Evaluation of Supplemental Education Services

Technical Report

Prepared for
Michigan Department of Education

Prepared by:

Public Policy Associates, Incorporated

119 Pere Marquette Drive

Lansing, MI 48912-1231

(517) 485-4477

Fax: 485-4488

March 2009

Table of Contents

Acknowledgments

Public Policy Research,

Development,

and Evaluation

Jeffrey D. Padden

President

119 Pere Marquette

Lansing, MI 48912-1231

517-485-4477

Fax: 485-4488

ppa@publicpolicy.com

www.publicpolicy.com

Introduction

This report describes the methods used in the second of a planned series of annual evaluations of Supplemental Education Services (SES) in Michigan. The evaluation was conducted by Public Policy Associates, Inc. (PPA), a national public policy research, evaluation, and program development firm located in Lansing, Michigan. This report covers the evaluation activities from Fall 2007 to Fall 2008.

SES is provided to students throughout Michigan under the federal No Child Left Behind Act (NCLB). Under NCLB, states are responsible for approving and monitoring providers, and ensuring that providers meet certain quality standards. The evaluation is a key source of data for monitoring the effectiveness and quality of providers. SES provider performance data gathered in this process is disseminated by MDE to parents and school districts throughout the state.

PPA staff and MDE staff collaboratively decided that the evaluation framework and data streams to be used would remain the same as the previous year. This team also adapted instruments from the prior evaluation. However, the administration of some of the data collection changed, and these, along with lessons learned, are discussed throughout this report.

The evaluation methodology included the following:
- An analysis of Michigan Education Assessment Program (MEAP) scores estimated the impact of SES on student achievement in math and English language arts/reading (ELA).
- A survey of parents regarding perceived changes in student behaviors, quality of communication from the SES provider, and overall satisfaction with tutoring.
- A survey of teachers inquiring into the degree and quality of communications from providers, perceived improvements in student performance, and an overall assessment of providers.
- A survey of district SES coordinators looking at how well providers met administrative requirements of their contracts, and perceptions of program quality and program fidelity.

Findings from the evaluation were reported through provider-specific profiles as well as a statewide summary report. This report was designed as an accompaniment to the statewide summary report and provides a detailed review of the data collection and analysis methods.

The report is organized as follows:
- This section, the *Introduction*, provides an overview of the evaluation.
- *Survey Process* describes the various surveys and their implementation.
- *Analysis of Impact of SES on MEAP Scores* reviews the data sources and methods used in a statistical exploration of the impact of SES delivered in 2006-2007 on participants' 2007 math and ELA MEAP scores.
- *Provider Profiles* describes the system by which providers were assigned performance ratings.
- *Recommendations for Change* highlights weaknesses in the evaluation approach and discusses potential improvements for future SES evaluation.

Survey Process

Source Data

Data on SES participants statewide, with some exceptions, are captured in the Center for Educational Performance Information (CEPI) statewide data collection system. The CEPI data, made available to PPA in April 2008, included student name, State of Michigan unique identification code (UIC), date of birth, gender, grade, district name and code, building name and code, and provider name. State staff had updated the CEPI system in 2007 to include three additional variables, which allowed for the identification of the subject of tutoring and the collection of actual service hours for tutoring in (a) math, (b) ELA, and (c) other subjects. However, these service data were not available until the end of the 2008 services.

Six districts were using relational database software to manage the service data for SES. Most of these districts, including the two largest districts, Flint and Detroit Public School District (DPS), were using the CAYEN system, which is a database designed to integrate information on student enrollment and provider services, and to support related administrative functions, such as invoicing.

Because districts must manually enter each student's information into the CEPI database via an online data-entry system, it was unwieldy for the largest districts enter their cases into CEPI. Therefore, PPA received SES data directly from the DPS technology office in April 2008, and from Flint Public School District in May 2008.

The three data files had some issues upon arrival such as missing data and duplicate cases. These were minor issues and subsequent updated data files eventually appeared to reflect the intact and complete cases in the system. Additionally, in the course of processing surveys it became apparent that certain providers' names were variably entered into the data system, complicating the process of reporting provider-specific evaluation findings to MDE.

The state staff ensured that districts completed entering data in CEPI, identifying the service provider for each student enrolled in SES by March. This CEPI list, along with the DPS list and the Flint list of students, formed the sample frame for the teacher and parent surveys.

Since the evaluation design called for it and because the data system is static, it was necessary to pull the sample frame as late in the school year as was practical.

Although MDE's work to advance data collection on student name and provider made the evaluation possible, the SES service-delivery system in the districts was both dynamic and ongoing at the time districts submitted their student lists, resulting in a sample file with notable inaccuracies. For example, the data systems did not necessarily reflect students who signed up but dropped out before services were delivered, students who changed providers, students who needed a new provider because the provider had dropped, and students who were offered SES very late in the school year. Districts varied widely in the timeline with which they implemented

SES, and some with winter and spring service-startup goals were delayed in initiating their programs due to contract problems and other administrative issues. For districts that implemented SES relatively late in the year, the data collected in winter was necessarily preliminary and subject to change.

The data on the number of actual service hours delivered were entered by districts at the end of August 2008, at the close of services. Final DPS service data files were received by PPA in September, and final CEPI service data files in October. These data were matched to the survey sample frame, and merged for use in the Summary Report for 2007-2008.

These service hours data were used to screen for eligibility, and cases in which zero hours of service were provided were removed from analyses of parent and teacher survey data. To be more specific, there were:
- 15,077 cases in the initial sample from CEPI spring 2008
- 3,382 cases with missing or zero hours of service; these cases represent students who enrolled at some point, but did not get services; were in the initial sample frame
- 11,695 de-duplicated cases, matched to initial sample frame, with non-zero service hour data

The initial sample frame drawn from CEPI then reflects the limitations of a point-in-time picture of enrollment. Of cases in the initial sample frame, 22.4% of cases did not receive services. As noted previously, these inappropriate cases were removed from the survey data prior to parent and teacher data analyses for the Summary Report.

Parent Survey

The parent survey was a hardcopy survey delivered via postal mailing to the home addresses of parents or guardians of students listed in various data systems as participants in SES. The survey was revised from the instrument that had been crafted in the previous year, with the input and guidance of MDE staff and an assessment of its performance and utility based on the previous year's data. The survey instrument was printed on one double-sided piece of paper and contained 18 questions, the last being open-ended. It was designed to be machine-readable (scannable) using Remark, a scanning software program, and was personalized using mail merge fields that supplied the students' full name, UIC number (both in numerals and in barcode), and provider name. The survey was written at 5th grade reading level, and its cover letter at the 8th grade reading level. The cover letter was on State of Michigan letterhead and included a footnote in Spanish directing Spanish-speaking parents to call PPA's toll-free number to receive assistance in completing the survey.

The subcontracted printing company packaged the survey, cover letter and a PPA pre-paid business reply envelope in a sealed standard-sized envelope with a PPA return address. The envelope was labeled with both the UIC number and "To the Parent/Guardian of STUDENT NAME."

PPA arranged with the two largest districts, DPS and Flint, to mail the parent survey directly from PPA. This process required these districts to provide files with parent contact information to PPA and freed them from any physical handling of the mailing.

For each of the other school districts in the state, PPA took the stuffed envelopes, grouped them in batches by building and alphabetically by student, and mailed them to District Coordinators with instructions for their distribution (District Coordinators also coordinated the teacher survey process, as described later in this report). District Coordinators were to generate, match, and affix mailing labels to these prepared envelopes. PPA also included extra self-addressed stamped envelopes in anticipation of some parents returning their surveys to the schools rather than using the reply envelope.

Technical assistance to District Coordinators for the parent survey was provided in the form of a written instruction packet that described the materials in the mailing and their responsibilities in the survey process (in most cases, generating the mailing labels and dropping surveys in the mail). PPA had previously contacted each District Coordinator by e-mail or telephone to describe the survey process and their role within it and generally to set their expectations for the evaluation. Technical assistance was lent parents during the fielding of the survey via telephone; both a direct-dial and toll free number to the PPA offices were provided in the cover letter. A Spanish version of the survey was drafted for PPA staff so that staff could conduct the parent survey via telephone as needed. Twelve incoming calls to PPA were from Spanish-speaking parents requesting assistance in completing the survey.

District Coordinators were asked to mail parent surveys so that they would be in the field by May 6, 2008 at the latest. Progress was then verified through follow-up contact by PPA staff (via fax, e-mail or telephone).

There were challenges in fielding the survey from the point of view of executing the mailing. As was to be expected, some addresses on district lists were not current, and this resulted in undeliverable returned mail. While technical assistance was provided to parents, this activity did not present any undue challenge for execution. Most questions and comments pertained to the student's status with a given provider.

To encourage parent participation, MDE sent e-mails to providers requesting that they alert parents that surveys were forthcoming and encourage them to complete the surveys. This provider request included a pre-survey cover letter generically addressed to parents, for providers to share with parents. This was the only role that providers had in the evaluation process, and their completion of their assignment was voluntary and unmonitored. It is uncertain the degree to which this was effective in engendering additional parent participation.

Returned surveys were scanned using Remark software and analyzed in SPSS. The survey earned a 9% response rate or 6% response rate when parents reporting their child did not receive SES and surveys with missing data are excluded. The protocol used to administer the survey as described here was unchanged from the previous year, and therefore the team had expected and prepared for the level of technical assistance required and the logistical challenges involved.

Teacher Survey

Instrument and Mode

A single teacher survey instrument was implemented across all districts and all schools. The survey instrument was adapted from the instrument used in the previous year's evaluation. Minor changes were incorporated in the instrument based on MDE staff input and lessons learned from the data collection of the prior year. Each "teacher survey" was in fact a request for teacher reflections on a particular dyad of student and provider. The instrument asked about their role in relation to the student, type of communications received from provider, changes in student behavior, demeanor, and performance, whether they saw a positive impact, and would recommend the provider.

The data collection was conducted via a secured Web site, which was hosted and maintained by a subcontractor. When invitations to teachers to participate in the survey were sent out electronically by the host site, the "from" address associated with the e-mail was a PPA research assistant e-mail address. This ensured that PPA would be aware of undeliverable addresses that "bounced back" to the sender. The body of the e-mail alerted the reader that "The Michigan Department of Education is working in partnership with Public Policy Associates, Incorporated to evaluate Supplemental Education Services (SES) providers."

Assignment and Invitations

In teacher surveys, the goal was to gather feedback on each student enrolled in SES from a teacher who could speak to particular student's progress. A significant challenge in this process was to identify by name the appropriate teacher, as well as his or her contact information, who could then be invited to respond knowledgeably regarding the particular student's achievements.

Before the survey was implemented, where possible, appropriate teachers and their contact information were obtained from the District Coordinators. However, in some cases, this information was not available beforehand. A significant effort was required for staff to negotiate with districts to get contact information or to determine which mode would work, as well as to negotiate with the subcontractor the functionality required to manage multiple modes of contact, and varying levels of data elements.

The types of scenarios that had to be managed in providing invitations to teachers included the following:
- Known teacher name with known e-mail address
  - *Action:* set up an account and invite them electronically
- Known teacher name, unknown e-mail address
  - *Action*: District Coordinator receives hardcopy invitation per student, and either
    - ◆ Distributes hardcopy, or
    - ◆ Obtains e-mail address and sets up account.

- ■ Unknown teacher in district where all teachers are unknown in advance
  - ● *Action:* District Coordinator receives hardcopy invitation per student.
    - ♦ Must identify teacher, and either
      - ■ distribute hardcopy, or
      - ■ obtain e-mail address and set up account
- ■ Unknown teacher in DPS
  - ● *Action:* Assign Building Coordinator assigned to case electronically
    - ♦ Building Coordinator must identify appropriate teacher, obtain e-mail address, and set up account

Teachers received an e-mail instructing them to access the secured Web site where they could log on to evaluate the impact of SES on the students assigned to them. This method resulted in just a single e-mail request per teacher, regardless of the number of students assigned to them.

For teachers with a valid e-mail address, an automated reminder e-mail was sent to the teacher if their survey had not yet been completed. The subject line of any reminder e-mails was "Urgent SES reminder from MDE."

Interface

In the web-based system, a teacher account was set up in advance for each identified teacher, i.e., for those who where affiliated with a specific student and had contact information. Similarly, the system provided Building Coordinators a management interface permitting access to the status of each survey, and teachers assigned. This also included the right to reassign surveys to another teacher. PPA staff access to the system also included the right to view the status and reassign teachers. The PPA staff were responsible for securely communicating with all these parties their credentials, sometimes repeatedly, which proved to be a time-intensive effort that had not been anticipated.

When teachers with an account initially reached the Web site, they would log in to their existing account, and see a list of all of the students assigned to their username. When teachers without an account initially reached the Web site, they would set up an account. The Web site provided logged-in teachers a listing of the students "assigned" to them, and status information such as whether each survey was completed, unsubmitted, or untouched.

Eligibility and Reassignment

The survey began with a series of questions aimed at understanding if it was appropriate for the teacher to actually complete the survey for the student in question. Specifically, teachers were asked to respond to the following statements:
- ■ I cannot answer a survey about this student because I have a conflict of interest that prevents me from objectively evaluating the provider.
- ■ I cannot answer a survey about this student because the student is not enrolled in this school.
- ■ I cannot answer a survey about this student because this is not one of my students.
- ■ I cannot answer a survey about this student because the student rarely attends my class.
- ■ I cannot answer a survey about this student because I know for certain that the student rarely attended SES.

- I cannot answer a survey about this student because I know for certain that the student was not enrolled in SES.

Any "true" response to these statements prompted an automated process in which (1) the survey for that student case was automatically terminated, (2) the case was coded as a "return", and (3) the Building Coordinators received an e-mail alerting them that a student case needed reassignment to another teacher. On-line, Building Coordinators were able to view the return cases, and in the case of a wrongly assigned teacher, reassign the case to another teacher. The new teacher in turn received an electronic invitation to complete the survey for this case.

Technical Assistance and Challenges

PPA monitored building-specific response rates from surveys and communicated progress to District Coordinators. A contact number for technical or content questions was provided to all teachers in the e-mail invitations and on the Web site. PPA provided technical assistance to both teachers and Building Coordinators during the data collection phase via telephone and e-mail. PPA worked in collaboration with the District Coordinators to address issues on an as-needed basis.

There were several challenges and lessons learned in fielding the survey. The foremost issue was a series of subcontractor delays, some of which went unexplained, and some due to miscommunication. The delay in the final launch of the survey resulted in a severe compression of the window of opportunity for data collection. This resulted in a lower response rate than would otherwise have been possible. From this, the evaluation team learned several lessons, including the value of ensuring an earlier beta version, and using multiple mechanisms to ensure that technical and functional requirements, and system limitations, are well-understood by all parties.

A limitation of this particular Web system was the complication of having multiple modes of initial invitations, meaning some teachers were invited via e-mail and some via hardcopy. This resulted in some teachers mistakenly creating an additional account in the system, and required multiple versions of instructions to coordinators regarding the distribution and management of teacher accounts.

There were known drawbacks to distributing hardcopy invitation letters, chief among them being that PPA did not know which teachers were requested to respond unless they accepted the invitation, went on-line, and set up a user account, and identified in that account the name(s) of the students that they had been assigned. Therefore, in some cases, PPA did not have this linking information, and could not then provide reminders as follow-up, nor could a paper trail indicate whether the papers had been successfully distributed by District Coordinators.

The chief advantage of the e-mail system was the ability to automate the e-mail reminders and to link a student with a particular assigned teacher, thereby allowing targeted follow-up when teachers and their e-mail addresses were known. It also was convenient for teachers to reach PPA staff for technical assistance via e-mail inquiry. On the downside, this system required district staff to maintain a role in the tracking and follow-up contact with individual teachers. This effort was met with varying degrees of commitment and consistency. Dealing with returned

e-mails due to invalid or misspelled e-mail addresses also represented an unanticipated time-intensive effort.

The results from both surveys were merged and analyzed in SPSS. The survey earned a 23% response rate; however, nearly a third of the submitted surveys were entirely composed of missing data and had to be discarded. Further, as mentioned earlier, cases in which parents reported that the child did not receive SES (about 3% of completed teacher surveys) were also discarded. After these exclusions, the response rate was 16%.

District Coordinator Survey

The District Coordinator survey was a two-page survey administered in hard copy and printed on one double-sided piece of paper. The number of surveys given to each District Coordinator was equal to the number of providers in his or her district as reported by districts in CEPI or CAYEN. The hardcopy surveys were not personalized with the provider or district name, but included spaces for capturing this information at the top of the page; a list of providers in each district was submitted along with the surveys for reference in personalizing them.

The District Coordinator surveys were provided to Coordinators via postal mailing. Technical assistance for this survey consisted of a handout entitled "District Coordinator Survey Instructions," which was provided with the surveys, as well as information on PPA staff availability for questions via e-mail and telephone.

The surveys were fielded in April 2008, and due to be completed by most districts by May 30. Reminder e-mails were sent on May 27 and June 13 to District Coordinators who had not returned surveys, requesting a reply as to their status. This was followed with reminder telephone calls in the third week of June.

One issue in the implementation was that District Coordinators did not have full knowledge of providers' contractual and operational information, which was a key feature of the inquiry.

The District Coordinator surveys were manually entered into Microsoft Excel and analyzed in SPSS. The survey earned a 79% response rate.

Analysis of Impact of SES on Michigan Education Assessment Program Scores

Analysis Overview

Unlike other elements of the 2008 SES Evaluation, analysis of the impact of SES on Michigan Education Assessment Program (MEAP) scores focused on services delivered in the 2006-2007 school year. Evaluation of the impact of SES on MEAP scores requires both a pre- and post-services MEAP score, and for students receiving SES in the 2007-2008 school year, no post-services score will be available until the fall 2008 MEAP tests have been taken and processed. Thus, the 2006-2007 school year was the most recent instance of SES delivery that could be evaluated.

The analysis of the impact of SES on MEAP scores was restricted to students in grades 3 through 7 in 2006. Students in other grades did not take MEAP tests in math and English language arts/reading in both 2006 and 2007 and, therefore, could not be included in the analysis.

The analysis was conducted using a hierarchical linear modeling (HLM) approach relying on a matched group of students that did not participate in SES. HLM is the most appropriate form of analysis in many types of educational research as it accommodates "nested" data: that is, where students are grouped into classrooms, school buildings, and districts and these settings are expected to influence student outcomes. Additional information about HLM and the execution of the analysis for the SES evaluation is provided in the sections that follow.

Source Data

Source data for the MEAP analysis included extracts from the 2006 and 2007 statewide research files, containing:
- Raw and scaled MEAP scores for individual students
- Student identifying information including name, unique identification code (UIC), and school-issued student identification number (where such existed)
- The names and codes of the building and district in which the student was enrolled
- The student's grade level
- Demographic information including economic disadvantage status, Limited English Proficiency (LEP status), special education status, Former Limited English Proficiency (FLEP) status (all coded "yes" or "no"), gender, and race.

These data were provided by the Michigan Department of Education for students in grades 3 through 8 in 2006 (747,843 cases) and for the same grades in 2007 (726,192 cases).

The statewide research file does not include information on whether a student received SES. Accordingly, PPA received a separate file from the MDE Office of School Improvement with data on 13,863 students receiving SES in the 2006-2007 school year. The file information included the student name and UIC, grade, date of birth, and SES provider name and was extracted from the Center for Educational Performance Information (CEPI) data collection.

The 2006 and 2007 statewide research files were matched to one another to facilitate use of both the 2006 and 2007 MEAP data. Subsequently, the SES student data was matched to the consolidated statewide research file to connect SES participants to their MEAP data.

Identifying the Sample

Many of the 13,863 SES recipients in 2006-2007 could not be considered in the analysis. Table 1 lists exclusion factors and the number of SES students dropped from the analysis at each stage.[1]

Table 1: Exclusion Factors for Analysis of SES Impact on MEAP Scores

Factors	Number of SES Recipients Dropped (*Total SES Recipients = 13,863*)
Student not in grades 3 – 7 in 2006	10,582
No MEAP record in 2006, 2007, or both*	671
*Missing MEAP data in 2006 or 2007 for subject in question***	
Math	46
ELA	48
*No match available****	
ELA	14
Math	16
*Only student with named provider in grade/building*****	
ELA	784
Math	785

---

[1] The table describes the process of eliminating students that could not be considered in the analysis, but readers interested in data limitations should note that for many students excluded, more than one factor could have been responsible. For example, a student who was retained between 2006 and 2007 might also have had missing MEAP data and been the only SES recipient with a given provider in his or her grade and building. The table describes the number of students excluded at each stage rather than the full count of students to which each exclusion circumstance applied.

Table 1: Exclusion Factors for Analysis of SES Impact on MEAP Scores

Factors	Number of SES Recipients Dropped (*Total SES Recipients = 13,863*)
Total available for analysis	
ELA	1764
Math	1763

Notes:
*This condition refers to the absence of any MEAP record for the student for the given year
**This condition refers to missing math or ELA MEAP scores, although other scores may be present.
***The matching protocol matched one unique student to one unique SES recipient. In some cases, SES students in the building and grade outnumbered nonparticipants, and the available pool of students for matching was exhausted before all SES participants could be assigned a match.
****The HLM analysis of provider-specific results, controlling for school building, cannot be performed in situations where there is only one SES student in a given building and grade.

As Table 1 shows, only approximately 13% of SES recipients could be considered in the MEAP score analysis. The most important exclusion reason was student grade, which eliminated 76% of the SES participants (59% were eliminated for this reason in the analysis for the 2005-2006 school year). Another 5% had missing MEAP data and 6% were eliminated due to requirements of the analysis approach, which focused on both provider and building factors. Embedded within the missing MEAP data category are students with missing or irregular UIC entries as well as students who were retained or promoted more than once between 2006 and 2007 (these students were eliminated from the full body of student MEAP records before these records were matched to the SES student population).

The final count of SES students used in the analysis is shown in Table 2 by grade and subject and overall.

Table 2: Final Count of Useable SES Student Records for MEAP Analysis

Grade	Useable SES Student Records	
	Math	*ELA*
3rd grade	275	271
4th grade	272	270
5th grade	193	193
6th grade	409	409
7th grade	614	621
Totals	1763	1764

Matched Control Group

The analysis relied on a matched control group drawn from the residual population in the merged 2006-2007 statewide research file. The raw data, including SES students, consisted of 747,843 student records for 2006 and 726,192 student records in 2007. Several exclusions were

implemented to the combined database to eliminate students inappropriate for matching. These exclusions included:

- SES students
- Students retained or promoted more than once between 2006 and 2007
- Duplicate records (based on UIC)
- Missing UICs
- Students not attending a school building with at least one SES student
- Students with missing data for the MEAP scores in question

These exclusions reduced the number of records available for matching to 21,154 for the analysis of math scores and to 21,112 for the analysis of ELA scores.

The matching protocol was originally developed in 2007, in consultation with MDE, and considered students' grades, buildings, baseline scaled math/ELA MEAP scores, economic disadvantage status, LEP status, and special education status. The protocol was implemented separately for math and ELA, resulting in different groups of matching students for each subject-matter area.

In order to develop the matched comparison group, SES participants were, by definition, matched to a student in the same grade attending the same building. Within this pool, the matching protocol selected all students with a 2006 scaled MEAP score within two points of the SES student's score (math and ELA considered separately).[2] Subsequently, it was found:

- If more than one student was available, the protocol selected the student with the same economic disadvantage status.
- If more than one student remained available, the protocol selected the student with the same special education status.
- If more than one student remained available, the protocol selected the student with the same LEP status.
- If more than one student remained available, the protocol selected the student with the same FLEP status.
- If more than one student remained available, the first available match was selected without further discrimination between available cases.

These priorities were consistent with the protocol developed last year, in which priorities were established on the basis of bivariate correlations between the variables and the outcomes of interest—the 2006 scaled math and ELA MEAP scores—which showed an extremely strong

---

[2] The selection of matches with a scaled MEAP score within two points, plus or minus, of the SES participant's score represents the only change to the matching protocol between the 2007 evaluation (focused on 2005-2006 scores) and the current evaluation (focused on 2006-2007 scores). For the 2007 evaluation, the matching protocol selected the students(s) with the closest baseline MEAP score, then matched on the basis of other characteristics: if a potential match had an identical MEAP score but differed from the SES participant on other qualities, the protocol preferred that match to one in which a one-point MEAP difference existed but other qualities considered were identical.

The change was implemented in an effort to increase the likelihood that matching students would be comparable on demographic characteristics including economic disadvantage, special education status, and English proficiency, as well as comparable on the baseline MEAP score.

correlation between the year-one and year-two scaled scores and modest relationships for the remaining variables.

No non-SES student was matched to more than one SES student and matches were implemented in the order of the SES students' 2006 scaled scores, with those scoring lowest matched first and those scoring highest matched last. The protocol was implemented using a macro written for SPSS.

Table 3 compares the characteristics of the SES population, by grade and MEAP subject matter, to the characteristics of the matched sample. Table entries in **bold italics** highlight the differences between the groups of four percentage points or greater. The most significant differences are found between SES participants and non-SES matched students in special education status. In all grades, and for analysis of both math and ELA scores, the SES population had a greater percentage of special education students than the matched non-SES students.[3]

Table 3: Comparison of Participants and Matched Control Group on Variables Used to Construct the Match

	Percentage of Group Members With Given Characteristic						2005 Mean MEAP Scores	
	Economic Disadvantage		*Special Education*		*Limited English Proficiency*			
Subject/Grade	*SES*	*Match*	*SES*	*Match*	*SES*	*Match*	*SES*	*Match*
Math								
3rd Grade	91	89	***9***	***5***	13	12	307.76	307.89
4th Grade	92	90	12	11	9	11	400.54	401.99
5th Grade	90	90	***16***	***8***	9	9	491.16	491.29
6th Grade	92	90	15	12	11	10	590.61	590.96
7th Grade	87	87	***14***	***10***	6	6	687.70	687.70
ELA								
3rd Grade	91	92	***10***	***4***	13	11	304.48	305.23
4th Grade	92	89	12	11	10	12	399.88	401.73
5th Grade	90	87	***17***	***8***	8	9	497.76	497.60
6th Grade	***92***	***87***	14	12	11	10	594.94	595.88
7th Grade	87	89	14	11	6	5	693.53	693.38

Exploratory Analysis

A limited exploratory analysis was conducted to confirm that the independent variables had a linear relationship with the dependent variables, to confirm that the dependent variables were normally distributed, and to identify outliers in the data. Scatterplots and histograms were reviewed for evidence of nonnormal distributions, curvilinear relationships, and outliers. All

---

[3] Analysis of variance showed statistically significant differences between the SES and non-SES groups on special education status for 5th grade math and 3rd and 5th grade ELA where $p<0.05$. Differences for 3rd and 7th grade math were statistically significant for $p<0.10$.

histograms revealed reasonably normal distributions in the independent variables and scatterplots revealed presumptively linear relationships between independent and dependent variables. No outliers of a magnitude suggesting exclusion from the analysis were observed.

Hierarchical Linear Modeling

The analysis of the impact of SES on student academic achievement was conducted using HLM 6.04, a software program developed exclusively for hierarchical linear modeling. Michigan's SES program represents a cross-nested structure, where students are grouped in school buildings and with SES providers. Yet, all students in a school building do not necessarily use the same provider and SES providers can and do work with numerous school buildings and districts. Accordingly, the analysis relied on the program's HCM2 model for cross-classified linear models.

The level-one model was specified as follows:

MATHSS_07 $= B_{0jk} + B_{1jk}(\textit{MATHSS}) + B_{2jk}(\textit{SES}) + B_{3jk}(\textit{SE}) + e_{ijk}$

- ■ Where
  - ● j = the student's home school building
  - ● k = the student's SES provider
  - ● B_0 = the intercept term
  - ● $B_{1,2,3}$ = the estimated impact (coefficient) associated with each independent variable
  - ● e = a residual or error term
  - ● *MATHSS_07* = the scaled score for math in 2007 (the dependent variable)
  - ● *MATHSS* = the scaled score for math in 2006[4]
  - ● *SES* = a dummy variable set to 1 for SES participants and 0 for non-SES matching students
  - ● *SE* = a dummy variable set to 1 for special education students and 0 for other students

The level 2 model was specified as follows:

$B_{0}jk = \theta_0 + \textit{SCHOOLCODE}_{00}$

$B_{1}jk = \theta_1 + \textit{SCHOOLCODE}_{10}$

$B_{2}jk = \theta_2 + \textit{PROVIDER}_{20}$

$B_{3}jk = \theta_3$

- ■ Where
  - ● θ_0 = The model intercept

---

[4] This term was grand-mean centered in the analysis.

- **SCHOOLCODE$_{00}$** = the unique increment to the intercept associated with the student's building
- **PROVIDER$_{00}$** = the unique increment to the intercept associated with a specific SES provider
- θ_1 = the model estimate for the impact of 2006 scaled scores on 2007 scaled scores
- **SCHOOLCODE$_{10}$** = the unique increment to the estimate of the impact of 2006 scaled scores on 2007 scaled scores associated with a specific school building
- θ_2 = the model estimate for the impact of SES participation
- **PROVIDER$_{20}$** = the unique increment to the estimate of the impact of SES participation associated with a specific provider
- θ_3 = the model estimate for the impact of special education status

More generally stated, at level 2, the intercept term was specified with random school effects, the 2006 MEAP score (prior achievement) was specified with random school effects, SES participation was specified with random provider effects, and special education status was modeled independently, without mediation by schools or providers.[5] Although other demographic variables were presumed controlled as a function of the matched control group, differences between SES participants and the control group were sizeable enough that the special education term was also incorporated in the model.

The analysis was conducted using standard, default settings in HLM.

Provider Coefficients

One of the recognized benefits of using an HLM approach was that it would specify unique coefficients associated with each provider. These estimates are produced in external SPPS files generated by the software, but must be tested for statistical significance through additional calculations in SPSS or another appropriate software package.

HLM generates an empirical Bayes (EB) parameter estimate and an associated posterior variance (pv) for each column-level variable (here, each specific provider) in the course of executing the model analysis. Confidence intervals around the EB estimate may be generated by multiplying the desired Z score by the square root of the posterior variance and both adding and subtracting the resultant figure from the EB estimate. The equation for a 95% confidence interval is thus:

$$EB \pm \left(1.96 * \sqrt{pv} \right)$$

Where both the minimum and maximum associated with the confidence interval exceeded zero (positive impact) or both the minimum and maximum were less than zero (negative impact), the

---

[5] The 2007 analysis included an interaction term between the baseline MEAP score and SES participation—a slope effect for SES participation. An examination of the proportional reduction in error associated with each of the terms in the model demonstrated that there were virtually no explanatory gains associated with the interaction term. Accordingly, the term was dropped from all analyses, and was not reintroduced for the 2008 analysis.

provider was deemed to have a significant impact on the MEAP scores of SES students in attendance.

Results of this analysis in 2007 found very limited identifiable impact of individual providers' delivery of SES on subsequent MEAP scores.[6] In consultation with MDE, an analysis protocol was developed in which statistically significant provider effects were identified using a range of confidence intervals: the categories considered included "possible impact" (50% confidence interval), "plausible impact" (67% confidence interval), "probable impact" (80% confidence interval) and "highly probable impact" (95% confidence interval). The equations used to generate the upper and lower bounds for each provider-specific parameter were as follows:

- 80% confidence interval: $EB \pm \left(1.28 * \sqrt{pv}\right)$
- 67% confidence interval: $EB \pm \left(0.975 * \sqrt{pv}\right)$
- 50% confidence interval: $EB \pm \left(0.675 * \sqrt{pv}\right)$

In total, 10 providers were associated with a measurable positive or negative impact on the MEAP score. In only one instance did the results for a provider indicate measurable impact in more than one grade level/subject matter combination. One of these impacts was discovered using a 95% confidence interval, three were associated with an 80% confidence interval, two were associated with a 67% confidence interval, and five were associated with a 50% confidence interval. All were associated with 3rd grade ELA, 5th grade ELA, or 3rd grade math.

---

[6] The original protocol, developed in consultation with MDE, was to identify providers with significant impacts at the 95% level of confidence and classify them based on the magnitude of the estimated impact: providers with impacts in excess of 7.50 points (one-half of a standard deviation on the MEAP) would be rated as "A" providers, providers with impacts between 3.75 and 7.50 points (one-quarter to one-half of a standard deviation) would be rated as "B" providers, providers with no measurable positive or negative impacts would be rated as "C" providers, and so on into the negative ranges of coefficient estimates. Initial results showed fewer than ten instances where providers had any measurable impact for a given grade and subject level. After review of the data, MDE hoped to find a means that better discriminated among providers, and ultimately recommended a new approach relying on looser confidence standards and different verbiage to describe the nature of the impact. This approach was retained in 2008.

Provider Profiles

Survey data, results of the MEAP analysis, and descriptive information about the SES program (submitted with providers' applications to serve students in 2007-2008) were drawn on to generate a profile of each SES provider. The profile included an overall grade for the provider as well as selected survey statistics and results of the analysis of the provider's impact on MEAP scores in 2007 for students in each of grades 3 through 7 and for both math and ELA scores.

While most of the data reported in the profile is based on individual survey questions and presented in its original units (i.e., percent "agreeing" or "strongly agreeing," average letter grade for program quality offered by parents), the overall grade considers several data elements, as does one teacher-survey item included in the profile. The methods used to calculate these items were as follows:

- *Letter Grade from Teachers for Effects on Classroom Performance.* The letter grade was derived from responses to seven scaled survey questions:
  - During the time tutoring was provided, did this student's attitude toward class improve, stay the same, or worsen?
  - During the time tutoring was provided, did this student's homework improve, stay the same, or worsen (e.g., quality or timeliness or frequency)?
  - During the time tutoring was provided, did this student's classroom achievement improve, stay the same, or worsen?
  - During the time tutoring was provided, did this student's class attendance improve, stay the same, or worsen?
  - During the time tutoring was provided, did this student's Math grades improve, stay the same, or worsen?
  - During the time tutoring was provided, did this student's ELA grades improve, stay the same, or worsen?

  Each question was scaled "improved," "somewhat improved," "stayed the same," "somewhat worsened," and "worsened." Responses of "improved" or "somewhat improved" were coded as 1, and all other responses coded as zero. The mean across all seven items was calculated for each respondent (the mean for cases with partial missing data was developed on the basis of the available responses), and these respondent-specific means were then averaged at the provider level. Providers in the top 20% of the distribution were assigned an "A" or 4.0, the next 20% were assigned a "B" or 3.0, and so on.

- *Overall Provider Letter Grades.* The overall provider letter grade was a weighted function of the following items:
  - Parent survey
    - "Overall, are you satisfied with this tutor?"
      - Parents could respond "yes," "no," or "not sure." The item grade was based on the percentage of "yes" responses; 90% - 100% was graded as "A" or 4.0, 80% - 89% was graded as "B" or 3.0, 70% - 79% was graded as "C" or 2.0, 60% - 69% was graded as "D" or 1.0, and below 60% was graded as "E" or 0.0.

- ♦ "What overall grade would you give your child's tutor?"
  - ▪ Parents were asked to provide a grade for the tutor's performance overall, with response options of "A," "B," "C," "D," or "E – Failing." Responses were converted to a four-point scale and averaged for each provider.
- ● The teacher letter grade for effects on classroom performance was derived as described in this section, above.
- ● Teacher survey.
  - ♦ "This tutor is positively impacting this student's learning."
    - ▪ Response options included "strongly agree," "agree," "disagree," and "strongly disagree." The percentage of respondents choosing "agree" or "strongly agree" was calculated and providers in the top 20% of the distribution were assigned an "A" or 4.0, providers in the next 20% were assigned a "B" or 3.0, and so on.
- ● MEAP analysis.
  - ♦ All providers were initially assigned a letter grade of "C" (2.0) for MEAP performance overall (including those without any data on MEAP impacts[7]). For findings of impact at the 50% or 67% levels of confidence, providers' grades were adjusted by 1/3 of a grade level (i.e., from "C" to "C-" or from "C-" to "B+") in the direction of the impact for that grade-level/subject matter combination. For findings of impact at the 80% or 95% levels of confidence, providers' grades were adjusted by one full grade for impacts estimated at 7.5 points or less and by two full grades for impacts estimated at greater than 7.5 points.
    - ▪ For example, a finding of a 12-point decrease in MEAP scores at the 95% level of confidence for third-graders in ELA would result in the default grade of "C" being lowered to "E" for that subject/grade. Overall grades for the MEAP were adjusted from the starting value of "C" in recognition of any variation from "C" in any grade/subject combination.
    - ▪ For example, a grade of "C-" for 4th-grade math and "Cs" or "not available" for other grades/subjects would result in an overall grade of "C-" for MEAP in general; a grade of "B+" for 5th grade ELA and of "C+" for 6th grade ELA would result in a grade of "A-," the sum total of all steps away from "C" associated with the provider's grade-level/subject-matter combinations. In general, providers' overall MEAP grades were commensurate with the lowest (or highest) grade for any of the grade-level/subject-matter combinations (e.g., third-grade ELA), as only a handful of providers had any detectable impact on the MEAP, positive or negative.

For the survey-based items, no minimum number of surveys was established; one completed survey was sufficient to establish a grade. The data elements were combined as follows:
- ▪ All five items were weighted at 20%.
- ▪ No grade was calculated for providers with fewer than four of the five elements.

---

[7] The purpose of assigning providers without any data a "C" or 2.0 was to ensure that providers without MEAP data were not advantaged by this fact. MEAP ratings for those providers with useable MEAP data were typically a "C," and in the absence of a comparable statistic for providers without data, would have exerted negative pressure on their overall grades relative to providers without useable MEAP data.

- Where other elements of the score were missing, the weight associated with the remaining items was proportionally upgraded. For example, a provider with four available items would have each weighted at 25% of the total.
- Resulting average ratings on a four-point scale were converted back to letter grades using the following protocol:
  - 3.67 – 4.00 = A
  - 3.34 – 3.66 = A-
  - 3.00 – 3.33 = B+
  - 2.67 – 2.99 = B
  - 2.34 – 2.66 = B-
  - 2.00 – 2.33 = C+
  - 1.67 – 1.99 = C
  - 1.34 – 1.66 = C-
  - 1.00 – 1.33 = D+
  - 0.67 – 0.99 = D
  - 0.33 – 0.66 = D-
  - Below 0.33 = Failing

Recommendations for Change

During this second year of evaluation of Michigan's SES program, several lessons were learned in the course of conducting the work. The following recommendations are focused on ways to improve the evaluation process in the interests of improving data quality and reducing burden on evaluation participants.

Improve the Ability to Identify in Real-Time the Pool of Students Who are Enrolled in SES.
- As it currently exists, CEPI is a static system that is updated sporadically during the year by district staff at the urging of MDE. This means that at any given time, save the end-of-year update, the data system can potentially include students who have dropped out, who never received any service at all, or are assigned to providers who have dropped out. This does not permit an accurate drawing of a sample frame. There appear to be three potential paths that MDE could take to improve identification of students.
  1) MDE implements the use of a real-time database to capture SES data. This would be a longer-term project, and not likely to become available for the next year of the evaluation.
  2) MDE accesses existing CAYEN systems in districts, to facilitate the identification of students for the evaluation. This is a more immediate and realistic aid. This year, it took multiple data file downloads to gain complete and accurate lists. More direct access could facilitate improve timeliness and accuracy of lists.
  3) MDE develop improvements to CEPI to improve timeliness of data. In lieu of developing a real-time data base, improvements to CEPI could be implemented. The critical component would be getting data into the CEPI system, and updating it frequently.
- To have utility for the evaluation, this would need to allow for actual amount of service hours to be entered on an ongoing basis, rather than at the completion of services. This process would also trigger updating of data for students who have dropped out or moved to a different provider.
- Another approach to modifying CEPI would be to increase the capacity to upload files from districts to the CEPI system. For districts with a large number of SES students, this would decrease the burden of documenting SES data, and would likely improve data quality.
- Another possible approach would be for MDE to develop and distribute a database for local use by districts, from which monthly uploads to CEPI could occur, or from which the district sample frame would be directly defined, even if not uploaded into CEPI.

A Rolling Data-Collection Procedure for Parent and Teacher Surveys.
- If a real-time database or data timeliness measure implemented, a rolling data collection process would become feasible. This would be premised upon knowledge of what time of year districts provide services. This would simplify some of the logistics of collecting data on thousands of cases annually, invite input when services are newly completed, and avoid data collection in the summer months when target audiences are likely to be less available and responsive.

■ Move survey data collection process to earlier in the spring, to avoid conflict with end-of-year activity. This is only feasible if services are at or near completion by January or February.
■ Have the online data collection system and protocols developed earlier in the year so that these can be shared at the March meeting of DPS building coordinators, and followed up in the April meeting.
■ Improve functionality of teacher online data collection tool in the following ways.
   ● When status/reason for ineligibility are entered, data should be saved instead of being written over.
   ● Tailor instructions to district coordinators to better match their process of identifying and assigning teachers
   ● Ensure that the online site is thoroughly tested and functioning prior to delivering instructions to teachers
   ● Ensure that vendor delivers on time, so that exhaustive testing of all functionality is completed prior to sharing instructions with districts and teachers

Identify and Implement Ways to Improve Parent Participation in the Survey.
■ Beyond the obvious utility of a cleaner sample frame on SES participants, which would provide an accurate denominator for calculating a response rate, there are other options for increasing the parent response rate:
   ● Continue to consult with districts about the best means of reaching parents of their particular student body and implement a process that allows for district-level variation. Districts were free to send parent surveys home with students or to distribute via postal mail. One suggestion was to distribute and/or have these completed at parent-teacher conferences.
   ● Implement a second mailing to non-responsive parents, a reminder post-card, or similar set of follow-up protocols. A follow-up with nonrespondents is advised to improve response rates.
   ● Identify non-English speaking households of SES students. Assess what language alternatives are needed for parent surveys, and find the resources to offer these. At present only the DPS district has these data. However, in the DPS district, the database element that captures language needs for individual students contains a large amount of missing data. Most recently, this source indicated that, of DPS households, about 63% were English-speaking, over one-third were missing data, 1.5% specified non-English languages, and 1.4% spoke "other" non-English languages. Presumably there are thousands of student households requiring non-English communication, possibly as high as 40% of the SES students in the DPS district. Reaching these households in a relevant language could significantly improve survey response rates and data quality. Implementing this would require a higher level of resources than is currently budgeted.
   ● Consultation with the school districts may also yield additional suggestions for improving response rates.

Identify and Implement Measures to Improve Teacher Participation in the Survey.
■ A cleaner sample frame of SES participants will promote improved feedback from teachers of SES students. Because teachers are often asked to complete multiple surveys, an accurate

sample frame would avert inviting teachers to complete surveys on students that did not actually receive SES.

- Teacher participation could also be enhanced by improving the functionality of the teacher online data collection tool, and the clarity of the instructions.
- Response rates in the teacher survey could also be enhanced by improving the process of identifying and inviting the teachers to participate in the online data collection tool.
  - Provide more lead time to districts to identify the most appropriate teacher to evaluate each student's progress. Help the coordinators to set realistic teacher expectations regarding the survey.
  - Improve the accuracy of e-mail addresses used in the e-mail distribution system. Work with coordinators to communicate the importance of activating school e-mail addresses.
  - The online teacher survey data collection system as implemented was a notable improvement over the prior year system, and eradicated previous issues. For example, the database-driven online system allowed teachers upon log in to view a list of students assigned to them, and the status of each survey. Similarly, it allowed for a process to reassign students to another teacher. However, the system also created other challenges.
- Continue to offer technical assistance at the current level, and use the e-mail and telephone format to deliver the assistance.

Improve the Content of the Parent, Teacher, and District Coordinator Surveys.
- The parent survey tool was improved to help parents document when a child did not ultimately participate in SES. This screening process should be retained in future parent data collection.
- The teacher survey tool had been improved by allowing teachers to document at the outset of the survey that the particular student a) did not receive SES, b) is no longer enrolled at the school, c) has not been in class to the degree necessary to support an evaluation of progress, or d) cannot be evaluated for other reasons. This screening process should be retained in future parent surveys.
- Within districts, responses to some parts of the District Coordinator survey were repetitive; responses appeared routinized, a reflection of a standard contract process. The team needs to explore modifications such that District Coordinators can share information about contract structure once rather than for each provider. This would reduce the burden to District Coordinators. The instrument could retain one or more items to allow understanding of any unique variation across a district's providers. Before finalizing an instrument, District Coordinators should be consulted regarding the reasons for nonresponse to questions about contractor program quality, curriculum alignment, and fidelity to the service plan.

Revise Analysis of MEAP Scores to Accommodate Quantity of Tutoring.
- Summary report data have been improved this year by the availability of data on actual number of service hours provided. These data, however, were not available at the time of the Provider Report Cards. These data will be merged into data sets and will improve the rationality of next year's MEAP analyses. SES could be specified as a ratio variable (amount of variable) as opposed to a simple yes-or-no and would likely yield more reliable and interesting findings. The analysis should also make use of subject-matter data now being captured in CEPI.

Consider Alternatives to Standardized Tests Upon Which Student Progress is Gauged.
- Tests that use pre- and post-service measurement of specific skills targeted in the Individual Learning Plan would be an appropriate and immediate gauge of provider effectiveness, which could be integrated into the assessment of providers.