

Conversion Magnet Schools Evaluation

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

April 2007

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
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.....	4
2.3. Degree of Accuracy Needed for the Purpose Described in the Justification.....	4
2.4. Unusual Problems Requiring Specialized Sampling Procedures.....	9
2.5. Use of Periodic (Less Frequent Than Annual) Data Collection Cycles.....	9
3. Procedures to Maximize Response Rates and Deal with Issues of Nonresponse.....	9
4. Tests of Procedures or Methods.....	10
5. Names and Telephone Numbers of Individuals Consulted.....	10

List of Exhibits

	Page
Exhibit 1. Minimum Detectable Effect Sizes for Different Scenarios.....	8
Exhibit 2. Minimum Number of Schools for Different Scenarios (Assumes a MDES of 0.2).....	8
Exhibit 3. Statistical and Methodological Experts Who Consulted on the Study Design.....	10

Supporting Statement for Paperwork Reduction Act Submission

B. Collections of Information Employing Statistical Methods

1. Respondent Universe

The Conversion Magnet Schools Evaluation consists of two components—one focusing on resident students and the other on non-resident students of MSAP-funded elementary schools. The two components have different respondent universes and will be discussed separately below.

Universe of MSAP Grantee Elementary Conversion Magnet Schools for the Study of Resident Students

The potential respondent universe of **MSAP-funded conversion magnet schools** includes all elementary schools in the United States that

- existed prior to their district’s MSAP 2004 or 2007 grant;
- used MSAP grant funds to establish new, whole-school magnet programs; and
- have clearly defined attendance areas (also often referred to as attendance zones) from which resident students are drawn.

The precise size of this universe is not known because the 2007 MSAP grants have not yet been awarded, and currently available information about the schools in the 2004 MSAP cohort lacks some necessary details. Preliminary information about the 2004 cohort identifies 76 grantee elementary schools in 27 districts in 14 states that existed in 2003-2004 and used MSAP funds to establish new magnet programs. It is anticipated that the grantee screening process during the feasibility phase will determine that most of these schools have clearly defined attendance areas and operate whole-school programs, as this is a common configuration for elementary school magnets funded by MSAP. If the mix of grantee schools in the 2007 cohort is similar, the total respondent universe would consist of about 140 schools. The study will attempt to recruit 50 conversion magnet schools from this universe and 100 non-magnet comparison schools from the same set of districts.

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. In addition to focusing on new conversion magnets established with federal funds, the design for the component of the study focusing on resident students’ achievement (an interrupted time-series design) imposes further requirements for inclusion in the study:

- Each included magnet school must be accompanied by one or more non-magnet comparison schools from the same district with similar demographic and achievement profiles.¹
- The magnet and comparison schools must have existed and administered standardized tests to their students for at least 3 years prior to and 3 years after the magnet conversion date. Furthermore, the same test should be used throughout the baseline (pre-grant) and follow-up periods.

¹ The number of non-magnet schools that are reasonably good demographic and achievement matches to the conversion magnets will be determined during the grantee screening process.

- 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 is expected to reduce the number of grantee schools that are eligible for inclusion in the study. A further consequence of these requirements is that the study will not be representative of all elementary magnet schools,² or even representative of all elementary conversion magnet schools in the United States.

Although the design requirements of the study necessarily limit the generalizability of the findings, the study will attempt to maximize the variability of the contexts in which the schools are 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, should there be sufficient schools to make the study feasible, 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 would 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 a common type of magnet school that is characterized by the same grade level, magnet program structure, and stage of development. 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 NCLB in particular.

Universe of Non-resident Students Applying via Lottery to Magnet Schools

The respondent universe for an experimental study of non-resident students' achievement includes all non-resident students who applied for admission to a conversion magnet school through a “true” lottery (one in which there were more applicants than available seats and in which all applicants were equally likely to be selected). Students selected by the lottery are the experimental group, and students not selected are the controls.³ If no pretests are available, the study will try to identify 5,300 lottery applicants. The availability of pretests with substantial predictive power could reduce

² Restricting the sample to elementary schools using 2004 or 2007 MSAP funds to convert to magnet status excludes from consideration elementary magnet schools in the grantee districts that have any of the following characteristics: (1) did not receive funds from MSAP grants awarded in 2004 or 2007; (2) operates a magnet program-within-a-school (PWS) that serves only part of the school's total enrollment; (3) requires all students to apply for admission and does not give preference to applicants based on their residence in an attendance area (e.g., a “dedicated” magnet school); (4) has an unclearly defined attendance area (e.g., is one of multiple schools in an attendance area that students can attend); (5) used MSAP funds to revise an existing magnet program; (6) opened as a brand new school rather than as a pre-existing school that had converted to magnet status. Also excluded from the universe are all elementary school magnets in districts that did not receive MSAP grants.

³ The existence of magnet schools that are so popular that they have more applicants than vacancies makes an experimental study of randomly selected students possible. However, the likely (but unmeasured) differences between these over-subscribed magnets and other magnets that are in less demand implies a limit to the generalizability of the study results as may be typical of studies that rely on voluntary lotteries as a source for randomization.

the number of lottery applicants identified for the study by about half, to approximately 2,600 applicants.⁴

The size of the respondent universe will be established during the feasibility study. Selection of applicants for a given school typically involves several lotteries—at least one per grade—of which not all are “true.” The proportion of seats in magnet schools that are filled by non-resident students as well as the number of students involved in “true” lotteries to fill them is expected to vary substantially by school and grade, and the details can only be discovered through discussions with district staff.

To be eligible for the study, the lottery applicants must have adequate test data and be located in districts that are able and willing to provide longitudinal individual student record data including demographic, residence, and achievement information as well as information on the lottery (i.e., who applied, who was selected/not selected, and where each applicant ultimately enrolled). With regard to achievement data, the minimal requirement for study eligibility would be one year of test scores in the year following their lottery selection or non-selection to the magnet school. However, multiple years of post-lottery test scores and at least one year of pre-lottery test scores are also highly desirable.

The number of lottery applicants eligible for the study is not expected to be much larger than the number needed to detect meaningful effect sizes; and the evaluation will not employ random sampling to select students for the study. If more students are found to be eligible than are needed for the study, priority will be given to balancing geographic diversity and budgetary constraints. The former would argue for multiple locations, while the latter tends to restrict analyses to districts with the largest number of lottery participants.⁵

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 will not use stratification or random sampling to select samples of magnet schools and non-resident students for the envisioned studies. Rather, the studies will be based on data from a set of conversion magnet schools and students 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 will be 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 can participate in the study. If this investigation shows that sufficient data are available, it is likely that all eligible and willing districts will be recruited for each part of the study.

⁴ Sample size estimates for the lottery study are discussed in more detail in section 2.3 of this document.

⁵ If an insufficient numbers of students are eligible for an experimental study of non-resident students, an alternative design will be considered that involves all students who switched schools, regardless of whether they participated in a lottery. Such a design is based on a student fixed effects model in which the achievement gains of individual student would be compared before and after the students switched between the traditional public school and the conversion magnet school. The size of the respondent universe might be established from information being gathered during the screening of districts.

Multivariate statistical methods will be 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.

Non-magnet elementary schools that meet these criteria will be matched to the MSAP-funded conversion magnet schools on the following pre-grant year characteristics: percent of African American students; percent of Hispanic students; percent of Asian-American students; percent of students eligible for free/reduced price meals; academic achievement;⁶ and school size. The statistically “nearest matching” comparison schools for each magnet school will be identified as those that have the smallest Euclidean or Mahalanobis distance to the magnet school. Before finalizing the matches, AIR/BPA will consult district officials to confirm that the candidate comparison schools do not differ from the magnets in unexpected ways (e.g., the presence of special programs or populations in the schools).

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.

Once the schools have been identified, 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. Similarly, baseline characteristics of the experimental and control students included in the lottery analysis will be compared to ensure that they are statistically equivalent.

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

⁶ The academic achievement indicators will be created by averaging mean scores on standardized English Language Arts and mathematics assessments across the grades tested in the year prior to the MSAP grant.

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.⁷

A second powerful reason for a focus on subgroups is that NCLB has brought a sharp focus onto inter-group differences in student achievement. Under this federal law, schools are accountable not only for overall student proficiency, but also for proficiency rates of numerically significant subgroups at each school, where subgroups are defined by characteristics such as family income, race and ethnicity, and English Language Learner status. Given that school choice is one of the key remedies for low-achieving students in low-achieving schools under NCLB, it is crucial that the present study evaluate the relation between school choice in the form of magnet schools and achievement of subgroups.

Being able to conduct such subgroup analyses is thus critical for the utility of this study. It is important to document whether the estimated effect of magnet conversion is concentrated among academically successful students who may have done well without the conversion, but also that it benefits students who enter with average academic records or who may come from educationally or economically disadvantaged backgrounds.

Successfully conducting such subgroup analyses requires a larger sample of schools than would be the case if we only examined effects for the full sample of students in each school. Exhibit 1 presents preliminary estimates of the MDES for a range of scenarios for subgroup representation, and Exhibit 2 provides preliminary information about how required sample sizes vary for these scenarios. Below we discuss in detail the assumptions underlying these MDES calculations.

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 the analysis of the estimated effects will be.

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 could be obtained if a stable and well-defined trend emerged post-conversion. This would allow 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.,⁸ 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.

⁷ 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.

⁸ 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.

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 0.15 for the full sample and 0.19 for a subgroup of 20 percent of students in the sample (assuming it would include a sample of 150 schools, consisting of 50 magnet schools and 100 comparison schools).

A number of important assumptions underlie these calculations, all of which are discussed in detail in the design report by Bloom et al.⁹ 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 variance among students that can be attributed to the students being in different cohorts. As part of a review of the design report for this study, a range of empirical estimates ρ was compiled. To protect ourselves from an underpowered study in case there would be more cohort-to-cohort variation in our sample, we used a value of 0.055 for ρ , which is at the upper end of the distribution of empirical values found during the MDRC/AIR design study. This value corresponds to the conservative estimates presented in Table 1 of that study.

Other key assumptions also follow the design report and the analyses conducted in preparation for it. We assume 100 students per school (for full sample analyses) and twice as many comparison schools as magnet schools. We also assumed we will have three pre-conversion time periods (cohorts) to build a time series. Lastly, we included a term to account for the variation in true impacts across schools.¹⁰

Exhibit 1 shows that the interrupted time series design has sufficient power to detect an effect of 0.2 standard deviations for a subgroup representing 20 percent of all students in the magnet schools with a sample of 50 magnet schools (and 100 comparison schools). Exhibit 2 shows what the minimum number of schools would be for an MDES of 0.2 for the full sample and subgroups of 20, 30, 40, and 50 percent of the sample, respectively.

⁹ Ibid.

¹⁰ 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:

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

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's derived from the assumed n of students per school (e.g., 100), n of baseline periods (3), and variation across cohorts or ρ , which we've assume to be 0.055.

Lottery Study of Non-resident Students Applying to Magnet Schools

The power of the estimates of effects from the lottery-based design is derived from a comparison of individual students' outcomes for those who "win" the lottery and those who do not. This corresponds to a simple individual-level random assignment design, with randomization conducted separately for each magnet school.

The primary constraint on the statistical power of the lottery design is the availability and predictive power of pretest measures. The availability of a pre-test is likely to depend on the grade level at which students apply to enroll in a magnet school. If most applicants apply to enroll at kindergarten or first grade, a pre-test is unlikely to be available. On the other hand, if students apply in later grades, a pre-test will probably be available and is likely to have substantial predictive power. Given the uncertainty about the availability and quality of pre-test data, MDES estimates are provided for the case of no pre-test, as well as for a range of different values of the R^2 for a post-test on pre-test regression. It is possible to approximate these expected R^2 with empirical data collected from existing data sources.¹¹

In addition to assumptions about the explanatory power of the pre-test (which are captured by the R^2 values included in Exhibits 1 and 2), it is necessary to make assumptions about the number of schools which hold lotteries. Because schools that are not oversubscribed will admit all applicants, they will not provide randomly assigned samples of lottery winners and losers. Our calculations suggest that for all but the smallest subsamples, we will require significantly fewer than the 50 magnet schools envisioned for the study of resident students described earlier. Our calculations assume that for each lottery we will allow a separate intercept, which is crucial for removing variations in the types of students across lotteries. Other assumptions are shown in Exhibits 1 and 2, in the panels on the right side.

For the lottery-based design, Exhibit 1 shows how the statistical power (expressed in minimum detectable effect sizes) varies with different assumptions about the explanatory power of the regression effects and the smallest subgroup breakdown in the analysis of estimated effects. Exhibit 2 shows how the required sample size for an MDES of 0.20 varies under different assumptions. Thus, for example, Exhibit 1 shows that, for a sample of 2,600 applicants to oversubscribed magnet schools, the MDES ranges from a high of 0.13 with no pretest to as small as 0.09 with a pretest having an R^2 of 0.5. Note that these results are based on the conservative assumption that 26 of our 50 magnet schools will have lotteries due to being oversubscribed. The MDES rises rapidly as the analysis focuses on smaller subgroups of the full sample. With no pretest, the MDES exceeds 0.20 for analyses of any subgroup made up of fewer than 40 percent of the full sample. An acceptable MDES of 0.20 or lower can be maintained for somewhat smaller subgroups with a pretest. However, for a sample of 2,600 applicants, a pretest with an R^2 of 0.5 or

¹¹ Such analyses to establish empirical values for the predictive power of pre-tests will be pursued during the feasibility phase of the study. The MDRC/AIR design study provides useful guidance. See also Bloom, H.S., Richburg-Hayes, L., & Black, A. R. (2005). *Using covariates to improve precision: Empirical guidance for studies that randomize schools to measure the impacts of educational intervention*. New York: MDRC.

higher is required to achieve a MDES of 0.20 on subgroups as small as 20 percent of the full sample.¹²

Exhibit 2 shows the number of oversubscribed magnet schools that would be needed for an MDES of 0.20 if one assumes an average of 100 applicants per school. When there is no pretest, a minimum of 11 schools, resulting in 1,100 applicants, are needed to conduct an analysis on the full sample. That number rises rapidly for analyses of subgroups, with as many as 53 schools with an average of 100 applicants each, resulting in 5,300 applicants, being needed for analyses of a subgroup that is 20 percent of the full sample. Depending on its predictive power, the availability of a pretest can substantially reduce the number of oversubscribed schools needed for subgroup analyses. With an R² of 0.5, the number of schools needed for analyses of 20 percent of the full sample is half (N=26) that needed when such a pretest is not available.¹³

Exhibit 1. Minimum Detectable Effect Sizes for Different Scenarios

Subgroup size as percent of full sample*	Interrupted time series design (assuming 150 schools)**	No pretest	Lottery design ***			
			R ²			
			0.2	0.3	0.4	0.5
20	0.19	0.28	0.25	0.24	0.22	0.20
30	0.18	0.23	0.21	0.19	0.18	0.16
40	0.17	0.20	0.18	0.17	0.16	0.14
50	0.16	0.18	0.16	0.15	0.14	0.13
Full sample	0.15	0.13	0.11	0.11	0.10	0.09

* Assumes subgroups equally distributed across schools.

** Assumes data collection for 50 magnet schools and 100 comparison schools, 100 students per school and intercohort variability of 0.055.

*** Assumes 26 schools, each with 100 lottery applicants; and that for each school, 75 percent of these applicants were selected for admission while 25 percent were not.

MDES based on 80 percent power, alpha of 0.05 (two-tailed test).

Exhibit 2. Minimum Number of Schools for Different Scenarios (Assumes a MDES of 0.2)

Subgroup size as percent of full sample*	Interrupted time series design (number of schools)	No pretest	Lottery design **			
			R ²			
			0.2	0.3	0.4	0.5
20	150	53	42	37	31	26
30	120	35	28	25	21	17
40	111	26	21	19	16	13
50	105	21	17	15	12	10
Full sample	93	11	8	7	6	5

* Assumes subgroups equally distributed across schools.

** Assumes 100 applicants per school; and that for each school, 75 percent of these applicants were selected for admission while 25 percent were not.

MDES based on 80 percent power, alpha of 0.05 (two-tailed test).

¹² In the event that an insufficient number of students are eligible for a randomized lottery-based study, an alternative is to consider conducting a student fixed effects analysis of students who switch into or out of magnet schools. Such an approach could include magnet schools regardless of whether or not their lotteries are oversubscribed. Estimating the power (and MDES) for student fixed-effect models is more difficult a priori than it is for the lottery-based analysis. Without knowing the number of students who “switch,” we cannot say for sure what the power of the analysis would be. However, we may be able to approximate the statistical properties of this fixed-effect analysis by referring to closely related work using the same methods that has been done on a different type of school choice, that of charter schools. (Betts, J.R., Tang, Y.E., & Zau, A.C. (2007). Madness in the method? A critical analysis of popular methods of estimating the effect of charter schools on student achievement. Manuscript. San Diego: Department of Economics, University of California, San Diego.)

¹³ The results shown are based on the assumption that in each lottery, 75% of the applicants are admitted to the magnet school, and 25% are not.

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** occurs only once, during the feasibility phase. Officials in districts that received 2004 MSAP grants will be interviewed between May 1 and August 1, 2007, and officials in districts that receive 2007 grants will be interviewed between July 2 and September 28, 2007. (If a district receives a 2007 grant as well as a 2004 grant, officials will be contacted during both screening waves, but during the second wave will only be asked questions pertaining to the newly funded elementary conversion magnet schools.)

Although districts collect **student records data** on an annual basis, the study will request several years of these data at once during the first year of the evaluation phase (i.e., 2001-2002 through 2006-2007 for districts in the 2004 MSAP cohort, and 2003-2004 through 2006-2007 for the 2007 cohort). Data for the 2007-2008 through 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—in 2007 for the 2004 cohort and again in 2010 for both the 2004 and 2007 cohorts.

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 will be screened out during the feasibility phase. Project staff will work closely 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 each **principal survey** is at least 85 percent. Several procedures will be 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 will receive a small amount of compensation for their time in completing the principal survey. This should encourage respondents to participate and thus increase the response rate.
- AIR and BPA staff will be 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 tests 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 is being 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 most of the items in the survey are based on items from operational NCEES surveys (SASS and ECLS-K), and thus have already been piloted and administered to national samples of principals.

The semi-structured grantee screening protocols and the district data requests are not being piloted. However, project staff will take note of difficulties encountered during the initial uses of these instruments and will make any adjustments necessary to make questions clearer and less burdensome in later interviews.

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 3. 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	CEO and Principal Research Scientist, BPA	(510) 465-7884	Hans@bpacal.com
Fred Doolittle*	Vice President and Director of Policy Research and Evaluation, MDRC	(212) 532-3200	Fred.Doolittle@mdrc.org
Michael Garet	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 BPA. IES staff overseeing the study Marsha Silverberg and Lauren Angelo.

