Response to Terms of Clearance for Conversion Magnet Schools Evaluation (1850-0832)

QUESTION 1 - NCEE provided three criteria (response of 6/19/07 to OMB questions) that would help it determine whether the proposed study was feasible. It appears that at most one of these was met. Please provide a more detailed discussion of the feasibility results, specifically in light of those criteria.

As shared in the 6/19/07 response to OMB questions, "…the determination of whether or not to implement the evaluation [was] based on the availability of data to support the **interrupted times series (ITS)**." The response specified that the necessary data for this analysis included 50 magnet schools and 100 non-magnet comparison schools and that (1) each magnet school must be accompanied by one or more non-magnet comparison schools from the same district with similar demographic and achievement profiles, (2) the magnet and comparison schools must have existed and administered the same standardized tests to their students for at least 3 years prior to and 3 years after the magnet conversion date, and (3) the districts must be able and willing to provide longitudinal individual student records data. These criteria were established based on prior power calculations that demonstrated this overall sample (50 magnets, 100 comparison schools) would be sufficient to detect an MDE of .19 for a sub-sample of approximately 20%.

However, subsequent to the submission of that response, we and our contractor refined the power calculations for the ITS and tailored these calculations to focus on (1) estimation of effects on the large group of resident students (ED's greatest policy interest), rather than smaller subsamples, and (2) the particular schools that are eligible and willing to participate in the study. The original calculations had been overly conservative in the assumptions (about the R-squared, intra-cohort correlation, etc.) because there were based on a limited set of published data that were not particularly aligned with our study parameters. The new power calculations draw on a wider set of information, including published data for the specific sample of magnet schools recruited; these new calculations indicate that we would need substantially fewer schools, 15-16 magnet schools and 32-34 comparison schools (depending on reading or math outcome), to achieve an MDE of 0.20 for the **resident student** sample, even if not all of the schools have a full 3 years of baseline (pre-grant) achievement data (see appendix).

According to the criteria established earlier:

1) We have identified 23 conversion magnet schools and 48 comparison schools in 13 districts, an average of 2.2 comparison schools per magnet school. That full set of schools will be used *for an analysis of* math achievement gains, while 21 have the data to conduct the analysis of reading achievement gains.

2) Among the identified schools/districts, we have an average of 2.6 years of baseline data and expect to collect the full three years of post-grant data.

3) The 13 districts in the identified sample have agreed to provide the longitudinal data. We have another 2 districts we believe are eligible for the study, and are pursuing their cooperation; if they are included in the study, the MDE will be reduced.

Overall, with our current sample of schools that meet criteria and are willing to participate, we will be able to detect an MDE of .167 for resident students over the three-year period of the MSAP grant (see power analysis results – Appendix B Table 2). Although or primary analysis will focus on the resident students as a whole, we will still be able to detect effects for subgroups of 30% and likely less. This would allow us the opportunity to conduct analyses for specific grade levels and some minority groups.

QUESTION 2 - In addition, please clarify whether the number of identified schools represents those for which participation (via the districts) has been secured, or merely the universe from which NCEE must secure agreement to participate.

All 23 + 48 schools in the 13 districts that we previously identified have been screened, determined to have the necessary data, and are willing to participate. These schools/districts have received MSAP grants through the Office of Innovation and Improvement (OII), and OII has encouraged grantee cooperation (e.g., EDGAR requires grantees to participate in a program evaluation if one is conducted). As noted above, there are two other districts that appear eligible but for whom we are seeking their agreement to participate.

APPENDIX

REVISED POWER CALCULATIONS

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 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 in our list of eligible magnet schools. For this sample, we calculated the average number of students tested in each of math and English Language Arts in the most recent year available, for the magnet schools and the provisional sample of comparison schools. We assumed that (1) 80% of the students at each school would be resident, (2) on average, there were 2.5 years of baseline test-score data available before the year of magnet conversion, and (3) on average there were two comparison schools for each magnet. (The actual sample means were slightly larger, at 2.6 and 2.2 respectively, but we wanted to be somewhat conservative in our estimates.)

One particularly important parameter in the power calculation is ρ, defined as the proportion of total test score variance that is between cohorts within schools. Unfortunately, there is very little published data to help guide a choice ρ. For this parameter, for math we used an estimate of 0.02, which is the median estimate obtained by Bloom (1999) in his study of grade 2 and grade 6 math test scores in Rochester, New York. (He obtained the same estimate for both grades.) For reading, we took a simple average of Bloom's median estimates for grades 2 and 6 in Rochester, plus six other estimates for grade 2 from six other districts around the country, kindly provided by Michael Garet of AIR (with permission of ED). The average of these was 0.022, which is considerably above the Rochester results, of roughly 0.0025. We emphasize that we have used all the estimates of ρ of which we are aware. (We checked with Howard Bloom, for example, and he confirmed that the Rochester estimates in his 1999 paper are the only estimates of which he is aware.)

Another important parameter is the variance across magnet schools in the true effects of converting a school into a magnet, which is referred to as τ^2 in Appendix A of the design document for this study (Bloom, Doolittle, Garet, Christenson and Eaton, 2004). The design document, lacking any information on the value of τ, "guesstimated" a value of 0.01, which is what we have used in our main power calculations. The authors chose this figure on the presumption that a reasonable 95% confidence interval for the true effects of magnets might be -0.05 to 0.35, (centered on a mean effect of 0.15, which as cited elsewhere in their report is the effect size of a full year of school on math achievement and the effects estimated in the Tennessee class-size reduction experiment). The 95% confidence interval suggests τ has a standard deviation of 0.1, and a variance of $0.1^2 = .01$. This estimate of variance in the true effects is fairly large, in the sense that sometimes a school that becomes a magnet performs slightly worse, and in some cases substantially better (+0.35 effect size). This is a conservative estimate in terms of our power analysis because the number of schools needed to obtain a given MDES rises with τ^2.

Our estimates of the number of schools needed to reach a MDES of 0.2 or lower are probably conservative (that is, on the high side). First, recall from above that we assume 2.5 years of baseline data on average and 2 comparison schools per magnet, both of which are below the means of 2.6 and 2.2 respectively. We also assume that only 80% of students tested will be relevant. This is likely to be true for the magnet schools, when we study only resident students. It is less clear to us that we will want to exclude nonresident students from comparison schools, or even that many of the comparison schools will have any nonresident students to speak of. A less conservative but still reasonable estimate is that 90% of tested students could be included in our analysis. Finally, the design document assumed for τ a value of 0.01, which reflects an interest in estimating the average impact for the population of magnet schools from which the schools in the sample were drawn. If we instead set $\tau=0$, we are focusing instead on estimating the average effect for the particular schools in our sample, which may be more appropriate.

Sample Required for an MDES of 0.2 for Resident Students

To calculate the number of magnet schools required to provide a Minimum Detectable Effect Size (MDES) of 0.2 or less, we drew on data from the sample of magnet schools that met study eligibility requirements. In particular, we based several key assumptions on characteristics of the sample, including the number of students tested in each of math and English Language Arts in the most recent year available; the number of years of baseline achievement data available; and the number of comparison schools available.

Based on these calculations, Table 1 shows that, at a minimum, we need 15-16 magnet schools in order to detect an effect size of 0.2 for the overall **resident** student population (pooled across all grades tested).

Table 1 Number of Magnet Schools Needed to Yield a MDES of 0.2 or Less, Based on Characteristics of Magnet Schools Already in Our Sample

Subgroup Size as % of Full Student Sample	***English Language Arts*** **Minimum Number of Magnet Schools Needed**	***Mathematics*** **Minimum Number of Magnet Schools Needed**
20	25	24
30	21	20
40	19	18
50	18	17
Full Resident Sample	**16**	**15**

Notes: Calculations assume subgroups are equally distributed across magnet and comparison schools. MDES based on 80% power and alpha of 0.05. Our calculations are based on the characteristics of all magnet schools and comparison schools in our sample.

Power Calculations Based on Screened and Willing Sample

What is the MDES for the sample of magnet schools that we in fact have recruited?

We have identified 23 eligible magnets (compared to the 16 needed for an MDES of 0.20). All of these are able to provide consistent data for analysis of math achievement, while 21 can provide data for analysis of reading achievement.

Table 2 MDES Based on Characteristics of Magnet Schools Eligible for Inclusion in Our Sample

Subgroup Size as % of Full Resident Student Sample	*English Language Arts* MDES	*Mathematics* MDES
20	0.211	0.209
30	0.194	0.193
40	0.185	0.184
50	0.179	0.178
Full Resident Sample	**0.167**	**0.167**

Summary: Feasibility of a Comparative Interrupted Time Series Study of Resident Students

The power analysis suggests that we should be able to detect effect sizes as small as 0.167 when we test for an overall effect on resident students. We obtain a MDES smaller than 0.20 when we have sub-samples of 30% or even less. This finding opens up the strong possibility that we can obtain fairly precise estimates of the effects of magnetization for students in individual grades, rather than pooled across grades. Alternatively, we could obtain estimates for demographic subgroups when we pool across grades. We will almost certainly be able to test for an effect on non-white students and white students separately. Depending on the demographics in our final sub-sample, we may be able to break down the non-white category at least into its larger subgroups.

-