGE students achieve them, and does not report similar results when comparison students actually performed better than SAGE students, even if the result was not statistically significant.

67. *See* Molnar, Alex. "Smaller Classes, Not Vouchers, Increase Student Achievement," Harrisburg PA: Keystone Research Center, January 1998.

68. Maier, et. al. First Year Results of SAGE, UW-Milwaukee, December, 1997.

69. Figure is derived from the following calculation: percentage gain of African-American students [percentage of African-American students in SAGE] + percentage gain of non-African-American students [percentage of non-African-American students in SAGE] = total percentage gain of students in SAGE. Result: .40(.25) + x(.75) = .13

70. Figures from 1996-97 SAGE Evaluation, Table 4; 1997-98 SAGE Evaluation, Table 2; 1998-99 SAGE Evaluation, Table 2.

71. 1998-99 SAGE Evaluation, p. 55.

72. Adjusted effects sizes were computed by controlling for the effect of students' pre-tests, socio-economic status and attendance differences on the mean scores of both SAGE and comparison students.

73. For more information on the differences of effect sizes between smaller class size policies and other education policies *see* Mayer, Susan E. & Paul E. Peterson. (1999) "The Costs and Benefits of School Reform." In *Earning and Learning: How Schools Matter.* S. E. Mayer & P. E. Peterson (eds.) Washington DC: Brookings Institution; *and* Greenwald, Rob, Larry V. Hedges, Richard D. Laine. (Fall 1996). "The Effect of School Resources on Student Achievement," *Review of Educational Research*, v. 66, no. 3, pp. 361-396.

74. For example, the third grade average SAGE score for math was 607.53, with a standard deviation of 36.69, while the third grade average comparison score for math was 595.29, with a standard deviation of 38.23. The pooled standard deviation is (36.69+38.23)/2 = 37.46. This figure is then multiplied by .193 (the effect size), for an average of 7.23 points better.

75. For example, the first grade test for reading is the CTBS Level 11 sub-test. Due to the test system's scoring, the lowest obtainable score on that sub-test is 407 and the highest obtainable score is 701, for a possible range of 356 points. A year later in the second grade, students will take the CTBS Level 12 sub-test, where the lowest obtainable score is 423 and the highest score possible is 722, for a total range of 299 points.

76. The lowest and highest obtainable scores for the reading, mathematics, and Language Arts sub-test are taken from Table 45 of the 1998-99 UWM SAGE Evaluation. Numbers were not given in that table for the lowest and possible *total* scale scores for each level. For the total scale score tests, I have therefore used the lowest and highest possible scores for those tests from the 1999-2000 Winter Norms CTB book, even though the tests administered would have been from the spring 1999 version.

77. Radio address from the State Superintendent, December 14, 1999.

78. SAGE Newsletter, January 1997, Vol. 1, No. 1. Department of Public Instruction.

79. For a more detailed explanation of how low-income students are distributed between MPS and non-MPS schools, see, Wisconsin Legislative Fiscal Bureau (May 1999) Paper #778: Student Achievement Guarantee in Education.

80. 1997-98 SAGE Evaluation, p. 6.

81. According to Mayer and Peterson, reducing class size by a third in any particular grade requires a 50 percent increase in expenditures on the salaries and benefits of teachers for that grade. *See* Mayer and Peterson. (1999). p. 351.

82. The figure for the average regular instruction cost per pupil comes from a State of Wisconsin Legislative Fiscal Bureau's report, *Study of School District Costs, 1996-97.* Madison, WI: Legislative Fiscal Bureau, April 1999, p. 6. This figure is used instead of the average total per-pupil expenditure in the state because it represents a more accurate figure of what precisely the SAGE funds supplement, which is expenditures on direct classroom functions. According to the LFB, "regular instruction consists of activities directly associated with teaching interactions between staff (teachers, classroom aides and others) and students at all grade levels. . . It includes not only the salaries and benefits of instructional staff, but also supplies and materials, such as textbooks, used in the process of teaching. With the exception of special education and co-curricular activities, all instructional activities are included within this category. . . " This number does not include the administrative costs associated with public education, which comprise almost half of the standard per-pupil expenditure.

83. Reported in *The Milwaukee Journal Sentinel,* February 27, 2000.

84. Greenwald, Rob, Larry V. Hedges, Richard D. Laine. (Fall 1996). "The Effect of School Resources on Student Achievement," *Review of Educational Research*, v. 66, no. 3, pp. 361-396.

85. Rivkin, Steven G., Eric A. Hanushek, and John F. Kain. "Teachers, Schools, and Academic Achievement." July 1998 (revised April, 2000).

86. Heather Hardner. (1990). "ACritical Look at Reduced Class Sizes." *Contemporary Education*, Vol. 62, No. 1. p. 29.

87. Brewer, et. al. (1999). p. 188.

88. Chester E. Finn Jr. and Michael J. Petrilli. "The Elixir of Class Size," *Weekly Standard*, March 9, 1998.

89. Smith, Philip & William Kritek. (May 1999). "The Effects of Class Size on Student Achievement: A Closer Look at Conventional Wisdom" Report prepared for the Institute for Excellence in Urban Education, University of Wisconsin-Milwaukee School of Education.

90. *See*, Susan E. Mayer and Paul E. Peterson. (1999). "The Costs and Benefits of School Reform." In *Earning and Learning: How Schools Matte*r. S. E. Mayer & P. E. Peterson (eds.). Washington DC: Brookings Institution.

ABOUT THE INSTITUTE

The **Wisconsin Policy Research Institute** is a not-for-profit institute established to study public-policy issues affecting the state of Wisconsin.

Under the new federalism, government policy increasingly is made at the state and local levels. These public-policy decisions affect the life of every citizen in the state. Our goal is to provide nonpartisan research on key issues affecting Wisconsinites, so that their elected representatives can make informed decisions to improve the quality of life and future of the state.

Our major priority is to increase the accountability of Wisconsin's government. State and local governments must be responsive to the citizenry, both in terms of the programs they devise and the tax money they spend. Accountability should apply in every area to which the state devotes the public's funds.

The Institute's agenda encompasses the following issues: education, welfare and social services, criminal justice, taxes and spending, and economic development.

We believe that the views of the citizens of Wisconsin should guide the decisions of government officials. To help accomplish this, we also conduct regular public-opinion polls that are designed to inform public officials about how the citizenry views major statewide issues. These polls are disseminated through the media and are made available to the general public and the legislative and executive branches of state government. It is essential that elected officials remember that all of the programs they create and all of the money they spend comes from the citizens of Wisconsin and is made available through their taxes. Public policy should reflect the real needs and concerns of all of the citizens of the state and not those of specific special-interest groups.