

WWC Review of the Report “School Turnarounds: Evidence From the 2009 Stimulus”¹

The findings from this review do not reflect the full body of research evidence on *School Improvement Grants*.

What is this study about?

The study examined the effects of *School Improvement Grant (SIG)* eligibility and receipt on school-wide achievement of students in low-performing California public schools.

The study was conducted using a regression discontinuity design. The sample consisted of 2,892 schools, including 168 schools that were eligible for a *SIG* in 2010–11 as a result of being identified as “persistently lowest achieving” (PLA). To be identified as PLA, schools needed to fall below two thresholds: (1) “lowest achieving” (defined by schoolwide average test score levels), and (2) “lack of progress” (defined by changes in schoolwide average test score levels). As a result, this study used a regression discontinuity design where these two rating scores defined eligibility to receive a *SIG*.

The study author estimated the effect of *SIG* eligibility at each threshold, comparing *SIG*-eligible schools with schools that were ineligible. These estimates controlled statistically for differences in positions of the *SIG*-eligible and ineligible schools on either the “lowest achieving” or “lack of progress” continuum.

To measure the impact of *SIG* eligibility on school-wide achievement levels, the study author analyzed the state of California’s Academic Performance Index (API) scores from the 2010–11 school year as the main outcome of interest.²

Features of *School Improvement Grants (SIGs)* in California

Federally funded *SIGs* are offered on a competitive basis to “persistently lowest achieving” schools. Each school in California that received a *SIG* was given nearly \$1.5 million (approximately \$1,500 per student) and was required to adopt a whole-school reform strategy of transformation, turnaround, or restart. Schools were determined to be eligible to receive a *SIG* by having student achievement levels in the lowest 5% in the state (i.e., “lowest achieving”), and low rates of growth on school API scores (i.e., “lack of progress”).

What research questions does this study answer?

The study estimates the impact of *SIGs* on school-wide achievement. Only 82 of the 168 *SIG*-eligible schools ultimately received a *SIG*. As a result, the study estimated both the average impact of being *eligible to receive a SIG* on school API, as well as the average impact of *actually receiving a SIG* on school API. This WWC review focuses on the estimates of the impact of *SIG* eligibility, regardless of whether or not a school received a grant.

The study presented several analyses that examined the effect of being *eligible to receive a SIG* on school API. This WWC review focuses on one set of analyses that provided the most compelling evidence of the effect of *SIG*-eligibility at each of the two thresholds (“lowest achieving” and “lack of progress”).³

To estimate the effect of being eligible to receive a *SIG* for schools at the “lowest achieving” threshold, the authors categorized all schools below the lowest achieving threshold as intervention schools ($n = 542$), and all schools above the threshold as comparison schools ($n = 2,350$). Notably, in this comparison, only a subset of the schools that were categorized as intervention schools also met the “lack of progress” criterion. Furthermore, only a subset of the schools that met both criteria ultimately received a *SIG*. As a result, this analysis provides an estimate of the effect of being eligible to receive a *SIG* at the “lowest achieving” threshold; however, only a small proportion (82 out of 542 schools, or 15%) of the schools categorized as “intervention schools” actually received a *SIG*.

Similarly, to estimate the effect of being eligible to receive a *SIG* for schools at the “lack of progress” threshold, the author categorized all schools below the lack of progress threshold as intervention schools ($n = 1,155$), and all schools above the threshold as comparison schools ($n = 1,737$). As with the analysis estimating the effect at the “lowest achieving” threshold, only a small proportion of the schools categorized as intervention schools for the analysis of the effect of *SIG* eligibility at the “lack of progress” cutpoint (82 out of 1,155 schools, or 7%), ultimately received a *SIG*.

As a result, this WWC report provides two sets of estimates for the impact of *SIGs* on school API: one at the “lowest achieving” threshold and one at the “lack of progress” threshold.

What did the study find?

The study found, and this WWC review of evidence confirmed, that there were positive and statistically significant impacts of *SIGs* on schools at the “lowest achieving” threshold, but not for schools at the “lack of progress” threshold, on levels of schoolwide API in the year that the *SIGs* were received. These results were consistent across several specifications and sensitivity tests.

WWC Rating

The research described in this report meets WWC regression discontinuity design standards with reservations

Strengths: This study is a regression discontinuity design.

Cautions: Both sets of analyses did not pass the baseline equivalence standard. That is, the schools on either side of the “lowest achieving” threshold or the “lack of progress” threshold were not shown to be equivalent on all variables related to school level student achievement as measured by the Achievement Performance Index (API).

The changes in observed API for all reported results may be in part due to (1) improved student learning at *SIG*-eligible schools, (2) high-achieving students moving into the *SIG*-eligible schools and low-achieving students moving out of *SIG*-eligible schools, or (3) a combination of these effects. The analysis cannot separate effects from these various sources.

Additionally, since the study analyzed school-level effects, the magnitude of the effect size reported cannot be directly compared to the magnitude of an effect size of an analysis that uses student-level data.

Finally, as a result of the design used for the study, the reported impacts are only valid at the thresholds that define the eligibility criteria, and do not generalize to all *SIG*-eligible schools.

Appendix A: Study details

Dee, T. (2012). *School turnarounds: Evidence from the 2009 stimulus* (NBER Working Paper No. 17990). Cambridge, MA: National Bureau of Economic Research. Retrieved from <http://www.nber.org>

Setting The study took place in California.

Study sample The sample includes 2,892 public schools in California that met the broad federal criteria for Tier 1 or Tier 2 eligibility for receiving a 2010–11 *School Improvement Grant*. Tier 1 *SIG*-eligible schools include Title I schools that repeatedly did not meet adequate yearly progress under No Child Left Behind and Tier 2 *SIG*-eligible schools include schools that were eligible for Title I support, but were not receiving it.

In the 2010–11 school year, the 2,892 schools in the study sample averaged 824 students per school (standard deviation = 624). The majority of students (66%) were Hispanic, and were receiving free or reduced-price lunch (76%). Approximately 33% of the students were English language learners, and 11% of the students had disabilities.

This study uses school-level data for all analyses.

Intervention group There are two different variables used to determine *SIG* eligibility for the California schools in the study sample: the “lowest achieving” threshold and the “lack of progress” threshold:

- **Lowest achieving:** Baseline (defined as the period between 2007–09) achievement in English/language arts and mathematics was in the bottom 5% of all schools in California. For the purpose of this WWC report, the 542 schools (18.7% of all potentially eligible schools) below the “lowest achieving” threshold are described as Sample A (for achieving).
- **Lack of progress:** Baseline (defined as the period between 2004–08) school Academic Performance Index growth was less than 50 cumulative points across the 5 years prior to *SIG* eligibility determination. For the purpose of this WWC report, the 1,155 schools (39.9% of all potentially eligible schools) below the “lack of progress” threshold are described as Sample P (for progress).

Schools below the cutoff value on *both* thresholds (i.e., schools that were in both Sample A and Sample P) were labeled as “persistently lowest achieving,” and therefore eligible to receive *SIGs*. Only a small subset (168, or 5.8%) of California schools in this study sample met this requirement, and only 82 (2.8%) of these schools ultimately received a *SIG* in 2010–11.

Schools that received *SIG* funding were required to implement one of the three *SIG*-funded whole-school reform models:

- **Transformation:** This model emphasizes (1) teacher and principal effectiveness, (2) comprehensive instructional reform, (3) extended learning time and community engagement, (4) operational flexibility and support, (5) the use of social-emotional and community oriented services and supports (e.g., health and nutrition), (6) data-driven and differentiated instructional strategies, and (7) extending the school day/year. Schools choosing this model were required to replace the principal and introduce teacher evaluations that are based in part on student performance and are used in personnel decisions (e.g., rewards, promotion, retention, and firing).

- **Turnaround:** The components of this model are largely the same school reforms that define the transformation model. However, in addition to replacing the principal, this model also requires replacing at least 50% of the school’s staff.
- **Restart:** This model requires that the school must reopen under the management of a charter-school operator, a charter-management organization, or an educational management organization.

Comparison group

Schools in the study sample that were not identified as “persistently lowest achieving” did not receive *SIG* funding or resources to implement a whole-school reform model, and as such, were considered a “business-as-usual” comparison condition.

Outcomes and measurement

The API is an index of school-level performance based on statewide student testing. For a more detailed description of this outcome measure, see Appendix B.

Support for implementation

No information about support for implementation was provided in the study report.

Reason for review

This study was reviewed by the WWC in response to a request by the Institute of Education Sciences.

Appendix B: Outcome measure for the schoolwide academic performance domain

Schoolwide academic performance

<i>Academic Performance Index (API)</i>	This index is based on statewide student testing. A school's annual API can take on a value ranging from 200 to 1,000 and is calculated by converting schoolwide student performance levels (i.e., advanced, proficient, basic, below basic, and far below basic) on California Standards Tests in English/language arts, mathematics, social studies, and science for students in grades 2–11 into values on the API scale.
---	--

Table Notes: The author also presented impacts on mediators of schoolwide achievement, including rates of schoolwide suspension, truancy, the likelihood of hiring a new principal, average years of teacher experience, the prevalence of teachers with graduate degrees, the prevalence of teachers of particular races/ethnicities (Black, Hispanic, and Asian), and the student–teacher ratio. These outcomes are not included in this WWC report because the study provided insufficient information on these outcomes to meet WWC standards.

Appendix C: Study findings for the schoolwide academic performance domain

Domain and outcome measure	Study sample	Sample size	WWC calculations			
			Estimated impact (standard error)	Effect size	Improvement index	p-value
Schoolwide academic performance						
<i>Academic Performance Index (API)^a</i>	All schools in California. The intervention schools are those in Sample A (lowest achieving).	542 schools meet the lowest achieving criteria and serve as the intervention condition, and 2,350 schools serve as the comparison condition	0.07 (0.02)	0.07	+3	< 0.01
<i>Academic Performance Index (API)^b</i>	All schools in California. The intervention schools are those in Sample P (lack of progress).	1,155 schools meet the lack of progress criteria and serve as the intervention condition, and 1,737 schools serve as the comparison condition	0.00 (0.02)	0.00	+0	0.86

Table Notes: Positive results for effect size and improvement index favor the intervention group; negative results favor the comparison group. The effect size is a standardized measure of the effect of an intervention on school outcomes, representing the change (measured in standard deviations) in a school’s outcome that can be expected if the school is eligible for the intervention (provided that the school is at the threshold in question). The improvement index is an alternate presentation of the effect size, reflecting the change in a school’s percentile rank that can be expected if the school is eligible for the intervention (provided that the school is at the threshold in question).

Study Notes: The analysis was conducted on a standardized version of API scores, and thus, effects are presented in terms of school standard deviation units. The author indicated that the impacts shown above were considered to be the most appropriate tests of the impact of SIG eligibility at each threshold. The impacts in this table were estimated using baseline API as the forcing variable, with linear splines, and student, teacher, and school controls. A correction for multiple comparisons was needed but did not affect significance levels. The p-values presented here were reported by the author in email correspondence. The WWC regression discontinuity standards do not currently provide an approach for aggregating impacts across two thresholds, so no domain average is included in this table. The study is characterized as having statistically significant positive effects because univariate statistical tests are reported for each outcome measure, the effect for at least one measure within the domain is positive and statistically significant, and no effects are negative and statistically significant, accounting for multiple comparisons.

^a Using all schools in the state that met the broad federal criteria for Tier 1 or Tier 2 eligibility, this row examines the impact of SIG eligibility for schools surrounding the “lowest achieving” threshold. In this table, this is estimated by comparing schools in Sample A against all schools not in Sample A.

^b Using all schools in the state that met the broad federal criteria for Tier 1 or Tier 2 eligibility, this row examines the impact of SIG eligibility for schools surrounding the “lack of progress” threshold. In this table, this is estimated by comparing schools in Sample P against all schools not in Sample P.

Endnotes

¹ Single study reviews examine evidence published in a study (supplemented, if necessary, by information obtained directly from the author[s]) to assess whether the study design meets WWC evidence standards. The review reports the WWC's assessment of whether the study meets WWC evidence standards and summarizes the study findings following WWC conventions for reporting evidence on effectiveness. This study was reviewed using the single study review protocol, version 2.0. The WWC rating applies only to the results that were eligible under this review protocol and meet WWC regression discontinuity standards without reservations or meet WWC regression discontinuity standards with reservations, and not necessarily to all results presented in the study.

² There were several additional outcomes included in the study that are not described in this WWC report. See the table notes in Appendix B for more information.

³ In addition to the analysis at each eligibility threshold that used the complete analysis sample of 2,852 schools, the study also conducted impact analyses at one threshold, conditional on being eligible under the threshold. In these sensitivity analyses, the sample was first limited to only those schools that fell below one eligibility threshold. Schools were then divided into two groups using the *other* threshold: all schools below the second threshold were categorized as intervention schools (i.e., *S/G*-eligible), and the remaining schools above the second threshold were considered comparison schools. The author also conducted an impact analysis at the “lowest achieving” threshold that focused on the 1,671 schools in the “optimal bandwidth” around the threshold. These sensitivity analyses are not included in this WWC report.

Recommended Citation

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2013, November). *WWC review of the report: School turnarounds: Evidence from the 2009 stimulus*. Retrieved from <http://whatworks.ed.gov>

Glossary of Terms

Attrition	Attrition occurs when an outcome variable is not available for all participants initially assigned to the intervention and comparison groups. The WWC considers the total attrition rate and the difference in attrition rates across groups within a study.
Clustering adjustment	If intervention assignment is made at a cluster level and the analysis is conducted at the student level, the WWC will adjust the statistical significance to account for this mismatch, if necessary.
Confounding factor	A confounding factor is a component of a study that is completely aligned with one of the study conditions, making it impossible to separate how much of the observed effect was due to the intervention and how much was due to the factor.
Design	The design of a study is the method by which intervention and comparison groups were assigned.
Domain	A domain is a group of closely related outcomes.
Effect size	The effect size is a measure of the magnitude of an effect. The WWC uses a standardized measure to facilitate comparisons across studies and outcomes.
Eligibility	A study is eligible for review if it falls within the scope of the review protocol and uses either an experimental or matched comparison group design.
Equivalence	A demonstration that the analysis sample groups are similar on observed characteristics defined in the review area protocol.
Improvement index	Along a percentile distribution of students, the improvement index represents the gain or loss of the average student due to the intervention. As the average student starts at the 50th percentile, the measure ranges from -50 to +50.
Multiple comparison adjustment	When a study includes multiple outcomes or comparison groups, the WWC will adjust the statistical significance to account for the multiple comparisons, if necessary.
Quasi-experimental design (QED)	A quasi-experimental design (QED) is a research design in which subjects are assigned to intervention and comparison groups through a process that is not random.
Randomized controlled trial (RCT)	A randomized controlled trial (RCT) is an experiment in which investigators randomly assign eligible participants into intervention and comparison groups.
Single-case design (SCD)	A research approach in which an outcome variable is measured repeatedly within and across different conditions that are defined by the presence or absence of an intervention.
Standard deviation	The standard deviation of a measure shows how much variation exists across observations in the sample. A low standard deviation indicates that the observations in the sample tend to be very close to the mean; a high standard deviation indicates that the observations in the sample are spread out over a large range of values.
Statistical significance	Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups. The WWC labels a finding statistically significant if the likelihood that the difference is due to chance is less than 5% ($p < 0.05$).
Substantively important	A substantively important finding is one that has an effect size of 0.25 or greater, regardless of statistical significance.

Please see the [WWC Procedures and Standards Handbook \(version 2.1\)](#) for additional details.