

Conversion Magnet Schools Evaluation

*OMB Clearance Request:
Part B of the Supporting Statement for
Paperwork Reduction Act Submission*

January 2010

PREPARED FOR:
Institute of Education Sciences
United States Department of Education
Contract No. ED-04-CO-0025/0009

PREPARED BY:
American Institutes for Research[®]

Table of Contents

	Page
List of Exhibits.....	ii
Supporting Statement for Paperwork Reduction Act Submission.....	1
B. Collections of Information Employing Statistical Methods.....	1
1. Respondent Universe	1
2. Procedures for the Collection of Information/Limitations of the Study.....	3
2.1 Statistical Methodology for Stratification and Sample Selection.....	3
2.2 Estimation procedure	3
2.3 Degree of Accuracy Needed for the Purpose Described in the Justification.....	4
2.4 Unusual Problems Requiring Specialized Sampling Procedures.....	6
2.5 Use of Periodic (Less Frequent Than Annual) Data Collection Cycles.....	6
3. Procedures to Maximize Response Rates and Deal with Issues of Nonresponse.....	7
4. Tests of Procedures or Methods	7
5. Names and Telephone Numbers of Individuals Consulted.....	8

List of Exhibits

Exhibit 1. Minimum Detectable Effect Sizes for an Interrupted Time Series Study*	5
Exhibit 2. Statistical and Methodological Experts Who Consulted on the Study Design	8

Supporting Statement for Paperwork Reduction Act Submission

B. Collections of Information Employing Statistical Methods

1. Respondent Universe

Universe of MSAP Grantee Elementary Conversion Magnet Schools for the Study

The Conversion Magnet School Evaluation is conducting a comparative interrupted time series (CITS) study to determine if converting to a Magnet Schools Assistance Program (MSAP) supported magnet school is associated with improved achievement and a reduction in minority group isolation. Thus, the study focuses on MSAP-funded magnet schools rather than all magnet schools in the United States. The potential respondent universe consists of all elementary schools in the U. S. that

- existed prior to their district's MSAP 2004 or 2007 grant;
- used MSAP grant funds to establish new, whole-school magnet programs.

The precise size of this universe is unknown because the screening process that identified schools as eligible for the study eliminated some grantee districts from consideration before information about the program structure of their MSAP-funded magnets was collected. However, the upper bound on the size of the respondent universe is the number of new elementary magnet schools of all types funded by the MSAP in the 2004 and 2007 grant cycles. There were 138 such elementary schools in 45 districts and district consortia located in 22 states.¹

In addition to focusing on new conversion magnets established with federal funds, the comparative interrupted time series (CITS) study design imposes further requirements for inclusion in the study:

- Each included school must have a clearly defined attendance area (also often referred to as attendance zone) whose residents are permitted to enroll in the school based on their resident status.
- Each included magnet school must be accompanied by one or more non-magnet comparison schools from the same district.
- The magnet and comparison schools must have existed and administered standardized tests to their students for at least 2 years prior to and 2 years after the magnet conversion date.
- The same test should be used throughout the baseline (pre-grant) and follow-up periods.
- The grantee districts in which the schools are located must be able and willing to provide longitudinal individual student records data (including demographic variables, residence information, and test scores) to the study.

Each of these additional requirements reduced the number of grantee schools eligible for inclusion in the study. The screening process yielded a sample of 25 MSAP-funded conversion elementary magnet schools in 14 grantee districts in 9 states eligible for and able to participate in the CITS study.²

¹ While most funded elementary schools operate whole school programs, a few operate programs-within-schools (PWSs). Based on the characteristics of the funded elementary schools in districts that were not eliminated in the early stages of screening, it is likely that nearly all of the elementary schools funded by the MSAP introduced whole-school magnet programs.

²

These 14 districts will also comprise the potential sample for a fixed-effects analysis of the academic achievement of *non-resident* students in these schools if there are a sufficient number of students switching between other public schools and the 25 MSAP magnet schools³.

Generalizability of the Sample of MSAP Grantee Elementary Conversion Magnet Schools

A consequence of the requirements for inclusion of the magnet schools in our sample is that it is not representative of all elementary magnet schools,⁴ or even representative of all elementary conversion magnet schools in the United States. The fact that they were part of a winning application for an MSAP grant suggests that as a group, these schools may represent a “best case scenario” for new conversion magnet schools.

Although the design requirements of the study necessarily limit the generalizability of the findings, the study attempted to maximize the variability of the contexts in which the schools were found (districts that vary in terms of geographic region, size, urbanicity, and ethnic composition) by recruiting as many eligible schools as possible from the respondent universe. In addition, the Common Core Data (CCD) and NCES surveys will be used to compare conversion magnet schools and non-magnet schools in the study to schools nationwide. The comparison will provide a context for understanding how the study schools differ from other magnet and non-magnet schools.

Despite the limitations on generalizability, carrying out the proposed study is an improvement over previous studies and will make an important contribution to the literature:

- The study has been designed to maximize internal validity and thus the potential of drawing valid conclusions about the relationship of the magnet schools included in the study to student achievement and minority groups isolation. A strength of the study is that it targets the most common type of MSAP magnet school that is characterized by the same grade level (i.e., all study schools are elementary schools), magnet program structure (i.e., all operate whole school programs), and stage of development (i.e., all convert to a magnet school at the start of the grant). This naturally reduces a number of extraneous factors that have often not been controlled in studies of magnet schools.
- The study focuses on a type of magnet school in which a large proportion of students are typically poor and minority—a group of primary concern to the education community and to ESEA in particular.

³ The fixed effects analyses focus on non-resident student who were admitted to one of the 25 conversion magnet schools after the school began its grant and have at least two consecutive years of test scores on the state’s standardized achievement test while attending a non-magnet school in the district and two consecutive years of test scores while attending the magnet school. States begin standardized testing in grade 2 or, more typically, grade 3, and many students enroll in magnet schools at grades K or 1. Consequently, the number of students eligible for the non-resident student study will be smaller than the total non-resident enrollment in the conversion magnet schools. Some of the 25 MSAP schools may be excluded from the analysis if they do not have non-resident students transitioning in grades that allow for two consecutive test scores in both the non-magnet and the magnet school. We will not know which schools can be included in the study until all of the record data have been obtained and reviewed. However, because we are already collecting student record data for all elementary students in the study districts, the analysis of non-resident students would not require the collection of any additional data.

2. Procedures for the Collection of Information/Limitations of the Study

2.1 Statistical Methodology for Stratification and Sample Selection

For the reasons outlined in the previous section, the Conversion Magnet Schools Evaluation has not used stratification or random sampling to select samples of magnet schools for the CITS analyses. Rather, analyses will be based on data from a set of conversion magnet schools drawn from districts that have the capacity to provide the student records data that meet study requirements. Although this approach will limit the generalizability of the results, the studies are designed to maximize internal validity and provide contextual information to help readers interpret study findings.

During the feasibility study, information about potential study schools and pools of participants in true lotteries was gathered from pre-existing sources (grantee applications and performance reports, CCD files, and the National Longitudinal School-Level State Assessment Score Database) and through semi-structured interviews with officials in grantee districts to identify districts and schools that could participate in the study. At the conclusion of the feasibility investigation, all eligible and willing districts were recruited for the evaluation.

Multivariate statistical methods are being used to identify the comparison schools used in the study of resident students. As indicated above, each conversion magnet school must be accompanied by at least one non-magnet comparison school matched as well as possible to the magnet school on key characteristics at baseline (the year before the MSAP grant was awarded). The pool of potential comparison schools includes elementary schools that

- are in the same districts as the conversion magnet school;
- have not operated a magnet program during the school years covered by the study; and
- have clearly defined attendance areas.

2.2 Estimation procedure

Procedures that will be used by the different components of the study to estimate the relationships between conversion magnet schools and student outcomes (academic achievement and minority group isolation) are discussed at length under item 16 in Part A. Statistical tests will be run to determine how similar the study schools are to other schools in their district and to the total set of conversion magnet schools funded in 2004 and 2007.

2.3. Degree of Accuracy Needed for the Purpose Described in the Justification

The desired degree of accuracy for the core analyses of magnet versus non-magnet differences is a minimum detectable effect size (MDES) of 0.20 standard deviations, which is an upper bound for a “small” effect in the evaluation literature. In addition to a basic analysis of magnet versus non-magnet effects, however, it is desirable to assess the academic success of subgroups of students in magnet schools versus non-magnet schools at a similar level of accuracy. The literature on the determinants of student achievement provides many examples in which the overall relation between a given school characteristic and achievement was quite weak, but a far stronger relation emerged between the same characteristic and achievement of one or more subgroup of students. As an example from the area of school choice, Howell and Peterson’s study of the impact of the New York voucher program found that winning a voucher lottery was associated with zero gains in achievement in many cases, but positive gains in achievement for certain races and grades, with other races and grade levels exhibiting no effect. A student fixed-effect analysis of achievement gains of students in San Diego by Betts, Zau and Rice showed that variations in class size had twice as large an effect on English language learners compared with other students. Krueger and Whitmore reported from their re-analysis of the Tennessee STAR experiment that low-income students responded far more strongly to reductions in class size than did other students.⁵

Interrupted Time Series Study of Resident Students in Conversion Magnet Schools

The interrupted time series analysis derives its inferential power from a comparison of the average achievement of a sequence of cohorts over time within each sample school. The longer and more stable this time series before the conversion to a magnet school, the more powerful will be the analysis of the estimated effects.

The power of the estimates of effects from an interrupted time series analysis also depends on whether the years of student achievement data for the post-conversion period are aggregated into a single outcome or each year’s scores are treated as distinct outcomes. The most powerful estimates can be obtained when a stable and well-defined trend emerges post-conversion. This allows a comparison of the slope or intercept shift defining the trend and the pre-conversion trend. In reality, however, a post-conversion trend is not always stable. The MDRC/AIR design report of 2004 by Bloom et al.,⁶ estimated the difference between achievement in any single follow-up year and the mean achievement during the baseline period. While this measure is inherently less stable (because it is affected by year-to-year variation in tests, policy contexts, demographics, and so forth), it requires fewer assumptions about the shape of the post-conversion trend in achievement, and therefore has been adopted in the power analysis presented here.

Exhibit 1 shows how the statistical power of the interrupted time series design varies for the full sample and different subgroups. It shows that the interrupted time series design has sufficient statistical power to detect significant effects at a minimum detectable effect size (MDES) of 0.17 for the full sample and 0.21 for a subgroup of 20 percent of students in the sample. To estimate the number of magnet schools needed to yield an MDES of 0.2 or less for the resident student population and various sub-samples, we assumed that the desired sample would resemble, in

⁵ Howell, William G., & Paul E. Peterson (2002) *The education gap: Vouchers and urban schools*. Washington, D.C.: Brookings Institution; Betts, Julian R., Zau, Andrew, & Rice, Lorien (2003). *Determinants of student achievement: New evidence from San Diego*. San Francisco: Public Policy Institute of California; Krueger, Alan & Whitmore, Diane (2000). “The effect of attending a small class in the early grades on college test-taking and middle school test results: Evidence from project STAR.” National Bureau of Economic Research Working Paper 7656.

⁶ Bloom, H., Doolittle, F., Garet, M., Christenson, B., & Eaton, M. (2004). *Designing a study of the impact of magnet schools on student achievement: Alternative designs and tradeoffs*. Submitted to U.S. Department of Education, Institute of Education Sciences under task order #ED-01-CO-0060/0002 by MDRC and American Institutes for Research.

number of students tested, average number of years of baseline data, and average number of comparison schools, the average characteristics of the sample of magnet schools we had identified for the study. We based our estimate of the average number of students tested in reading and in mathematics in the most recent years available, for the magnet schools and a provisional sample of comparison schools.

Exhibit 1. Minimum Detectable Effect Sizes for an Interrupted Time Series Study*

Subgroup size as percent of full resident student sample	<i>Reading</i> MDES	<i>Mathematics</i> MDES
20	0.206	0.193
30	0.188	0.176
40	0.178	0.166
50	0.172	0.160
Full Resident Sample	0.160	0.148

*Based on characteristics of 23 magnet schools with usable reading scores and 25 magnet schools with usable mathematics scores in the study sample

A number of important assumptions underlie these calculations. These are discussed in detail in the design report by Bloom et al.⁷ as well as in the evaluation's feasibility memo and NCEE's responses to OMB questions in March 2008 (Appendix Q). The first is the natural cohort-to-cohort variation in the outcome variables. Such natural variance significantly reduces the statistical power of an interrupted time series design, because it seeks to identify an interruption in an otherwise stable trend of outcomes. In calculations of statistical power, this cohort-to-cohort variation is expressed in terms of an intra-class correlation coefficient ρ . This coefficient denotes the proportion of the total test score variance among students that is between cohorts within schools. Because there is little published data to guide a choice of ρ , we estimated this coefficient based on all known empirical estimates of ρ for elementary reading (for which we estimated $\rho=0.022$) and elementary mathematics (for which we estimated $\rho=0.02$).⁸

Other key assumptions also follow the design report and the analyses conducted in preparation for it. We assumed that (1) 80 percent of the students at each school would be resident, (2) on average 2.5 years of baseline test score data were available before the year of magnet conversion, and (3) on average there were two comparison schools for each magnet. Lastly, we included a term to account for the variation in true impacts across schools.⁹

⁷ Ibid.

⁸ For mathematics, we used the median estimate for ρ obtained by Bloom (1999) in his study of grade 2 and grade mathematics scores in Rochester, New York. For reading, we took a simple average of Bloom's median estimates for grades 2 and 6 in Rochester and estimates for grade 2 from six other districts around the country provided by Michael Garet of AIR (with permission of ED). These estimated values of ρ are lower than the estimated value of 0.055 (which was based on fewer empirical estimates) that was used by Bloom, et al. in the design report to produce the estimate that an initial estimate that a sample of 50 magnet and 100 comparison schools were required to obtain an MDES of 0.2.

⁹ The formulation of MDES for the interrupted time series is:

$$\text{MDES} = \text{Multiplier} (\alpha, 1-\beta) * \text{sqrt} (\text{variance of the impact estimate})$$

The Multiplier ($\alpha, 1-\beta$) term indicates how many standard errors from zero the impact must be to be sure of detecting it ($p < \alpha$, power = $1-\beta$). We have assumed $\alpha = .05$ for a two-tailed test, $\beta = .20$, so power = .80). The multiplier is approximated as the sum of two t 's $t(0.05, xdf)$, and $t(.40, xdf)$, where each is the tail probability for a t with x df. We've approximated the df as the n of schools. Strictly speaking, the second term should be a non-central t , but t is a good approximation.

The variance of the impact estimate is the sum of two terms:

Exhibit 1 shows that the interrupted time series design has sufficient power to detect an effect of 0.21 standard deviations for a subgroup representing 20 percent of all students in the magnet schools in the sample of schools identified for the study (23 magnets and 46 comparison schools for reading and 25 magnet schools and 50 comparison schools for mathematics).

Fixed-Effects Study of Non-resident Students Attending Conversion Magnet Schools

The original request for OMB clearance included a detailed discussion of the power of estimates of effects from a lottery based experimental study design. Although an insufficient number of students are eligible for a randomized lottery-based study, it may be possible to conduct a quasi-experimental fixed effects analysis of students who switch between traditional public schools and the conversion magnet schools in our study. Such an approach could include these magnet schools regardless of whether or not their lotteries are oversubscribed. Until we know the number of students who actually switched, however, we cannot provide estimates of the power (and MDES) for the student fixed-effect models that would be used in this analysis. However, because we are already collecting student record data for all elementary students in the study districts, the analysis of non-resident students would not require the collection of any additional data.

2.4. Unusual Problems Requiring Specialized Sampling Procedures

Aside from the methods described above for identifying comparison schools for the conversion magnets, no specialized sampling procedures will be needed.

2.5. Use of Periodic (Less Frequent Than Annual) Data Collection Cycles

Most of the data used by this study will be collected less frequently than annually.

The **grantee screening process** occurred only once, during the feasibility phase. Officials in districts that received 2004 MSAP grants were interviewed during the summer and fall of 2007, and officials in districts that received 2007 grants were interviewed during the fall and winter of 2007. (In districts that received a 2007 grant as well as a 2004 grant, officials were contacted during both screening waves, but during the second wave were only asked questions pertaining to the newly funded elementary conversion magnet schools.)

Although districts collect **student records data** on an annual basis, the study has requested several years of these data at once during the first wave of data collection (i.e., 2001-2002 through 2007-2008 for districts in the 2004 MSAP cohort, and 2003-2004 through 2007-2008 for the 2007 cohort). Data for the 2008-2009 and 2009-2010 school years will be requested in annual installments, but the study will accommodate districts' requests to deliver these data in fewer installments.

The **principal survey** will be administered only twice during the 46-month evaluation study—the 2004 cohort was surveyed in 2008, and both the 2004 and 2007 cohorts will be surveyed in 2010.

$$[(\text{variance of impact for one magnet}/n) + (\text{variance of true impacts across magnet school}/n)]$$

The second part of this is the random effect assumption; we're assuming the numerator there is 0.01. The first part is the variance in the impact for a single magnet school and its two comparisons. It is derived from the average n of students per school in the sample, average number of baseline periods (2.5), and variation across cohorts or ρ , which we are estimating as 0.022 for reading and 0.02 for mathematics.

3. Procedures to Maximize Response Rates and Deal with Issues of Nonresponse

The expected response rate for *student records data requests* from the districts is 100 percent. This is because districts that are unable or unwilling to supply data to the study were screened out during the feasibility phase. In the evaluation phase, the project staff are working with data managers in the participating districts to ensure that requests are well-understood and aligned with data elements maintained in each district, and that data delivery schedules are realistic.

The expected response rate for the *principal surveys* is 85 percent. Several procedures are being used to ensure high response rates:

- Obtaining high response rates depends in part on the quality of the survey instruments. The principal survey has been pre-tested to ensure that the questions are clear and as user-friendly as possible (in particular, many of the items are answered by checking off boxes rather than writing in responses), skip patterns are easy to follow, and the survey can be completed quickly. It has also been kept short by excluding requests for information that can be obtained from other data sources.
- Respondents receive a small amount of compensation for their time in completing the principal survey. This encourages some respondents to participate and thus increase the response rate.
- AIR and BPA project staff are responsible for maintaining contact with respondents in an effort to track returns and follow up with non-respondents. Two weeks after the surveys have been sent, a reminder letter and a second copy of the survey will be sent to principals who have not yet responded. Four weeks after the surveys have been sent, phone calls will be made to principals who have not yet responded, and a staff member will offer to take responses over the phone. The study will also enlist the support of the local MSAP project directors to encourage principals to return the surveys.

4. Tests of Procedures or Methods

The first phase of the Conversion Magnet Schools Evaluation tested the feasibility of the methods proposed for the evaluation studies by assessing the availability of schools, students, and district data systems meeting the requirements of these methods.

The principal survey was tested by a small sample of respondents (fewer than 10) for two purposes—to ensure that the instrument and procedures work effectively, and to verify estimates of respondent burden. It should be noted that many items in the survey are based on items from operational NCES surveys (SASS and ECLS-K), and thus have already been piloted and administered to national samples of principals.

The protocol for the close-out interview with local MSAP project directors has not been piloted. However, many of the items in this protocol are based on items from the district screening protocols used during the feasibility study. Project staff took note of difficulties encountered during the screening interviews and has made adjustments in order to clarify questions and make them less burdensome. This data collection is being conducted as a semi-structured interview. Project staff who conduct the interviews will take care to limit the duration of these conversations to a half hour.

5. Names and Telephone Numbers of Individuals Consulted

The conceptual framework for the study of the relationship between magnet school programs and student achievement was developed by an IES-commissioned work group that produced a report on

potential research designs in 2004. The principal statistical and methodological experts who participated in that design study were Drs. Howard Bloom and Fred Doolittle from MDRC and Michael Garet from AIR. The statistical and methodological experts who were subsequently consulted for the design for the Conversion Magnet Schools Evaluation were Drs. Julian Betts (University of California at San Diego), Johannes Bos (Berkeley Policy Associates) and Michael Garet. Contact information is shown in Exhibit 3.

Exhibit 2. Statistical and Methodological Experts Who Consulted on the Study Design

Name	Position	Telephone	Email
Julian Betts	Professor of Economics, University of California, San Diego	(858) 534-3369	JBetts@ucsd.edu
Howard Bloom*	Chief Social Scientist, MDRC	(212) 532-3200	Howard.Bloom@mdrc.org
Johannes Bos	Formerly CEO and Principal Research Scientist, BPA; now Vice President, AIR	(510) 465-7884 now (650)-843-8110	Hans@bpacal.com now jbos@air.org
Fred Doolittle*	Vice President and Director of Policy Research and Evaluation, MDRC	(212) 532-3200	Fred.Doolittle@mdrc.org
Michael Garet	Vice President and Managing Research Scientist, AIR	(202) 403-5345	MGaret@air.org

*Drs. Bloom and Doolittle had lead roles in developing the design report from 2004 commissioned by IES. Although that report provides a conceptual framework and informs estimations for this proposed study, neither of these experts have been directly consulted on this proposed study.

The data will be collected and analyzed by staff of the American Institutes for Research (AIR) and Berkeley Policy Associates (BPA) under contract to IES. Key staff on the project include Dr. Julian Betts (Principal Investigator) of UC San Diego; Drs. Bruce Christenson (Project Director), Marian Eaton (Deputy Director) of AIR; and Dr. Hans Bos (Senior Advisor) of AIR. IES staff overseeing the study are Marsha Silverberg and Lauren Angelo.