Is the 'Evidence-Based Approach' a Good Guide to School Finance Policy?*

Eric A. Hanushek
Stanford University

March 2007

* This paper is an extended version of the peer review paper commissioned by Washington Learns.

Is the 'Evidence-Based Approach' a Good Guide to Finance Policy?

by Eric A. Hanushek

Washington, like many other states, is struggling with the twin problems of how to improve its schools and how to finance them. As recognized in the establishment of Washington Learns, perhaps no policy issue is more important for the future of the State of Washington.

To aid in answering these questions, Picus and Associates have prepared a report for the K-12 Advisory Committee of Washington Learns. This report utilizes their "evidence-based approach" (Odden, Picus, Goetz, and Fermanich (2006)) [OPGF]. Unfortunately, were Washington to implement these recommendations, there is little reason to expect student achievement to improve measurably. The only outcome that could be expected with any certainty is that expenditures would dramatically increase. Moreover, their recommendations would make Washington schools substantially less efficient in their operations and perhaps make them less prone to change.

Previous analyses have described how each of the "costing out" methods applied to determine the "adequacy" of school spending is fundamentally flawed.[1] These costing out studies (of which the evidence-based approach is one example) lack a sound scientific basis and typically point to inefficient and ineffective policies. Nonetheless, it is frequently difficult for people not thoroughly knowledgeable about the intricacies of such studies to see the application of the general principles to specific consultant reports and recommendations.

This paper provides a critique of the evidenced-based approach as provided to Washington Learns. The main points are:

1. The OPGF evidence is highly selected and generally of insufficient quality to be the basis for policy decisions.

2. The results in terms of student achievement summarized by the OPGF report are not credible.

3. The suggested approach to choosing among policies guarantees substantial inefficiencies in school spending and operations.

4. The programs creating the underlying OPGF rational for funding likely bear little relationship to the actions districts would take if provided with the funding changes suggested.

[1] See Hanushek (2006, (2007); available at http://www.hanushek.net.

5. The funding structure recommended does little to change the incentives facing schools to use money wisely or to improve student performance.

The OPGF finance plan suggests substituting consultants' judgments for the political appropriations process. And, it points toward a regulatory solution that would have the state dictate details of school operations to districts. Both of these are highly inappropriate given the state of knowledge about "what works" and the weak evidence that forms the basis for the proposals. There is **no existing evidence of any gains** in student achievement to be found from other states that have followed the costing out studies of these and other consultants.

The alternative that Washington Learns should consider is building upon the current strong performance accountability in Washington and relying much more on incentives that are related to student achievement. The accountability system is designed to provide feedback on where performance is good and where it is improving. The OPGF proposal would ignore on-going evidence on performance. More importantly, they would not alter incentives to schools, teachers, and students. Indeed, the importance of improving incentives to local schools, especially when the best approach is uncertain, was long ago recognized by Odden and Picus (Picus (1992)), but here it is relegated to being a side comment. Their proposed approach could actually weaken achievement in Washington if schools faced the current incentives but were given less flexibility in their programmatic choices. In the face of programmatic uncertainty, the correct answer is never to lock in a rigid set of policies.

Finally, the policy process is more sensibly viewed as an on-going series of adjustments, not as a one-time fix. From this perspective, ways to learn from the process must be built in if the schools are to improve over time. Because OPGF want to stress that they now have sufficient evidence to make policy, they downplay the importance of systematically evaluating and learning from the programs that are introduced. While they are comfortable with mandating a set of programs for all of the schools in the state, policy makers should recognize that some (many? all?) of their program suggestions are likely to have much less impact than they suggest. If, however, this package of proposals is implemented across the entire state and if there is no explicit program of evaluation, the state will not know what to change in the future in order to improve student performance or the efficiency of state school policies. The inevitable follow-on commission would have no better information than Washington Learns and would again have to cull ideas from research evidence that does not directly relate to state policies.

1. The OPGF evidence is highly selected and generally of insufficient quality to be the basis for policy decisions.

The main message of the OPGF report is that it identifies solid scientific evidence that could dramatically increase student performance. It lists various reports and publications that are identified as forming the basis for their recommendations. And it

has frequent discussions of the importance of research quality, the value of random assignment studies, and the like. But in the end their report fails to provide reliable program evidence.

The studies that OPGF cite prove to be remarkably poor in scientific quality. There is a single random assignment study in their main findings – the Tennessee STAR study of class size conducted in the mid1980s.[2] The remaining major areas cited rely on much less reliable research. For example, the discussion of school and district size relies on research in the late 1970s and early 1980s – research that tends to be highly suspect because of ignoring the impacts on student achievement while concentrating just on average spending. The discussion of school based instructional coaches relies on some case study evidence in unpublished materials or in a publication of the Association for Supervision and Curriculum Development – the professional association for the coaches. Even though professional development across schools is broadly thought to be highly variable in its effectiveness, OPGF are convinced that they know the "right way" and that it would then be very effective.

Most importantly, OPGF tend to select studies by their results. The studies that form the basis for their recommendations tend to be those that show large impacts on student achievement (see below). While it would be appropriate to weight high quality studies more than low quality studies in arriving at a judgment about the research, it is completely inappropriate from a scientific viewpoint to choose studies on the basis of their results. OPGF tend to do the latter.

The largest concern with their review is that it does not select or concentrate on studies that credibly identify the *causal impact* of the policies considered. Much of the existing research in education relies upon observed correlations of particular factors and student outcomes. This research does not provide much confidence that instituting a particular studied policy would have anywhere near the same impact on students when implemented in a new setting. The primary reason is that the research has not been successful at identifying the factors that are truly driving achievement.

This issue is frequently discussed under the heading of "the problem of going to scale," because of past disappointments from instituting 'research based' policies but seeing little of the predicted results. Fundamentally, it is almost always a problem of inappropriately assuming that the research has identified true causal factors.

A simple example can help to understand the issue. A large number of studies have addressed specific kinds of professional development programs for teachers. They typically will consider whether a particular approach or program will improve student performance. But, it is often very difficult to distinguish between the program and the specific people providing the professional development. If in fact the outcome of the program is most dependent on the quality of the people and not the specific program, the research would give the misleading impression that the program should be broadly implemented. When different people actually provide the professional development, the

[2] Importantly, the STAR study has been subject to considerable discussion and debate and is just one piece of the available evidence on class size. The extensive debates on class size policy of the last decade underscore a degree of policy uncertainty that is ignored by OPGF. See Ehrenberg, Brewer, Gamoran, and Willms (2001a, (2001b), Hanushek (1999a), Mishel and Rothstein (2002).

results are likely to be very different. In other words, the research that concentrates on the differences among programs mistakes what is the underlying causal factor that drives achievement, which in this example is the specific trainers.

It is now well known that study of selective samples of students and schools can lead to very biased results in the analysis of effectiveness. In general, new programs are not introduced to a random selection of students and schools, and they are not run by the typical person who might provide the program subsequently. If this selectivity is not adequately considered, the results of any evaluation studies will provide very misleading guidance about the potential impacts of any program.

A variety of approaches have been developed to identify the causal impacts of different factors, but they have not been widely applied in the areas that OPGF consider.[3] In fact, very few of the studies cited by them even address the issue, let alone deal with it in a satisfactory manner.

What the OPGF report fails to acknowledge is that the research base in important areas of their proposals is not sufficient for their purposes. It is simply the case that even the best research in some areas cannot support current policy decisions that involve substantial expenditures. Low quality research means that the outcomes that can be expected from any policy change are highly uncertain and/or likely to be wrong. When faced with low quality research, the answer is seldom to "follow program X because it is the best evidence available." Such strategies generally waste money. Moreover, the OPGF proposals do not provide a method for learning about better possible approaches.

A much better approach, as discussed below, it to introduce programs on a more limited, trial basis where the results can be assessed. Doing so promises a much more rational decision process.

2. The results in terms of student achievement summarized by the OPGF report are not credible.

One way to understand the selective nature of the OPGF study is simply to take their research at face value, i.e., to ignore the problems previously identified and to look at the results that they cite. Their reported program impacts are not credible, suggesting either poor underlying research and/or selective choice of evidence.

Their studies are easiest to summarize by looking at the "effect sizes" that they identify. The effect size is the standard deviations in student outcomes that are purported to result from the given policy.

The interpretation of effect sizes can be seen by looking at how students would be moved in the achievement distribution from a given program. Consider the average student in Washington State – the student at the 50th percentile of the achievement distribution. Table 1 provides a description of where this student would move with

[3] Random assignment methodology is just one of a wide number of approaches used for dealing with these issues. Various researchers have applied combinations of statistical analyses (regression discontinuity, matching, etc.), panel data analysis, difference-in-differences methodology, natural experiments, and so forth to deal with sample selection issues and the problems of causal inference. These approaches are well known in the modern program evaluation literature but almost completely absent from the references in the long but undistinguished bibliography of the OPGF report.

programs of varying effect sizes. Without the program (zero effect size), the person stays in the middle of the distribution. If she is provided with a program that has an effect size of 0.5, she would move up in the distribution by one-half standard deviation, and this would place her at the 69th percentile of the distribution. An effect size of 1.0 would place her in the 84th percentile – i.e., only 16 percent of the students perform better. An effect size of 2.0 places the average student at the 98th percentile. Above 3.0, it is very difficult to find *any* existing student who does better than this formerly average student.

Another way to understand the magnitude of effect sizes is to recognize the performance of students on the National Assessment of Educational Progress (NAEP). NAEP attempts to put all students on a common scale. When this is done, the typical eighth grade student is approximately one standard deviation ahead of the typical fourth grade student in the same subject. In other words, one standard deviation of knowledge is approximately the average learning in four years of schooling.

It is also possible to characterize the magnitude of achievement gaps in terms of standard deviations. On the NAEP, the magnitude of the black-white achievement gaps varies by test and grade from approximately ¾ of a standard deviation to somewhat over one standard deviation. Thus, a change of one standard deviation in performance by black students could essentially close the racial achievement gap.

Now consider the effect sizes identified by OPGF and cited as the basis for their set of policy proposals. Table 2 is an expanded version of their table (p. 64) that gives the estimated effective sizes of a subset of OPGF recommendations. Table 2 gives the range of estimated effect sizes that they report.

To understand the impact of the separate programs on overall achievement in Washington, it is important to note that some of their programs apply just to a portion of the student population. For example, several apply just to disadvantaged populations (which here is taken to be the 36 percent of the Washington student population that is eligible for free or reduced price lunch). The final two columns translates the OPGF effect sizes for students in the separate programs into the impact on overall achievement of Washington students by taking into account what proportion of students are affected by the program. This latter calculation simply involves calculating what portion of the students receives the program and then averaging across all students in the state. For example, a program that yielded a one standard deviation increase in performance for just disadvantaged students would lead to a 0.36 standard deviation increase when averaged across both disadvantaged and nondisadvantaged students.

As Table 2 shows, full day kindergarten is estimated to have an effect size of 0.77 for both advantaged and disadvantaged students. This implies that just moving from half day to full day kindergarten would move the average student to the 78th percentile of the current distribution (which is the precise impact of an effect size of 0.77).

Some of their programs have differential effects, as also identified in Table 2. Class size reduction has a larger impact on disadvantaged students (effect size of 0.5) than advantaged (0.25). When averaged across the two groups of students (see the last two columns), this implies that universal class size reduction across the state would have an overall impact of 0.34 standard deviations, or moving the average student to the 63rd percentile of what is observed today.

Table 1. How Programs of Different Effect Sizes Move Students in the Achievement Distribution

Effect Size	New position in distribution (percentile)
0	50th
0.25	60th
0.5	69th
1.0	84th
1.5	93rd
2.0	98th
2.5	99th
3.0	99.9th

Table 2. The Estimated Benefits of the OPGF "Evidence-Based" Programs are Not Credible (Estimated effect sizes for OPGF programs and impacts on average Washington achievement)

	Estimated Effect Sizes of Policies[a]		Effect of fully implemented program on overall WA achievement (s.d.)[b]	
Policy	NonDisadvantaged	Disadvantaged	Minimum	Maximum
Full day kindergarten	0.77	0.77	0.77	0.77
Class size reduction	0.25	0.5	0.34	0.34
Multi-age classrooms	0.0-0.5	0.0-0.5	0	0.5
Classroom coaches	1.25-2.7	1.25-2.7	1.25	2.7
Tutoring, 1-1	0	0.4-2.5	0.144	0.9
Summer school (1/2 FRL)	n.a.	0.45	0.081	0.081
Embedded technology	0.30-0.38	0.30-0.38	0.3	0.38
Gifted and talented (5%)	0.5-1.0	0.5-1.0	0.025	0.05
Accelerated instruction (2%)	0.4-0.7	0.4-0.7	0.008	0.014
		TOTALS	3.08	5.897

Notes:

a. The effect size (i.e., change in standard deviations of student performance) attributed to students in each program. Some estimates are given by OPGF as a range, and some differ for disadvantaged students (36 percent of the population) and more advantaged students (64 percent of the population).

b. The effect on overall Washington achievement recognizes that some programs (e.g., tutoring) apply just to disadvantaged students or have a differential effect on some students (e.g., class size reduction). The overall effect is how much the average achievement in Washington is estimated to change (in standard deviations) from the OPGF proposal. The minimum and maximum values reflect the range of estimates in the first two columns.

Source: Author calculations from Odden, Picus, Goetz, and Fermanich (2006), p. 64.

Finally, OPGF cite uncertainty about the precise effect size to be expected from some programs. Thus, the table gives their range of estimates and again translates these into an estimate of the overall impacts on average achievement in Washington. Look, for example, at professional development with classroom coaches. There is a range of estimates provided (by ACSD, the professional association). When translated into overall impact in the last two columns, they indicate the average student exposed to teachers with such a program would move to somewhere between the 89th percentile and the 99.7th percentile of the distribution.

They concentrate, as another example, on summer school programs for disadvantaged students. These are identified as having an effective size of 0.45, which translates into an overall impact on average achievement of 0.081 standard deviations assuming that advantaged students do not get any gains from the program. Interestingly, however, in Odden's peer review in 2004 of the Ohio study of the costs of implementing NCLB, he castigates the study's authors who called for expanded summer school because they "reference no research to support this assertion, when in fact most research shows that summer school typically administered has little if any impact on learning." [4] This contrast underscores the selective nature of the "evidence" cited in this report.

OPGF develop a program that simultaneously implements all of these listed programs (and some others). Moreover, they imply that the programs would work independently, so that doing all simultaneously would yield the sum of all of the effect sizes.[5] These are given at the bottom of Table 2. By their estimates, student performance would improve by 3 to 6 standard deviations.[6] In other words, thinking back to Table 1, the average student subjected to this set of policies would achieve beyond that of the very best students found in the state today.

With a six standard deviation improvement, Albert Einstein would look below average. But this, of course, is not credible. It merely highlights the fact that the research base of OPGF – much of it unpublished or in obscure places – does not provide a reliable basis for policy.

Nonetheless, Table 2 understates how truly unbelievable the estimates are and fails to indicate accurately just how fanciful their evidence and subsequent proposals are. Their proposals are actually a mixture of things that have only the initial, one time impact on students and those that would be expected to have continuing and cumulative impacts. For example, the move to full day kindergarten would be done once in the schooling career of each student, and its impact (estimated at an effect size of 0.77) would happen once. Similarly, class size reduction to 15 students would only happen in K-3 and, according to the STAR evidence, would yield a one-time boost in each student's achievement. On the other hand, professional development with classroom coaches would be applied to all teachers on a continuous basis, so that each student would get an ever-larger achievement gain as she went from one trained teacher to another better

[4] See the Odden statement in http://www.ccsso.org/content/pdfs/CostOfImplementing.pdf.
[5] The OPGF report actually suggests in some places that the combination of programs will be more effective than each individually, but this is ignored in the calculations in Table 2.
[6] This is reminiscent of the "six sigma" program in industry that is designed to achieve almost perfect quality from some industrial process. To achieve Six Sigma, a process must not produce more than 3.4 defects per million opportunities. See http://www.isixsigma.com/sixsigma/six_sigma.asp.

trained. Or, tutoring one-to-one would have an estimated impact on disadvantaged students of 0.4-2.5 standard deviations *in each year*, and the OPGF program would introduce tutors for students throughout their schooling. This on-going, across-the-grades program would imply that they would get continual impacts of 0.4-2.5 standard deviations for each year in the program. Over twelve years, the estimates for one-to-one tutoring indicate possibly a 30 standard deviation improvement for disadvantaged students; this would translate into a 10.8 standard deviation improvement in overall average achievement in Washington (recognizing again that disadvantaged students are only 36 percent of the total population).

Table 3 begins with the range of effect sizes that identifies (and reproduced in Table 2) and then calculates in the last two columns the cumulative impact for running programs that go for the entire schooling career of a student (i.e., noted in bold are those that are described by OPGF as being an on-going program with implied cumulative effects). The cumulative impact on Washington schools, following the implications of the evidence cited in the OPGF report, would be *an increase in state achievement of 23 to 57 standard deviations*. If true, it would be a remarkable accomplishment. But of course this is the kind of thing written about in science fiction. It is not at all plausible.

To put this in perspective, national achievement has been assessed by the National Assessment of Educational Progress (NAEP), and this gives a consistent measure of how achievement has changed. Between 1973 and 2004, average mathematics performance of 17-year-olds in the U.S. improved by 0.1 standard deviations. While this shows the challenges facing the nation, and Washington, it also shows the difficulties. National spending per student (adjusted for inflation) more than tripled between 1960 and 2000, thus showing that resources alone were not the issue. Nor is simply uncovering the right program currently buried in the research literature, since much the same rational of "knowing what to do" (albeit by different people) has motivated previous increases in spending.

To be clear, OPGF never say that one can accumulate the impacts of their policies, but there is little other way to interpret their proposals. After all, the evidence they cite indicates the results from instituting the policies once for each student. But, they go on to say that they would not stop there but would provide these programs across all (appropriate) years. It would not make sense to do this unless one thought that even greater gains were possible from a continuous dose of the programs as opposed to a single dose. Of course, the effects might not be simply additive, but, even if subsequent treatments by their programs offered just half of the effect obtained during the first year, the total impacts would still top 12 standard deviations movement for the students of Washington.

It might be nice to think that, had the educators across the nation just been given the evidence that OPGF has accumulated, NAEP scores would go off the current scale. But this is almost certainly the wrong conclusion.

One must believe that they have identified the true impacts of these various reforms and that these are causal impacts. Why might this not be the case?

First, while they do not indicate their procedures for selecting among available research studies, it is important that they did not select studies on the basis of the studies

Table 3. Cumulative Impacts Implied for OPGF On-going Programs are Impossible (Estimated Initial and Cumulative Benefits of Select OPGF Programs)

(Bold indicates on-going program with cumulative impacts.)

	Initial effect on average WA achievement (s.d.)[a]		Cumulative effect for average WA achievement across all grades (s.d.)[b]	
Policy	Minimum	Maximum	Minimum	Maximum
Full day kindergarten	0.77	0.77	0.77	0.77
Class size reduction	0.34	0.34	0.34	0.34
Multi-age classrooms	0	0.5	**0**	**6**
Classroom coaches	1.25	2.7	**15**	**32.4**
Tutoring, 1-1	0.144	0.9	**1.728**	**10.8**
Summer school (1/2 FRL)	0.081	0.081	**0.972**	**0.972**
Embedded technology	0.3	0.38	**3.6**	**4.56**
Gifted and talented (5%)	0.025	0.05	**0.3**	**0.6**
Accelerated instruction (2%)	0.008	0.014	0.008	0.014
TOTAL EFFECTS	3.08	5.897	22.88	56.618

Notes: **Bold indicates on-going program with cumulative impacts**.

a. The initial effects are the estimated range of effect sizes in the last two columns of Table 2. These are the immediate impacts of programs and correspond to the effect sizes in OPGF report when translated to standard deviations for the average achievement for all students in Washington.

b. Cumulative impacts indicate the implied increase in average achievement for Washington students when an individual program is applied to all grades instead of just once. These are calculated as twelve times the one-year impact and are measured in standard deviations.

Source: Author calculations from Odden, Picus, Goetz, and Fermanich (2006), p. 64

results – because that would surely enter bias into their estimated effects. Yet, because their reported results appear at the end of the distribution of estimates across studies, it appears that they are indeed selecting on the results. Specifically, they make no case that the extreme values from among the available results come form better or more reliable studies than the estimates with less magnitude or that are statistically insignificant.

Second, they generally imply that the estimated effect sizes represent causal impacts and not just correlations. This finding is extraordinarily important for policy purposes, because – if implemented through legislative action – one would be relying on the assumption that changing the specific factor would indeed cause an increase in achievement. With few exceptions, little in the methodology suggests that these are really reliable estimates of causal impacts. Take full-day kindergarten, for example, which OPGF estimate to have by itself a 0.77 standard deviation impact on student achievement for advantaged and disadvantaged students alike. OPGF cite a 1997 meta-analysis of John Fusaro that shows such an impact (Fusaro (1997)). But they ignore Fusaro's own strong warning: "A seductive conclusion from these results is that attendance at full-day kindergartens causes students to achieve at a higher level than attendance at half-day kindergartens. It is imperative, however, that we strenuously resist succumbing to such a seduction." Meanwhile, OPGF ignore a large body of literature that shows little impact on advantaged students and smaller impacts on disadvantaged ones, to say nothing of the empirical reality that the 56 percent of students currently attending schools that have full-day kindergarten do not surpass the remaining 44 percent attending schools without full-day kindergarten by anything like a 0.77 margin. Note, for example, that black students and disadvantaged students are currently more likely to attend schools with full-day kindergarten than more advantaged students.

They place considerable weight on the class size estimate from Project STAR in Tennessee (Word et al. (1990)). This is reasonable, because it is designed as a random assignment study conducted in order to extract causal impacts. While there is considerable remaining controversy about the quality of this study and its interpretation (Hanushek (1999a, (1999b)), it remains the stud that is arguably closest to finding a causal impact, and its effect size is small in absolute terms and in terms of effectiveness relative to cost (see below).

Similar comments and concerns apply to the remaining items in the list. There is no apparent attempt to concentrate on reliable estimates of causal policy factors.

3. The suggested approach for choosing among policies guarantees substantial inefficiencies in school spending and operations.

The approach of OPGF to policies is simple: If a program shows a positive effect in one of the studies that they locate, it should be implemented across the state. This approach to policy is one of the most wasteful ones that could be imagined and violates all principles of efficiency in public programs.

The first thing to note is that the various programs suggested by OPGF have very different price tags associated with them. It is hard to tell from their report what prices

might go with each of the programs, because they bury them within the staffing of each prototypical school and do not separately identify the expenditures that might come from introducing each.

Obtaining an accurate estimate of the costs of any of these programs is difficult. The underlying research that they cite seldom identifies the range of expenditures that might go with a program. The impact on overhead and supervisory expenditures, on materials and supplies, and the like is generally even harder to ascertain.

Table 4 provides some rough information about how much their separate programs would cost. These are built on an average expenditure of $7,800 per pupil and an average teacher compensation (salary plus benefits) of $60,000.

In most of the program calculations, the minimum conceivable cost would come from just the compensation that is required for any new teachers. This calculation, of course, includes none of the additional costs of materials, support, maintenance, supervision, and the like. Thus, the expected cost displayed in Table 4 takes these items into account. Expected cost either adds an amount to the teacher compensation (assuming in general that the teacher salary is 60 percent of the full cost) or directly considers how much of the current average expenditure would be increased.

In calculating costs, two separate figures are provided in Table 4. The first is the cost per participant, and the second is the impact on the overall average spending for the system. (These cost calculations parallel the translation of effect sizes in Table 2 from the impact on the specific students in the program to the implications for overall achievement in the state taking into account all students). For example, with full day kindergarten, the cost for each participant is estimated to range between $2,000 and $3,900. The first value, or minimum cost, comes from adding one-half of a teacher (because it is an added half day) and spreading this expense across 15 students in the class, or $2,000. The second, or expected cost, is calculated instead on the basis that each kindergarten student would get another half year of schooling, or $3,900 at the current average of $7,800. Students only do this added kindergarten once in their career and the cost of the remaining 12 years is unaffected. Thus, the increase of the current average spending across all grades is $1/13^{th}$ of the cost during actual participation.[7] These are shown in the final columns of Table 4.

The final two columns show the variation in impacts on average spending per student with a subset of the programs proposed by OPGF. Note that these programs are sometimes calculated as if they apply once to each student once over the schooling career (e.g., kindergarten), sometimes as if they apply every year to all students (e.g., technology), and sometimes to just a special populations (e.g., tutoring for disadvantaged students). The right two columns suggest the range of impact on the average spending per student in Washington from each program. Thus, the full day kindergarten would increase average spending by $154-300 per student, while the K-3 class size reduction would increase average spending by $410-800 per student.

[7] One way to see these calculations is to consider that the total amount currently spent on a graduate with a half day kindergarten would be 7800 x 12.5 = $97,500. With full day kindergarten the spending would be 7800 x 13 = $101,400. The average spending spread across all 13 years in the latter case would be $300 more than in the former.

Table 4. Added Per Student Costs of OPGF Programs Vary Widely

Policy	Cost per participant[a] Minimum[c]	Cost per participant[a] Expected[d]	Impact on average spending per pupil in WA schools[b] Minimum[c]	Impact on average spending per pupil in WA schools[b] Expected[d]
Full day kindergarten	$2,000	$3,900	$154	$300
Class size reduction	$1,333	$2,600	$410	$800
Multi-age classrooms	$0	$0	$0	$0
Classroom coaches	$300	$500	$300	$500
Tutoring, 1-1	$600	$1,000	$216	$360
Summer school (1/2 FRL)	$1,000	$1,667	$180	$300
Embedded technology[e]	$250	$250	$250	$250
Gifted and talented (5%)[f]	$250	$250	$250	$250
Accelerated instruction (2%)[g]	($7,800)	($7,800)		
TOTAL ADDED COSTS per pupil			$1,760	$2,760
PERCENTAGE INCREASE[h]			22.6%	35.4%

Notes:

a. Cost per participant estimates the additional costs that go directly to the program on a per pupil basis.

b. Impact on average spending per pupil in Washington schools recognizes that some programs apply only to a subset of students and calculates how much the specific proposal would add to the average spending per pupil in the state (assumed to be $7,800/student).

c. Minimum cost in general calculates just the cost of any added teachers (where total compensation is assumed to be $60,000). For example, full day kindergarten requires one-half of a teacher for 15 students, or $2,000/participant.

d. Expected cost recognizes that there are overhead costs, supervision requirements, training, material costs, and so forth that increase when new programs and teachers are added. When programs can be translated into similar normal operations, these are calculated from average per pupil spending; e.g., full day kindergarten is equivalent to adding one-half of a new student, or $3,900. When programs cannot readily be put in terms of overall spending, teacher compensation is assumed to be 60 percent of the total; e.g., tutoring.

e. Embedded technology uses the OPGF recommendation of $250 per student per year.

f. Gifted and talented is assumed to be the OPGF funding for all students of $250 per year (benefits assumed to apply to 5% of the students).

g. Accelerated instruction resulting in skipping grades would save one year of spending (not translated to overall spending).

h. Percentage increase is total increase in per pupil spending compared to $7,800 per pupil.

Some programs have no obvious costs. For example, multi-age classrooms have no obvious costs, and might reasonably be taken as free. Similarly, changes in curriculum do not in general have significant added costs (past, say, an initial training period). Other programs, such as skipping grades, would actually be a cost saving, since the student would spend 12 rather than 13 years in the system.

The programs in the table suggested by OPGF, if fully applied, would be estimated to increase the average spending in Washington by $1,760-2,760 per student, or 23-35 percent (see bottom row). This estimate comes from summing the added cost of each program across the identified programs. (Note that a number of other components of their proposed policies are not included, so these increases apply to just the specifically identified programs).

The important aspect of separating out the costs for each program is that one can immediately see the variation that exists and can make judgments about where money is better (more efficiently) spent. OPGF on the other hand strongly propose that all programs, regardless of cost, be simultaneously undertaken. And their method of calculating costs – going through their prototypical schools – gives no information on the efficiency of investing in the various components.

Tables 5 and 6 clearly show that there are very different achievement effects of money allocated to their different proposals. Table 5 employs the minimum expected effect size for each program (from Table 2), while Table 6 employs the maximum. These estimated effects are then compared to the overall system costs in Table 4. The combination of Table 5 and Table 6 gives the range of investment returns implied by the OGPF report. (Note: because the cumulative effects are not credible, these calculations rely just on the initial effects).

The final two columns give the improvements in student achievement (measured in standard deviations) that could, by the OPGF estimates of benefits, be expected for each $100 addition to spending per pupil on each of the separate programs.

By their analysis, each $100 spent on classroom coaches would be expected to yield at least 0.25 standard deviations gain in achievement, very similar to that for full day kindergarten (Table 5). On the other hand, their class size reduction proposal would yield only *one-sixth* that gain, or 0.04 standard deviations, an effect very similar to that for one-to-one tutoring.

Using the upper range of their effect estimates (Table 6) yields an even larger disparity in the payoff to investments in the different programs. One hundred dollars on classroom coaches would, by these alternative estimates, yield over one-half standard deviation in student achievement, and one-to-one tutoring would have one-quarter standard deviation improvement. According to their estimates, some of their favored programs (such as classroom coaches are more than ten times as cost efficient as others such as class size reduction for K-3). Yet the OPGF advice is to do them all without regard to cost or effectiveness.

Unless there are unlimited funds to spend on educational programs, it would not make sense to put the money in all programs, regardless of cost. The programs they

Table 5. Some OPGF Proposals on a Cost of Investment Basis Have Ten Times the Impact of Others (Achievement Benefits from $100 Increase in Program Spending Assuming Minimum Initial Impact)

Policy	Minimum Initial Impact[a]	Cost per student of fully implemented program[b]		Student achievement gains per $100 spending[c]	
		Minimum cost estimate	Expected cost estimate	With min costs	With expected costs
Full day kindergarten	0.77	$154	$300	0.501	0.257
Class size reduction	0.34	$410	$800	0.083	0.043
Multi-age classrooms	0	0	0	0	0
Classroom coaches	1.25	$300	$500	0.417	0.250
Tutoring, 1-1	0.144	$216	$360	0.067	0.040
Summer school (1/2 FRL)	0.081	$180	$300	0.045	0.027
Embedded technology	0.3	$250	$250	0.120	0.120
Gifted and talented (5%)	0.025	$25	$25	0.100	0.100

Notes:

a. Minimum initial impacts are given in column 3 of Table 2. See explanations there.

b. Cost per student (averaged across all students in the state) from columns 4 and 5 of Table 4.

c. The impact in column 1 is divided by the cost estimates in columns 2 and 3 (in $100) to get the achievement gain per $100 invested in each program.

Table 6. The Variation in Returns to OPGF Programs is Larger with Their Maximum Effect Size Estimates (Achievement Benefits from $100 Increase in Program Spending Assuming Maximum Initial Impact)

Policy	Maximum Initial Impact[a]	Cost per student of fully implemented program[b]		Student achievement gains per $100 spending[c]	
		Minimum cost estimate	Expected cost estimate	With min costs	With expected costs
Full day kindergarten	0.77	$154	$300	0.501	0.257
Class size reduction	0.34	$410	$800	0.083	0.043
Multi-age classrooms	0.5	$0	$0	Infinite[a]	Infinite[a]
Classroom coaches	2.7	$300	$500	0.900	0.540
Tutoring, 1-1	0.9	$216	$360	0.417	0.250
Summer school (1/2 FRL)	0.081	$180	$300	0.045	0.027
Embedded technology	0.38	$250	$250	0.152	0.152
Gifted and talented (5%)	0.05	$25	$25	0.200	0.200

Notes:

a. Maximum initial impacts are given in column 4 of Table 2. See explanations there.

b. Cost per student (averaged across all students in the state) from columns 4 and 5 of Table 4.

c. The impact in column 1 is divided by the cost estimates in columns 2 and 3 (in $100) to get the achievement gain per $100 invested in each program.

identify have very different returns on any given spending. Putting money equally into low return and high return programs does not make economic sense.

Moreover, pursuing this undifferentiated kind of policy risks locking in a set of bad resource decisions.

4. The programs creating the underlying OPGF rational for funding likely bear little relationship to the actions districts would take if provided with the funding changes suggested.

Their identification of programs permits them to give some rational, however thin, to the development of prototype schools. But there is no sense that, if this money were sent out to districts, the districts would spend it on the items identified by OPGF. In fact, we should be quite certain that they would not. For example, between 1990 and 2001, U.S. expenditures per pupil grew by over 15 percent (in real terms, above inflation). Yet schools did not invest in the programs identified by OPGF, even though much of the research cited was available.

Their report notes that the most important issue is *how* the money is spent. They claim to have identified from research particularly productive uses of funds. One must then ask, "why weren't these programs followed more often?" And, "why haven't we seen the results in terms of achievement?"

If it is just that school officials are ignorant of the research, then it makes sense to provide them with the answers along with the funds, or even before the funds. In fact, if schools knew how to spend their money better, they could get substantial gains out of existing funds, even if these effects are not described in the OPGF report. In fact, in earlier writing, Odden emphasized the importance of improving the productivity of existing funding; see, for example, Odden and Busch (1998).

If the objective were to get schools to use these precise programs, it could be accomplished more reliably by insisting through statute or regulations that they institute these specific programs. But this would make sense only if one was very certain that the programs would work universally.

Given the serious doubts about the efficacy of the policy recommendations, it would not make sense to lock them into school operations. As the prior discussion makes clear, there are many reasons to be suspicious of the research and of whether these programs would actually have the effects identified.

Yet, if these programs are not instituted in the schools, perhaps the added funding contained in the OPGF prototype schools is not appropriate. The added resources are predicated on these specific programs. In other words, why not provide schools with funds for less expensive programs (ones that they also might not employ)? Basing funding on a hypothetical that is unrelated to what might happen means that there are a great number of other potential programs that just as legitimately could be used as the basis for funding.

A major problem also arises from looking for empirical evidence on the success of the OPGF proposals. There exists no actual demonstration that these programs will be

instituted or that they will be effective. For example, virtually the same programs with the same rational have been proposed to Arkansas[8] and to Wyoming[9] by this group. Arkansas in fact raised is spending to meet the proposals. Is there any evidence that Arkansas districts, when provided the added money of such a proposal, put in place anything like the "evidence-based" proposal? Is there any evidence of improved student performance? No such evidence exists.

But, note also that the OPGF report never predicts that student achievement will actually increase with the funding they propose. They only assert that the gains would be possible, based on their reading of past research. By phrasing everything in terms of the "opportunity to learn," their programs can never in the future be judged against the actual achievement.

[8] Odden, Picus, and Fermanich (2003)
[9] Odden et al. (2005)

5. The funding structure recommended does little to change the incentives facing schools to use money wisely or to improve student performance.

OPGF rightfully point out that finance and school policy are closely intertwined and that improvements in the schools must recognize this. Following this message through their work is nonetheless difficult. They identify "good" programs but only as a means to adding up resources. Their policy is essentially to send the money to districts largely as identified by their prototype models. They do not propose a finance system that would support other school policies in the improvement of student achievement.

If there is one lesson of the past decades of reform efforts, it is that funds sent to districts without regard to performance do not have, on average, a powerful impact on student learning.[10] Yet none of their finance proposals include any incentives to improve student performance.[11] There are no rewards for student gains in achievement. There are no sanctions for a lack of student learning. There is only a calculation of how much money should be sent to districts with the underlying argument 'it should be a lot if Washington wishes large improvements in student achievement.'

The reality is that there are noticeable gaps in our knowledge of what works and how improvements can be gained on a large scale. A great deal has been written about how to proceed in this policy environment, and it is not possible to provide detail here.[12]

Nonetheless, in broad terms, this situation seems to call for three things:

First, there should be a strong accountability system that identifies accurately which schools are doing well and which are not. Washington has established a strong accountability system, so it has the basis for improvement.

Second, there should be a finance system that includes direct rewards for success as measured by student learning. This incentive structure should be combined with other forms of incentives, such a parental choice of schools, so that pressure is put on to improve performance. Incentives should also include direct incentives to students to improve their achievement.

Third, there should be an active learning program that provides feedback to schools and districts on what is working. This idea implies that, instead of rushing in with a set of programs with questionable research base, a program of broader experimentation and learning should be instituted in the schools.

---

[10] For example, the large expansion of programs and spending following *A Nation at Risk* (National Commission on Excellence in Education (1983)) yielded few tangible results; see Peterson (2003) and Hanushek (2003).

[11] In an earlier time, Picus (1992) was an early adherent to the idea that incentives were important to improving schools.

[12] See, for example, Hanushek (1994) or Moe (2001). For specific application to school finance policies, see http://www.korettaskforce.org/books/texas/.

The latter two elements are not currently central to policies in Washington. Reform along these lines has a much higher probability of success in raising the achievement of Washington students than does the proposals of the OPGF report.

Conclusions

The importance of improving schools in the U.S. and in the State of Washington cannot be denied. The importance of high quality education for both individuals and society as a whole is well established. That objective should not, however, motivate bad public policies.

The evidence presented by OPGF is simply not sufficient to guide policy. The research relied on by OPGF is on average of low quality. It is also very selectively chosen from available work, apparently to make the expected outcomes look large.

The easiest demonstration of the quality of the evidence put forward is the incredible outcomes that are attributed to the identified programs. It is unreasonable to believe that, without changing any of the incentives and structure of the existing schools but by simply adding substantially more resources, the schools of Washington will improve at all – let alone that the average student will jump above today's best student. This is, nonetheless, what the OPGF report implies.

A major problem that also comes from evaluating their set of proposals is that they have no regard for efficiency of public expenditures. Their advice is simply to spend whatever it takes, even if a number of programs are very wasteful. This is a fundamentally flawed view of how to make good policy decisions. It is, however, a basic building block of the available range of "adequacy studies."

Washington will be well served by recognizing that silver bullets do not exist. There is not a stockpile of good ideas undiscovered by policy makers and school personnel but known to a small cadre of consultants.

Instead, it should be pursuing a system of finance that introduces strong accountability, incentives for student performance, and a program of learning from experience.

References:

Ehrenberg, Ronald G., Dominic J. Brewer, Adam Gamoran, and J. Douglas Willms. 2001a. "Class size and student achievement." *Psychological Science in the Public Interest* 2,no.1 (May):1-30.

———. 2001b. "Does class size matter?" *Scientific American*(November):83-89.

Fusaro, J. A. 1997. "The effect of full-day kindergarten on student achievement: A meta-analysis." *Child Study Journal* 27,no.4:269-277.

Hanushek, Eric A. 1994. *Making Schools Work: Improving Performance and Controlling Costs*. Washington, D.C.: The Brookings Institution.

———. 1999a. "The evidence on class size." In *Earning and learning: How schools matter*, edited by Susan E. Mayer and Paul E. Peterson. Washington, DC: Brookings Institution:131-168.

———. 1999b. "Some findings from an independent investigation of the Tennessee STAR experiment and from other investigations of class size effects." *Educational Evaluation and Policy Analysis* 21,no.2 (Summer):143-163.

———. 2003. "The failure of input-based schooling policies." *Economic Journal* 113,no.485 (February):F64-F98.

———. 2006. "Science Violated: Spending Projections and the “Costing Out” of an Adequate Education." In *Courting Failure: How School Finance Lawsuits Exploit Judges’ Good Intentions and Harm Our Children*, edited by Eric A. Hanushek. Stanford: Education Next Books:257-311.

———. 2007. "The alchemy of 'costing out' an adequate education." In *School Money Trials: The Legal Pursuit of Educational Adequacy*, edited by Martin R. West and Paul E. Peterson. Washington: Brookings:77-101.

Mishel, Lawrence, and Richard Rothstein, eds. 2002. *The class size debate*. Washington, DC: Economic Policy Institute.

Moe, Terry M., ed. 2001. *A primer on America's schools*. Stanford, CA: Hoover Institution Press.

National Commission on Excellence in Education. 1983. *A nation at risk: the imperative for educational reform*. Washington, D.C.: U.S. Government Printing Office.

Odden, Allan, Lawrence O. Picus, and Mark Fermanich. 2003. *A Evidence-Based Approach to School Finance Adequacy in Arkansas*. Lawrence O. Picus and Associates, (September).

Odden, Allan, Lawrence O. Picus, Michael Goetz, and Mark Fermanich. 2006. *An Evidenced-Based Approach to School Finance Adequacy in Washington*. Lawrence O. Picus and Associates, (June 30).

Odden, Allan, Lawrence O. Picus, Michael Goetz, Mark Fermanich, Richard C. Seder, William Glenn, and Robert Nelli. 2005. *An evidence-based approach to recalibrating Wyoming's block grant school funding formula*. Lawrence O. Picus and Associates, (November 30).

Odden, Allan R., and Carolyn Busch. 1998. *Financing Schools for High Performance: Strategies for Improving the Use of Educational Resources*. San Francisco, CA: Jossey-Bass Publishers.

Peterson, Paul E., ed. 2003. *Our schools and our future: Are we still at risk?* Stanford, CA: Hoover Press.

Picus, Lawrence O. 1992. "Using incentives to promote school improvement." In *Rethinking school finance: an agenda for the 1990s,* edited by Allan R. Odden. San Francisco: Jossey-Bass Publishers:166-200.

Word, Elizabeth, John Johnston, Helen Pate Bain, B. DeWayne Fulton, Jayne Boyd Zaharies, Martha Nannette Lintz, Charles M. Achilles, John Folger, and Carolyn Breda. 1990. *Student/teacher achievement ratio (STAR), Tennessee's K-3 class size study: Final summary report, 1985-1990*. Nashville, TN: Tennessee State Department of Education.