



REVIEW OF *ORGANIZING SCHOOLS TO IMPROVE STUDENT ACHIEVEMENT*

Reviewed By

Patrick J. McEwan

Wellesley College

November 2011

Summary of Review

This report carefully reviews high-quality empirical evidence from the last several years on the test score effects of three approaches to modifying the organization of schools: (1) starting schools later in the morning, (2) favoring K-8 grade configuration instead of junior high or middle school configurations, and (3) increasing teacher specialization by grade and subject. It estimates the earnings benefits of each intervention and, for interventions (1) and (2), compares monetary benefits to costs. The report concludes that benefits outweigh costs, although the rough cost estimates suggest that better data are required to draw definite conclusions. The report's main conclusion is that organizational reforms deserve a more prominent place in education debates, and that individual school districts should carefully consider them alongside more popular reform options. The review points to a few shortcomings but concludes that the report's analyses are solid and helpful and that the results are presented carefully and cautiously.

Kevin Welner

Editor

William Mathis

Managing Director

Erik Gunn

Managing Editor

National Education Policy Center

School of Education, University of Colorado
Boulder, CO 80309-0249
Telephone: 303-735-5290
Fax: 303-492-7090

Email: NEPC@colorado.edu
<http://nepc.colorado.edu>

Publishing Director: Alex Molnar



This is one of a series of Think Twice think tank reviews made possible in part by funding from the Great Lakes Center for Education Research and Practice. It is also available at <http://greatlakescenter.org>.

This material is provided free of cost to NEPC's readers, who may make non-commercial use of the material as long as NEPC and its author is credited as the source. For inquiries about commercial use, please contact NEPC at nepc@colorado.edu.

REVIEW OF ORGANIZING SCHOOLS TO IMPROVE STUDENT ACHIEVEMENT

Patrick J. McEwan, Wellesley College

I. Introduction

Recent debates over education policy have focused on the merits of headline-grabbing reforms such as charter schools, school accountability, and the recruitment and motivation of teachers. In contrast, a new report from the Hamilton Project at the Brookings Institution sheds light on important but overlooked policies related to the organization and management of schools.¹ It highlights a recent flurry of empirical research that has examined the impact on students of modifying school start times (from earlier to later), of reconfiguring schools (from middle schools to K-8), and of modifying teacher assignments (from less-specialized in grade and subject assignments to more specialized).

Organizing Schools to Improve Student Achievement reviews the extant research on the effects of these three reforms on student achievement, estimates the eventual earnings benefits, and compares the benefits to the available evidence on costs. The report concludes that benefit-cost ratios are positive for two of the reforms: later start-time and K-8 interventions. This review identifies some caveats to these findings, most notably the absence of detailed cost data on the interventions. Overall, however, the report is a nuanced and cautious review of neglected areas of policy. Its main conclusion is justified: that organizational reforms should be “a more prominent part of the conversation on how to raise student achievement” (p. 2).

II. Findings and Conclusions of the Report

The report summarizes recent empirical studies on three potential interventions in the organization and management of schools: (1) instituting later start times for middle and high schools, (2) configuring schools differently by encouraging K-8 schools rather than junior high and middle school configurations, and (3) managing teacher assignments differently by increasing teachers’ grade-level or subject specialization. Based on the available empirical research and their own calculations, the report concludes that the three interventions have effects on student test scores of 0.1, 0.175, and 0.02 standard deviations, respectively (p. 6). Such effect sizes, expressed in standard deviation units, are a common way of expressing the magnitude of test score effects. By way of comparison, a

well-known overview of the Tennessee STAR class size reduction plan reported an effect of 0.15 standard deviations.²

The report conducts a simple cost-benefit analysis, converting each effect size into monetary benefits and comparing it with an estimate of costs. In the case of modifying the start time of schools, the estimated ratio of benefits to costs is large (\$9 or more of benefit for each \$1 of costs). It is even larger for converting schools to a K-8 configuration (from \$40 to \$200 of benefits per \$1 of costs, with the upper and lower figures derived from two different studies). The report does not include a cost estimate for the teacher assignment intervention.

Overall, the report concludes that organizational interventions have potential to increase student achievement at a comparatively low cost, although the recommendations acknowledge that the effects, costs, and feasibility of specific interventions may vary by district. The authors encourage piloting of interventions where they might be expected to have greater benefits or lower costs. More generally, the report concludes that such interventions have been relatively neglected in policy debates. Further, the report “encourage[s] school, district, and state education leaders to consider these reforms carefully, and...to make the management organization, and operation of schools an integral part of the conversation on how to raise student achievement nationwide” (p. 19).

III. The Report’s Rationale for Its Findings and Conclusions

The report is divided into three main sections, each of which reviews the research on one of the three organizational interventions and then estimates preferred effect sizes.³ In each case the report applies a secondary analysis to convert effect sizes into monetary benefits (as described below). Finally, these benefits are compared with estimates of monetary costs. The report summarizes these results and the overall benefit-cost ratios (p. 6).

Effect Sizes of Interventions

Chapter Two documents the substantial variation in school start times. It describes the role that student transportation schedules often play in such decisions. Specifically, minimizing the costs of bus transportation sometimes involves staggered school start times. This allows more intensive use of the bus fleet. In particular, so-called tiered busing systems usually involve earlier start times for middle and high schools relative to elementary schools.

In the absence of randomized experiments, the report discusses several “natural” experiments available in the economics literature. The report’s preferred effect size is based on two papers: a study of the U.S. Air Force Academy, in which students’ start times varied due to random assignment to classes and a policy shift in start times;⁴ and a study of a large North Carolina district whose students experienced changes over time in start times due to rapid growth and changes in school attendance boundaries.⁵ The report assumes that a one-year effect size of 0.1—roughly half-way between the results of the two

papers—will be cumulative over time, but that prior effects will “fade out” (p. 21). The authors multiply 0.1 by 7 years (from grades 6-12), but divide it by 4 to reflect fade-out, yielding a final effect size of 0.175 (p. 6). The report also cites a third study that finds no effects of a later start time.⁶

Chapter Three, on school grade configuration, notes that widespread use of junior high and middle schools is a relatively recent phenomenon in public school districts. Despite their ubiquity, the impact of these school configurations on student outcomes—as

The report is a nuanced and cautious review of neglected areas of policy.

compared with K-8 configurations—is not well understood. The report describes two recent papers using data from New York City⁷ and Florida.⁸ Both papers compare the achievement of students who move to a middle school or junior high school in grades 6 or 7 with that of students who attended a K-8 school. In each study, the test scores of students who attend grade 6 in a middle school (or grade 7 in a junior high) decline relative to students who attend the same grades in a K-8 school. The declines persist until grade 8. The report’s preferred effect size of 0.1 represents the increased achievement—by grade 8—of attending a K-8 school instead of a middle school. It is roughly in between each study’s estimated effects.

Finally, Chapter Four considers the broad issue of teacher assignments, focusing on teacher specialization in particular grades or in particular subjects. One cited study uses North Carolina data and reports that teachers assigned to the same grade level improve more quickly than teachers who switch grade assignments.⁹ Another study, also with North Carolina data, shows a high but not perfect correlation of approximately 0.7 between teacher effectiveness in English and teacher effectiveness in math. This, the authors contend, implies that students could benefit if teachers only taught the subject in which they are most effective.¹⁰ To arrive at a preferred effect size, the report’s authors apparently conducted their own empirical analysis of New York City data, although the results are not included or described in detail. Overall, the report concludes that complete elementary teacher specialization by grade and subject would result in math gains of 0.02 standard deviations, “with smaller effects in reading” (p. 18).

From Effect Sizes to Monetary Benefits

For each intervention, the effect sizes are converted to monetary benefits using a procedure that relies on additional assumptions (p. 10). The report assumes that a 1.0 standard deviation in test score will increase future earnings by 8% (that assumption is discussed later in this review). It further assumes that mean earnings of workers in 2008, assuming 1% real wage growth over time, provide a reasonable estimate of how earnings would evolve in the absence of test score increases. Finally, to calculate the present value of earnings benefits, the report assumes a discount rate of 4%. Given these assumptions, the report finds that the two favored interventions might increase the lifetime present

value of earnings by anywhere from \$2,000 (teacher specialization) to \$17,500 (later start times).

Costs

The report also estimates the costs of the two favored interventions. In the case of school start times, the report bases its estimates on rough figures reported in the transportation newsletter of the Wake County Public School System (WCPSS).¹¹ Between converting from a single-tier bus system in 1993 to a three-tier system with staggered scheduling, the WCPSS reported in 2004 that it had “...saved taxpayers more than \$100 million that would be needed to purchase and operate enough buses to allow all schools to have their own buses and open and close at the same time” (p. 3). To convert back to a single-tier system with a later start time, the report cites an unpublished manuscript suggesting that it would cost approximately \$150 per student per year to move two-thirds of WCPSS’s 100,000 students to a later start time ($\$100,000,000 \div 10 \text{ years} \div [2/3 \times 100,000 \text{ students}]$). The report estimates a total per-student cost over grades K-12 of \$1,950 ($\150×13). It further suggests that costs might be lower in cases where districts simply reversed the start times of schools in a typical three-tier system (with high school starting later), although this could create uncertain costs for elementary children who would then start at the earliest times.¹²

The report estimates the annual per-pupil cost of converting from a K-5 to a K-8 configuration to be roughly \$50 to \$250 per student per year. The low end is based upon a cost study conducted by the authors in Denver schools that accounted for increased classroom infrastructure and materials, as well as increased transportation costs. The higher end is based upon a study in New York City, where additional classroom space was constructed or remodeled. In both cases, few details are reported on data or calculations (although, importantly, the authors amortized infrastructure costs across time).

The report does not directly estimate costs of teacher specialization, and the authors acknowledge that it may vary substantially. Some grade-level or subject specialization might be accomplished with low costs, but complete specialization implied by the effect sizes could involve extensive teacher transfers and higher costs. The report concludes that “the scenario of complete specialization may not pass a cost-benefit test and is almost certainly politically infeasible” (p. 18).

IV. The Report’s Use of Research Literature

Effect Sizes

To estimate the effects and costs of organizational interventions, the report uses empirical studies from the last several years of school start times, grade configurations, and teacher assignments. With one exception—the U.S. Air Force study of start times—the studies cannot rely on randomized assignment to identify the effects of interventions. However,

they all use administrative, longitudinal data on students and schools, including datasets from Florida, North Carolina, and New York City. The availability of such data facilitates the application of statistical techniques that more credibly control for unobserved features of schools or students that are correlated with interventions. These approaches, however, might not fully control for biases in the estimates of program effects. Each of the cited papers carefully explores threats to internal validity such as selection bias, and each uses alternative statistical specifications that probe for such biases.

Costs

Ideally, a cost study should identify the incremental resources consumed in an intervention—from personnel to facilities—and systematically attach a price to each resource.¹³ It should strive to identify costs to the implementing agency, such as a school district, and also to other stakeholders such as families. Finally, it should ensure that costs (and benefits) are expressed in constant prices of a single year, and that costs (and benefits) are appropriately discounted to their present value.

The report relies on a cost estimate of modified start times from external sources.¹⁴ As those sources reveal, the estimate is a “back-of-the-envelope” calculation for which the method, data, and assumptions are not fully specified. It does not include costs to families, although the report carefully notes that such costs may exist (such as child-care costs incurred by a later start time). Overall, therefore, we must interpret the cost estimates with some caution.

V. Review of the Report’s Methods

The report is primarily a review of other studies, although it applies the above-described methods to arrive at estimates of effects, benefits, and costs. Overall, the report clearly states its methods as well as appropriate caveats.

Effect sizes

The effect size calculation for the start-time intervention relies on assumptions about how effects accumulate and “fade out” over time (see section III above). For example, the assumption about “fade-out” is based on empirical research suggesting that the achievement effects of having an effective teacher may decline by three-quarters over three years. Ultimately one cannot verify, without further empirical study, whether the assumption is valid in the specific context of school start times. Also, the preferred effect size estimate does not incorporate an empirical study of start-time in Minnesota that reported zero effects.¹⁵ The authors argue that the Minnesota findings are less credible because the learning outcomes are limited to ACT scores and because of concerns about selection bias. But, even assuming a much smaller effect size of 0.1, the benefits of the intervention outweigh its costs, conditional on other assumptions.

Regarding the third intervention, the report simulates the effect of teacher re-assignment, focusing an “extreme” case where New York City elementary teachers move to full specialization in a grade and subject. The simulation relies on unreported estimates by the report’s authors of the returns to teaching experience. The report would be stronger if these results were reported or cited.

Benefits

The report converts non-monetary test score effects into monetary earnings benefits by relying on assumptions. For example, it assumes that a 1.0 standard deviation increase in test scores is associated with an 8% increase in earnings. This estimate is taken from other empirical studies that estimate the statistical association between test scores and earnings, controlling for other variables. There is some precedent for conducting such an analysis. For example, a recent book applied similar methods and data to conduct a benefit-cost analysis of interventions that affect test scores, among other outcomes.¹⁶ Like the report, it used prior empirical work to inform assumptions about the empirical relationship between test scores and earnings, and it used household survey data to calculate workers’ earnings. While not perfect, such analyses provide the only available means of making “apples-to-apples” comparisons between monetary benefits and costs. Without doing so, one cannot judge whether an intervention is worth the investment.

Costs

The report conducts an analysis of the costs of converting middle schools to K-8 configurations, relying on budget data from Denver and New York City. The analysis seems carefully done, in that it includes a range of resources, such as materials, facilities, and transportation, and amortizes the costs over their useful life (to avoid imputing, for example, the entire cost of a classroom to a single cohort of students). It also highlights the importance of context in determining costs, since the upper range of costs reflect the costs of classroom renovation in New York City. The report does not include enough detail to fully evaluate or replicate the estimates, however.¹⁷

VI. Review of the Validity of the Findings and Conclusions

The report concludes that “...the ratio of benefits to costs is 9 to 1 for later school start times and 40 to 1 for middle school reform” (p. 2). It does not calculate a ratio for teacher reassignment, for which the effects are estimated to be quite small. The validity of these findings depends on the validity of individual estimates of effects, benefits, and costs.

The report’s discussion of effects is thorough and nuanced, and it draws from high-quality empirical studies conducted in the last several years. The report’s estimates of benefits are more speculative, if only because they rely on additional assumptions about the relationship between test-score effects and earnings. Though similar assumptions are also made in other cost-benefit analyses of education interventions, they nonetheless introduce

uncertainty into the conclusions. The cost data are the weakest link in the analytical chain. The cost estimate for the start-time intervention is based on a rough estimate from a school district; the cost estimate for grade configuration is based on data that are not fully described by the authors; and the costs of teacher assignment are not known.

Overall, the authors argue that their estimates of cost-benefit ratios are conservative, although this review has pointed out instances where even more conservative assumptions could be made, especially regarding costs. That said, two points should be emphasized. First, the reported benefit-cost ratios are quite large, and larger than “traditional” interventions such as class size reduction.¹⁸ The results might persist under more conservative assumptions, although more detailed sensitivity analysis would be required to draw this conclusion. Second, the report never claims to have identified a “correct” benefit-cost ratio for a particular intervention, and it identifies caveats to the interpretation of the results.

VII. Usefulness of the Report for Guidance of Policy and Practice

The report does not advocate for a blanket implementation of an intervention based on a single benefit-cost ratio. Rather, it calls for additional pilot studies in districts, especially those likely to have the greater incremental benefits or lowest costs. For example, research suggests that later start-times may have larger effects on more disadvantaged students. Likewise, it may have lower costs in districts that already have a single-tier bus schedule (or slack in their existing use of transportation). The incremental costs of converting to a K-8 configuration could vary substantially depending on the existing district infrastructure. Teacher assignment might be accomplished with relative ease in large districts where high rates of grade switching occur because of personal preference rather than a specific pedagogical strategy. It may be costly in districts where extensive teacher transfers must occur. In short, the report’s evidence supports its main conclusion that organizational interventions deserve careful consideration alongside more hotly debated or popular interventions such as charter schools or computer-assisted instruction.

Notes and References

- 1 Jacob, B. A., & Rockoff, J. E. (2011). *Organizing Schools to Improve Student Achievement: Start Times, Grade Configurations, and Teachers Assignments*. Washington, DC: The Hamilton Project at the Brookings Institution. Retrieved November 15, 2011, from http://www.brookings.edu/~media/Files/rc/papers/2011/09_organize_jacob_rockoff/092011_organize_jacob_rockoff_paper.pdf.
- 2 Schanzenbach, D. W. (2007). What Have Researchers Learned from Project STAR? In T. Loveless & F. M. Hess, eds., *Brookings Papers on Education Policy 2006/2007* (205-228). Washington, DC: Brookings Institution Press.
- 3 In fact, the original empirical studies usually report several effect sizes. The “preferred” effect size in the report represents the judgment of the report’s authors about the most credible or reasonable magnitude, given the weight of the evidence.
- 4 Carrell, S. E., Maghakian, T., & West, J. E. (2011). A’s from Zzzz’s? The Causal Effect of School Start Time on the Academic Achievement of Adolescents. *American Economic Journal: Economic Policy* 3(3), 62-81.
- 5 Edwards, F. (2011). *Early to Rise: The Effect of Daily Start Times on Academic Performance*. Unpublished manuscript, University of Illinois at Urbana-Champaign.
- 6 Hinrichs, P. (2011). When the Bell Tolls: The Effects of School Starting Times on Academic Achievement. *Education Finance and Policy* 6(4), pp. 1-22. While noting the zero findings, the report argues that the ACT scores used as the main outcome variable may not fully reflect student learning.
- 7 Rockoff, J. E., & Lockwood, B. B. (2010). Stuck in the Middle: Impacts of Grade Configuration in Public Schools. *Journal of Public Economics* 94(11-12), 1051-1061.
- 8 Schwerdt, G., & West, M. R. (2011). The Impact of Alternative Grade Configurations on Student Outcomes through Middle and High School. Unpublished manuscript, Harvard University.
- 9 Ost, B. (2010). How Do Teachers Improve? The Relative Importance of Specific and General Human Capital. Unpublished manuscript, Cornell University.
- 10 Condie, S., Lefgren, L., & Sims, D. (2011). Heterogeneous Match Quality and Teacher Value-Added: Theory and Empirics. Unpublished manuscript, Brigham Young University.
- 11 Wake County Public School System, Department of Transportation. (2004). *The Wheels of Education* 1(1).
- 12 Edwards, F. (2011). *Early to Rise: The Effect of Daily Start Times on Academic Performance*. Unpublished manuscript, University of Illinois at Urbana-Champaign. The Edwards study explored this issue, but elementary schools do not start at the earliest times in their data, preventing a careful answer to the question of how young children would be affected by reversing a three-tier system.
- 13 See, e.g., Levin, H. M., and McEwan, P. J. (2001). *Cost-Effectiveness Analysis: Methods and Applications*. Thousand Oaks, CA: Sage.
- 14 Wake County Public School System, Department of Transportation. (2004). *The Wheels of Education* 1(1).

Edwards, F. (2011). *Early to Rise: The Effect of Daily Start Times on Academic Performance*. Unpublished manuscript, University of Illinois at Urbana-Champaign.

15 Hinrichs, P. (2011). When the Bell Tolls: The Effects of School Starting Times on Academic Achievement. *Education Finance and Policy* 6(4), 1-22.

16 Levine, P. B., & Zimmerman, D. J. (2010). *Targeting Investments in Children: Fighting Poverty When Resources are Limited*. Chicago: University of Chicago Press. Also see studies of the Tennessee STAR experiment, cited in Schanzenbach and by the report.

17 The start time cost analysis is discussed above, under the “Literature Review” heading.

18 Schanzenbach, D. W. (2007). What Have Researchers Learned from Project STAR? In T. Loveless & F. M. Hess, eds., *Brookings Papers on Education Policy 2006/2007* (205-228). Washington, DC: Brookings Institution Press.

DOCUMENT REVIEWED:

Organizing Schools to Improve Student Achievement: Start Times, Grade Configurations, and Teacher Assignments

AUTHORS:

Brian A. Jacob and Jonah E. Rockoff

PUBLISHER/THINK TANK:

The Hamilton Project at the Brookings Institution

DOCUMENT RELEASE DATE:

September 2011

REVIEW DATE:

November 29, 2011

REVIEWER:

Patrick J. McEwan, Wellesley College

E-MAIL ADDRESS:

pmcewan@wellesley.edu

PHONE NUMBER:

(781) 283-2987

SUGGESTED CITATION:

McEwan, P. J. (2011). *Review of "Organizing Schools to Improve Student Achievement: Start Times, Grade Configurations, and Teacher Assignments."* Boulder, CO: National Education Policy Center. Retrieved [date] from <http://nepc.colorado.edu/thinktank/review-organizing-schools>.