EngageME P.L.E.A.S.E

Impact Study Results

Prepared by

Program Evaluation Group

University of Georgia

Table of Contents

Introduction

In 2010, Forsyth County Schools (FCS) received an Investing in Innovation (i3) development grant to implement and evaluate a learning management system (LMS) with personalization features. The goal of the study was to determine if providing educators with a learning management system (LMS) with personalization features improves middle and high school academic achievement in Math and English/Language Arts (ELA) at the school level. This study was called EngageME P.L.E.A.S.E (Personalized Learning Experiences Accelerate Standards-based Education).

The tested intervention was the recommendation engine feature of the LMS, which was designed to assist teachers in providing personalized content and activities for students. It suggested resources based on an assessment of individual learning styles, learner preferences, and prior achievement. Teachers could then assign these resources to individual students as a way to remediate or enrich their learning.

Eight middle and four high schools in the district were randomly assigned to condition, with half as treatment schools (4 middle, 2 high), and half as control schools (4 middle, 2 high). At intervention schools, teachers had access to the LMS recommendation engine in all content areas, although the focus of the evaluation was achievement in Math and ELA.

Research Questions

The research questions for the impact study were:

Middle School ELA Domain:

- Does the use of a learning management system with personalization features have a positive impact on ELA achievement in grades 6-8 as measured by state-mandated tests after one year of implementation?

Middle School Math Domain:

- Does the use of a learning management system with personalization features have a positive impact on math achievement in grades 6-8 as measured by state-mandated tests after one year of implementation?

High School ELA Domain:

- Does the use of a learning management system with personalization features have a positive impact on ELA achievement in grades 9-12 as measured by state-mandated tests after one year of implementation?

High School Math Domain:

- Does the use of a learning management system with personalization features have a positive impact on math achievement in grades 9-12 as measured by state-mandated tests after one year of implementation?

Timeline

The funding period for the i3 grant was five years, which included time to develop the LMS and to evaluate its impact on student achievement within intervention schools. The grant was awarded in December 2010 and concluded in September 2015.

2010-11 The primary focus of Year 1 was planning and developing a beta version of the LMS to be tested in pilot schools in Year 2. At the end of Year 1, the district decided to drop the vendor who was developing the beta version of the LMS due to dissatisfaction with how the product was being developed. A second vendor was hired.

2011-12 In Year 2, schools were randomly assigned to treatment and control conditions. A pilot study of the LMS occurred in two schools designated as pilot sites in the spring of Year 2. The beta version of the LMS, which was implemented during the pilot test, did not meet the expectations of the district. Specifically, the LMS did not provide an adequate level of functionality and integration with the district's current data systems, as specified in the contract. The district decided to part ways with the second vendor and hire a third vendor. Because of this change, the impact study, initially scheduled for Year 3, was delayed until Year 4.

2012-13 The primary focus of Year 3 was completing the development of the LMS and recommendation engine with the new vendor. The third vendor developed the LMS which is the focus of this impact study. In the meantime, the district focused on providing professional development for teachers on topics related to creating a more personalized approach to instruction.

2013-14 The impact study occurred in Year 4. The tested intervention was the personalization features of the LMS provided through the recommendation engine. Teachers in treatment schools had access to the recommendation engine, and teachers in control schools did not. However, due to development delays caused by repeated changes in vendors, the recommendation engine was not available until February 2014. Therefore, teachers in treatment schools were not able to access or use the tested intervention until more than halfway through the school year. This delayed roll out date was also only two/three months prior to post-testing (i.e., state-mandated standardized tests, which occurred in April and May of 2014). Focus groups and interviews with teachers at treatment schools indicated that, in general, teachers at treatment schools did not use the recommendation engine before post-testing because 1) it was so late in the school year and 2) they were focused on preparing students for standardized testing and did not want to shift to using a new method. Therefore, while the focus of this impact study is the 2013-14 implementation year, the actual implementation (i.e., use of the recommendation engine) that occurred in this year was minimal.

2014-15 In Year 5, all schools (treatment and control) had access to the LMS and the recommendation engine.

Method

Setting

Forsyth County Schools (FCS) is a suburban district north of Atlanta, Georgia. In October 2013 (the year of the impact study), the district served 38,850 students in 36 schools. The population was 71% White (non-Hispanic), 13% Hispanic, 10% Asian, 3% Black (non-Hispanic), 3% multi-racial. The graduation rate was 88%, while it was 70% for the state of Georgia. In the same year, the free/reduced lunch rate was 17% in comparison to 62% in Georgia. The district has an approximate annual growth rate of 4%.

Design

This study was a school-level randomized control trial where schools were randomly assigned to treatment or comparison group.

All middle and high schools in the district were involved in the study. In Year 2, one middle and one high school participated in the pilot phase; these schools were excluded from the later impact study. The remaining middle (n = 8) and high schools (n = 4) were randomly assigned to the treatment group or to a control/delayed treatment group. The unit of assignment was the school, and all treatment and control schools participated in the study for the duration of the evaluation.

Prior to random assignment, the external evaluation team paired schools based on the percentage of high-needs students according to the classification provided by the school district. FCS considers a student as high needs if he or she has more than four of the following longitudinal and current year indicators: (1) not meeting standards on a state-mandated achievement test, (2) scoring in the 25th percentile or below on the norm-referenced Iowa Test of Basic Skills (ITBS) exams, (3-5) a history of retention, placement into the next grade, or being over-age for the grade level, (6-7) participation in special education or RTI due to learning difficulties, (8) more than four disciplinary incidents, (9) more than four unexcused or seven total absences, (10) failing any portion of the Georgia High School Graduation Test (GHSGT), or (11) failing any class at the high school level causing them to be off track for graduation.

Schools were rank ordered by their percentage of high-needs students and paired (i.e., the two high schools with the highest percentages were paired, as were the two high schools with the lowest percentages). Each school in the pair was assigned a random number, and the school with the lower number in the pair was assigned to the treatment group.

In Year 4 (the year that the impact study was conducted), all schools participating in the study were given access to the LMS, but only schools in the treatment condition had access to the recommendation engine and professional development on how to use it. Control schools could use the LMS, but they did not have access to the recommendation engine, which was the focus of the evaluation.

Participants

The study included eight schools at the middle school level (four treatment and four control) and four schools at the high school level (two treatment and two control). Table 1 depicts participant demographics at the middle school level. The treatment and control samples were similar in terms of gender, grade level, ethnicity, and percentage of high-needs students.

Table 2 depicts participant demographics at the high school level. The treatment and control samples were similar in terms of gender and grade level. Some differences across the samples were observed in terms of race/ethnicity and high-needs students. For ELA, there were somewhat fewer white and Hispanic students and more Asian students in the treatment sample. For math, there were somewhat fewer Hispanic and more Asian students in the treatment sample. Overall, there were fewer high-needs students in the treatment sample.

Table 1
Participant Demographics, Middle School

		ELA		Math	
		Treatment	Control	Treatment	Control
All	N	3839	4020	3811	3992
Female	N	1861	2066	1850	2054
	%	48%	51%	49%	51%
Male	N	1978	1954	1961	1938
	%	52%	49%	51%	49%
Grade 6	N	1211	1371	1195	1360
	%	32%	34%	31%	34%
Grade 7	N	1368	1360	1362	1354
	%	36%	34%	36%	34%
Grade 8	N	1260	1289	1254	1278
	%	33%	32%	33%	32%
White	N	2926	2834	2903	2809
	%	76%	70%	76%	70%
Asian	N	419	325	420	327
	%	11%	8%	11%	8%
Black	N	79	119	77	117
	%	2%	3%	2%	3%
Hispanic	N	310	619	307	620
	%	8%	15%	8%	16%
Indian	N	10	310	9	17
	%	0.30%	0.40%	0.20%	0.40%
Mixed	N	94	101	94	97
	%	2%	3%	2%	2%
Pacific Islander	N	1	5	1	5
	%	0.03%	0.01%	0.03%	0.10%
High Needs	N	283	264	236	262
	%	7%	7%	6%	7%

Table 2
Participant Demographics, High School

		ELA		Math	
		Treatment	Control	Treatment	Control
All	N	2345	1801	1858	1493
Female	N	1163	877	955	732
	%	50%	49%	51%	49%
Male	N	1182	924	903	761
	%	50%	51%	49%	51%
Grade 9	N	1276	1001	1083	918
	%	54%	56%	58%	62%
Grade 10	N	6	18	759	562
	%	0.30%	1%	41%	38%
Grade 11	N	1062	779	16	12
	%	45%	43%	0.90%	0.80%
Grade 12	N	1	3	0	1
	%	0.04	0.20%	0	0.10%
White	N	1716	1420	1408	1149
	%	73%	79%	76%	77%
Asian	N	297	42	144	34
	%	13%	2%	34%	2%
Black	N	52	51	61	47
	%	2%	3%	3%	3%
Hispanic	N	200	230	181	218
	%	9%	13%	10%	15%
Indian	N	8	9	11	7
	%	0.30%	0.50%	0.60%	0.50%
Mixed	N	70	49	52	36
	%	3%	3%	3%	2%
Pacific Islander	N	2	0	1	2
	%	0.10%	0	0.10%	0.10%
High Needs	N	194	222	172	199
	%	8%	12%	9%	13%

Measures

The measures for the impact study were student achievement scores in Math and ELA on state-mandated standardized tests. The Criterion-Referenced Competency Tests (CRCT) were the middle school assessments, and the End Of Course Tests (EOCTs) were the high school assessments. Because they were state-level assessments, they are assumed to meet standards for reliability and validity.

All test scores were standardized to address differences in the tests across grades. We calculated z-scores for each test (e.g., a z-score for 6th grade Math CRCT). All subsequent analyses were conducted using these z-scores.

Middle school measures. For middle school, the outcome measures were CRCT scores in Math and ELA for 6th, 7th, and 8th grade. All post-test measures were completed in April of the 2013-14 school year. The baseline measures were CRCT scores from the prior year, 2012-13. The scores were from the CRCTs administered when students were in 5th, 6th and 7th grade.

High school measures. All post-test measures were completed in May of the 2013-14 school year for both subject areas. EOCTs were administered in selected Math and ELA courses. For Math, the courses were Math I (later re-named to Coordinate Algebra) and Math II (later re-named to Analytic Geometry). Most students took these courses in 9th grade and 10th grade, respectively. For ELA, the courses were 9th grade Literature and American Literature. Most students took these courses in 9th grade and 11th grade, respectively.

Baseline measures for Math were the state-administered assessments from the prior school year (i.e., 2012-13). For Math I, the baseline measure was the 8th grade Math CRCT score. For Math II, the baseline measure was the Math I score.

For ELA, the baseline measure for 9th grade Literature was the 8th grade CRCT score collected in 2012-13. For American Literature, the baseline measure was the 9th grade Literature score. In most cases, this score was from 2011-12, because of the gap between when most students took 9th grade Literature (9th grade) and American Literature (11th grade).

Covariates. Covariates were measured at student level and the school level. *Student-level covariates* were grade level, high-need status, gender, ethnicity, and pre-test scores. *School-level covariates* were school size and pre-test scores.

Analytic Approach

Baseline equivalence. Hierarchical linear modeling was used to evaluate the baseline equivalence of students at the middle school level. The equivalence of the treatment and control groups prior to the implementation of the study was examined using a two-level structure of students nested within schools. The following model was used for this baseline equivalence testing:

Level 1:

$$Y_{pre.ij} = \beta_{0j} + \varepsilon_{ij}$$

Level 2:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(Trt) + \mu_{0j}$$

where $Y_{pre.ij}$ is the pre-test measure of the i^{th} student in the j^{th} school, *Trt* is the dummy variable for treatment (coded 1 for treatment and 0 for control), ε_{ij} represents residual error (assumed to be normally distributed with mean = 0 and variance = σ^2), and μ_{0j} = is a random intercept term for the jth school (assumed to be normally distributed with mean = 0 and variance = τ^2, and independent of ε_{ij}).

By fitting the above model to the data, $\hat{\gamma}_{01}$ is the estimated treatment-control difference in the pre-test measure. This is reported as the mean difference between the treatment and control groups. A standardized difference was also calculated, which is the mean difference divided by the pooled standard deviation of the treatment and control groups, and reported as the effect size.

Impact testing. The impact of the intervention was estimated using a two-level structure of students nested within schools. The following model was used for this impact testing:

Level 1:

$$Y_{ij} = \beta_{0j} + \beta_{1j}(Y_{pre.ij}) + \sum_{k=1}^{K} \beta_{k+1}(x_{ijk}) + \varepsilon_{ij}$$

Level 2:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(Trt) + \mu_{0j}$$
$$\beta_{1j} = \gamma_{10}$$
$$\cdots$$
$$\beta_{1+k,j} = \gamma_{1+k,0}$$

where

Y_{ij} is an outcome measure of the i^{th} student in the j^{th} school, $Y_{pre.ij}$ is the baseline, pre-test measure of the i^{th} student in the j^{th} school, *Trt* is the dummy variable for treatment (coded 1 for treatment and 0 for control), x_k represents k=1…K student-level covariates (e.g., demographics) included in the model, ε_{ij} represents residual error (assumed to be normally distributed with mean = 0 and variance = σ^2, and μ_{0j} is a random intercept term for the jth school (assumed to be normally distributed mean = 0 and variance = τ^2, and independent of ε_{ij}).

Covariates. We controlled for significant covariates in the impact analysis. First, we determined which covariates should be included in the final impact models using a backward elimination procedure. Initially, all variables are entered into the model. We then ran the model and identified the covariate with the highest *p*-value. If the *p*-value was higher than .20, the covariate

was removed and the process was repeated, eliminating covariates one by one. The process stopped when all remaining covariates were significant at the .20 level.

At the middle school level, we tested both student-level and school-level covariates. The tested *student-level covariates* were grade level, high-need status, gender, ethnicity, and pre-test scores. The tested *school-level covariates* were school size and pre-test scores. For both content areas, all five student-level covariates were significant at the .20 level. School-level pre-test scores were significant but school size was not. Therefore, the final models for ELA and Math included all five student-level covariates plus school-level pre-test scores.

At the high school level, we only tested for student-level covariates because the smaller number of high schools resulted in not enough degrees of freedom to test school-level covariates. For ELA and Math, all five student-level covariates were significant and were included in the final impact models.

Missing Data. Students with pre-test scores were included in the baseline analysis. Students with pre-test and post-test scores were included in the impact analysis. Listwise deletion was used in cases when students were missing data on any of the covariates included in the model; however, this was very rare as the covariate data were nearly complete.

Program Implementation

Fidelity of program implementation was assessed by the external evaluator, the Program Evaluation Group at the University of Georgia. Five key components of fidelity were identified:

1. Integrated learning platform and data management system (LMS) recommendation engine
2. Professional learning support teams: Train the trainer
3. Teacher professional development
4. School level support
5. District level support

Information to determine fidelity was compiled from teacher surveys, administrator surveys, student surveys, site visit observations, interviews, focus groups, and school/district records.

Table 3 presents the five key fidelity components, their definitions, and findings from the year of the impact study. Implementation of the program met fidelity in four out of the five areas evaluated. The first area, the LMS recommendation engine, was the sole area where fidelity was not met. This was due to the vendor changes and product development delays that caused the LMS not to be available at the beginning of the school year. Because of these delays, teachers in treatment schools did not have access to the intervention until February 2014. This delayed roll out date was two/three months prior to post-testing. Focus groups and interviews with teachers at treatment schools indicated that, in general, teachers did not use the recommendation engine before post-testing because they were focused on preparing students for standardized testing and did not want to shift to using a new method. Therefore, the actual use of the recommendation engine that occurred in the impact year was minimal.

Table 3

Fidelity of Program Implementation

Key Components	Definitions		Findings	
	Number of Indicators	Threshold for fidelity	% meeting fidelity threshold for component	Was project fidelity for component reached?
Integrated learning platform and data management system LMS (Input) ***Key component is Recommendation Engine*	4	100% of schools have score of 4 (out of a total possible of 4)	0 schools met fidelity. The recommendation engine was not available at the beginning of the school year due to vendor changes and product development delays.	No
Develop Professional Learning Support Teams: Train the Trainer(Activity)	6	67% schools score 10 (out of a total possible of 14)	6 schools met fidelity-100%	Yes
Provide Teacher Professional Development (Activity)	8	67% schools score 16 (out of a total possible of 24)	6 schools met fidelity-100%	Yes
School Level Support (Activity)	7	67% schools score 12 (out of a total possible of 19)	6 schools met fidelity-100%	Yes
District Level Support (Activity)	5	67% schools score 10 (out of a total possible of 15)	6 schools met fidelity-100%	Yes

Results

Baseline Equivalence

Baseline equivalence of students was evaluated at the middle-school level. Table 4 presents the sample sizes, means, and standard deviations of the pretest standardized scaled test scores for the baseline sample.

The treatment-comparison difference for ELA was .0487. The treatment-comparison difference for Math was .1022.

Table 4
Pretest Standardized Scaled Scores

	Middle School					
	ELA			Math		
	M	*SD*	*N*	*M*	*SD*	*N*
Treatment	0.044	0.994	3839	0.072	0.99	3811
Control	-0.042	1	4020	-0.068	1	3992

Impact Estimates

Table 5 presents the impact estimates for the final models.

Table 5

Final Impact Estimates

	Treatment			Comparison			Impact					
Contrast Name	N of Clusters	N of Students	Unadjusted SD	N of Clusters	N of Students	Unadjusted SD	Unadjusted Mean	Impact Estimate	Standardized Effect Size	Impact Standard Error	p-value	df
ELA Middle	4	3839	0.992	4	4020	1.0072	-0.014	-0.0526	-0.0526	0.032	0.159	5
Math Middle	4	3811	0.9957	4	3992	0.9997	-0.0646	-0.01	-0.01	0.057	0.865	5
ELA High	2	2345	0.8574	2	1801	1.1322	-0.187	0.12366	0.12809	0.0805	0.264	2
Math High	2	1858	0.9716	2	1493	0.9542	-0.2957	0.23004	0.24436	0.103	0.155	2

Middle School ELA. For middle school ELA, after controlling for significant covariates, treatment was not a significant predictor of ELA achievement. The value of the impact estimate was -.0526 with a *p*-value of .159.

Middle School Math. For middle school Math, after controlling for significant covariates, treatment was not a significant predictor of math achievement. The value of the impact estimate was -.01 with a *p*-value of .865.

High School ELA. For high school ELA, after controlling for significant covariates, treatment was not a significant predictor of ELA achievement. The value of the impact estimate was 0.12366 with a *p*-value of .264.

High School Math. For high school Math, after controlling for significant covariates, treatment was not a significant predictor of math achievement. The value of the impact estimate was .23 with a *p*-value of .155.

Summary

No significant differences were found between treatment and control schools on student achievement in ELA and Math given the tested intervention. This result was not unexpected given the context of the program implementation. Because of delays in product development and changes in vendors, the intervention was not available at the beginning of the impact year. Further, teachers in treatment schools generally stated that they did not use the LMS once it was available because it was too late in the year to use it effectively. Given the very low dosage of the intervention, it is unsurprising that outcomes were similar for treatment and control schools.

Appendix A: Final Analytical Models

The middle school ELA final analytical model was:

Level-1 Model

$$ELA\_z\text{-}score\_posttest14_{ij} = \beta_{0j} + \beta_{1j}*(Grade_{ij}) + \beta_{2j}*(ATRISKCO_{ij}) + \beta_{3j}*(Female_{ij}) + \beta_{4j}*(Asian_{ij}) + \beta_{5j}*(Black_{ij}) + \beta_{6j}*(Hispanic_{ij}) + \beta_{7j}*(AmericanIndian_{ij}) + \beta_{8j}*(Mixed_{ij}) + \beta_{9j}*(PacificIslander_{ij}) + \beta_{10j}*(ELA\_z\text{-}score\_pretest\ 13_{ij}) + r_{ij}$$

Level-2 Model

$$\beta_{0j} = \gamma_{00} + \gamma_{01}*(Treatment_j) + \gamma_{02}*(ELA\_z\text{-}score\_posttest13\_Mean_j) + u_{0j}$$
$$\beta_{1j} = \gamma_{10}$$
$$\beta_{2j} = \gamma_{20}$$
$$\beta_{3j} = \gamma_{30}$$
$$\beta_{4j} = \gamma_{40}$$
$$\beta_{5j} = \gamma_{50}$$
$$\beta_{6j} = \gamma_{60}$$
$$\beta_{7j} = \gamma_{70}$$
$$\beta_{8j} = \gamma_{80}$$
$$\beta_{9j} = \gamma_{90}$$
$$\beta_{10j} = \gamma_{100}$$

The middle school Math final analytical model was:

Level-1 Model

$$Math\_z\text{-}score\_posttest14_{ij} = \beta_{0j} + \beta_{1j}*(Grade_{ij}) + \beta_{2j}*(ATRISKCO_{ij}) + \beta_{3j}*(Female_{ij}) + \beta_{4j}*(Asian_{ij}) + \beta_{5j}*(Black_{ij}) + \beta_{6j}*(Hispanic_{ij}) + \beta_{7j}*(AmericanIndian_{ij}) + \beta_{8j}*(Mixed_{ij}) + \beta_{9j}*(PacificIslander_{ij}) + \beta_{10j}*(Math\_z\text{-}score\_pretest\ 13_{ij}) + r_{ij}$$

Level-2 Model

$$\beta_{0j} = \gamma_{00} + \gamma_{01}*(Treatment_j) + \gamma_{02}*(Math\_z\text{-}score\_pretest13\_ME_j) + u_{0j}$$
$$\beta_{1j} = \gamma_{10}$$
$$\beta_{2j} = \gamma_{20}$$
$$\beta_{3j} = \gamma_{30}$$
$$\beta_{4j} = \gamma_{40}$$
$$\beta_{5j} = \gamma_{50}$$
$$\beta_{6j} = \gamma_{60}$$
$$\beta_{7j} = \gamma_{70}$$
$$\beta_{8j} = \gamma_{80}$$
$$\beta_{9j} = \gamma_{90}$$
$$\beta_{10j} = \gamma_{100}$$

The high school ELA final analytical model was:

Level-1 Model

$Language\_z\text{-}score\_posttest14_{ij} = \beta_{0j} + \beta_{1j}*(Grade_{ij}) + \beta_{2j}*(ATRISKCO_{ij}) + \beta_{3j}*(Female_{ij}) + \beta_{4j}*(Asian_{ij}) + \beta_{5j}*(Black_{ij}) + \beta_{6j}*(Hispanic_{ij}) + \beta_{7j}*(AmericanIndian_{ij}) + \beta_{8j}*(Mixed_{ij}) + \beta_{9j}*(PasificIslander_{ij}) + \beta_{10j}*(Language\_z\text{-}score\_pretest\ 13_{ij}) + r_{ij}$

Level-2 Model

$\beta_{0j} = \gamma_{00} + \gamma_{01}*(Treatment_j) + u_{0j}$
$\beta_{1j} = \gamma_{10}$
$\beta_{2j} = \gamma_{20}$
$\beta_{3j} = \gamma_{30}$
$\beta_{4j} = \gamma_{40}$
$\beta_{5j} = \gamma_{50}$
$\beta_{6j} = \gamma_{60}$
$\beta_{7j} = \gamma_{70}$
$\beta_{8j} = \gamma_{80}$
$\beta_{9j} = \gamma_{90}$
$\beta_{10j} = \gamma_{100}$

The high school Math final analytical model was:

Level-1 Model

$Math\_z\text{-}score\_posttest14_{ij} = \beta_{0j} + \beta_{1j}*(Grade_{ij}) + \beta_{2j}*(ATRISKCO_{ij}) + \beta_{3j}*(Female_{ij}) + \beta_{4j}*(Asian_{ij}) + \beta_{5j}*(Black_{ij}) + \beta_{6j}*(Hispanic_{ij}) + \beta_{7j}*(\mathrm{AmericanIndian}_{ij}) + \beta_{8j}*(Mixed_{ij}) + \beta_{9j}*(PasificIslander13_{ij}) + \beta_{10j}*(Math\_z\text{-}score\_pretest13_{ij}) + r_{ij}$

Level-2 Model

$\beta_{0j} = \gamma_{00} + \gamma_{01}*(Treatment_j) + u_{0j}$
$\beta_{1j} = \gamma_{10}$
$\beta_{2j} = \gamma_{20}$
$\beta_{3j} = \gamma_{30}$
$\beta_{4j} = \gamma_{40}$
$\beta_{5j} = \gamma_{50}$
$\beta_{6j} = \gamma_{60}$
$\beta_{7j} = \gamma_{70}$
$\beta_{8j} = \gamma_{80}$
$\beta_{9j} = \gamma_{90}$
$\beta_{10j} = \gamma_{100}$

Appendix B: Final Reporting Shells for the National Evaluation of i3

Table 1: Master List of Contrasts

Replace with the Contrast Table for Grant attached to the September E-mail from the AR Team. Contrast table in e-mail will have numbers for the confirmatory contrasts. You need to add numbers for any exploratory contrasts added to the table.

Table 1: Contrasts for ELA Outcomes

NOTE: In this example table, contrast IDs that are confirmatory have "C-" prefixes on contrast IDs. Exploratory contrasts have "E-" for confirmatory contrast try to use the contrast IDs assigned by the AR team. Be clear about which contrasts are confirmatory and which are exploratory.

			Treatment Group			Comparison Group	Outcome				Baseline		
Contrast ID #a	Contrast Name [Expected Reporting Date]	Design	[Condition] Description	Age/grade during intervention	Exposure	[Condition] Description	Domain	Measure [Scale]	Unit of Observation	Timing of Measurement	Measure [Scale]	Unit of Observation	Timing of Measurement
	1. General Reading Achievement (MS) [Fall 2014]	Cluster RCT with assignment at the school level	[Engage Me Middle School] All grade 6-8 students in 2013-14 intervention schools a. 6th b. 7th c. 8th	a. 6th b. 7th c. 8th	1 year	[Business-As-Usual Middle School] All grade 6-8 students in 2013-14 comparison schools a. 6th b. 7th c. 8th	General Reading Achievement (Middle School)	Criterion-Referenced Competency Test (CRCT) English Language Arts (state standardized test) [Continuous]	Student	Spring 2014 a. end of 6th b. end of 7th c. end of 8th	CRCT English Language Arts (state standardized test) [Continuous]	Student	Spring 2013 a. end of 5th b. end of 6th c. end of 7th

Contrast ID #[a]	Contrast Name [Expected Reporting Date]	Design	Treatment Group			Comparison Group	Outcome				Baseline		
			[Condition] Description	Age/grade during intervention	Exposure	[Condition] Description	Domain	Measure [Scale]	Unit of Observation	Timing of Measurement	Measure [Scale]	Unit of Observation	Timing of Measurement
	2. General Mathematics Achievement (MS) [Fall 2014]	Cluster RCT with assignment at the school level	[Engage Me Middle School] All grade 6-8 students in 2013-14 intervention schools a. 6th b. 7th c. 8th	a. 6th b. 7th c. 8th	1 year	[Business-As-Usual Middle School] All grade 6-8 students in 2013-14 comparison schools a. 6th b. 7th c. 8th	Mathematics Achievement (Middle School)	CRCT math (state standardized test) [Continuous]	Student	Spring 2014 a. end of 6th b. end of 7th c. end of 8th	CRCT Math (state standardized test) [Continuous]	Student	Spring 2013 a. end of 5th b. end of 6th c. end of 7th
	3. General Literacy Achievement (HS) [Fall 2014]	Cluster RCT with assignment at the school level	[Engage Me High School] All grade 9-12 students in 2013-14 intervention schools a. 9th b. 10th c. 11th d. 12th	a. 9th b. 10th c. 11th d. 12th	1 year	[Business-As-Usual High School] All grade 9-12 students in 2013-14 comparison schools a. 9th b. 10th c. 11th d. 12th	General Literacy Achievement (High School)	End of Course Tests (EOCT): i.) 9th grade literature and composition ii.) American literature and composition (Assume this is two distinct tests associated with two courses: (a) 9th grade literature and composition and (b)	Student	Spring 2014, end of course for all students in grades 9-12 in the associated course a. end of 9th b. end of 10th	"Pretest English/Language Arts achievement (state test)" (Assume this means state grade 8 CRCT ELA test for 9th grade literature and composition outcome and 9th grade literature and composition test for	Student	Spring 2013 or Spring 2012 a. end of 8th b. end of 9th c. end of 10th d. end of 11th (Assumes prior year

Contrast ID #[a]	Contrast Name [Expected Reporting Date]	Design	Treatment Group			Comparison Group	Outcome				Baseline		
			[Condition] Description	Age/grade during intervention	Exposure	[Condition] Description	Domain	Measure [Scale]	Unit of Observation	Timing of Measurement	Measure [Scale]	Unit of Observation	Timing of Measurement
								American literature and composition.) [Continuous]		c. end of 11th d. end of 12th (Assumes courses are full year with EOCT at the end)	American literature and composition outcome.) [Continuous]		courses are full year with EOCT at the end)
	4. Mathematics Achievement (HS) [Fall 2014]	Cluster RCT with assignment at the school level	[Engage Me High School] All grade 9-12 students in 2013-14 intervention schools a. 9th b. 10th c. 11th d. 12th	a. 9th b. 10th c. 11th d. 12th	1 year	[Business-As-Usual High School] All grade 9-12 students in 2013-14 comparison schools a. 9th b. 10th c. 11th d. 12th	Mathematics Achievement (High School)	End of Course Tests (EOCT) : i.) Math I ii.) Math II (Assume this is two distinct tests associated with two courses: (a) Math I and (b) Math II.) [Continuous]	Student	Spring 2014, end of course for all students in grades 9-12 in the associated course a. end of 9th b. end of 10th c. end of 11th d. end of 12th	"Pretest math achievement (state test)" (Assume this means grade 8 state CRCT Math test for Math I outcome and Math I EOCT for Math II outcome.) [Continuous]	Student	Spring 2013 a. end of 8th b. end of 9th c. end of 10th d. end of 11th (Assumes prior year courses are full year with EOCT at the end)

Contrast ID #[a]	Contrast Name [Expected Reporting Date]	Design	Treatment Group			Comparison Group	Outcome				Baseline		
			[Condition] Description	Age/grade during intervention	Exposure	[Condition] Description	Domain	Measure [Scale]	Unit of Observation	Timing of Measurement	Measure [Scale]	Unit of Observation	Timing of Measurement
										(Assumes courses are full year with EOCT at the end)			

Table 2: Impact Estimates
Table 2: Impact Estimates

Table 2: Impact Estimates

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
Contrast ID #	Contrast Name (Optional)	Post-test Measure Name	Treatment Group N of Clusters	Treatment Group N of Students	Comparison Group N of Clusters	Comparison Group N of Students	Unadjusted Treatment Group SD	Unadjusted Comparison Group SD	Standard Deviation Source (Code)	Unadjusted Comparison Group Mean (Optional)	Impact Estimate	Standardized Effect Size (Optional)	Impact Standard Error	p-value	Code for Impact Model Description	Degrees of Freedom	Source of Data (Optional)	Level of Inference (Optional)
	ELA Middle	CRCT	4	3839	4	4020	.9920	1.0072	A.1	-.0140	-.0526	-.0526	.032	.159	B	5		
	Math Middle	CRCT	4	3811	4	3992	.9957	.9997	A.1	-.0646	-.01	-.01	.057	.865	C	5		
	ELA High	9TH Grade Lit/ American Lit	2	2345	2	1801	.8574	1.1322	A.1	-.1870	.12366	.12809	.0805	.264	D	2		
	Math High	Coordinate Algebra / Analytic Geometry	2	1858	2	1493	.9716	.9542	A.1	-.2957	.23004	.24436	.103	.155	E	2		

A.1. Student level SDs calculated from Study data (full sample), and z-scored prior to analysis (so that SD =1)

B. The model used to estimate this impact is shown below

Level-1 Model

$$ELA\_z\text{-}score\_posttest14_{ij} = \beta_{0j} + \beta_{1j}*(Grade_{ij}) + \beta_{2j}*(ATRISKCO_{ij}) + \beta_{3j}*(Female_{ij}) + \beta_{4j}*(Asian_{ij}) + \beta_{5j}*(Black_{ij}) + \beta_{6j}*(Hispanic_{ij}) + \beta_{7j}*(AmericanIndian_{ij}) + \beta_{8j}*(Mixed_{ij}) + \beta_{9j}*(PacificIslander_{ij}) + \beta_{10j}*(ELA\_z\text{-}score\_pretest\ 13_{ij}) + r_{ij}$$

Level-2 Model

$$\beta_{0j} = \gamma_{00} + \gamma_{01}*(Treatment_j) + \gamma_{02}*(ELA\_z\text{-}score\_posttest13\_Mean_j) + u_{0j}$$
$$\beta_{1j} = \gamma_{10}$$
$$\beta_{2j} = \gamma_{20}$$
$$\beta_{3j} = \gamma_{30}$$
$$\beta_{4j} = \gamma_{40}$$

$\beta_{5j} = \gamma_{50}$
$\beta_{6j} = \gamma_{60}$
$\beta_{7j} = \gamma_{70}$
$\beta_{8j} = \gamma_{80}$
$\beta_{9j} = \gamma_{90}$
$\beta_{10j} = \gamma_{100}$

C. The model used to estimate this impact is shown below

Level-1 Model

$Math\_z\text{-}score\_posttest14_{ij} = \beta_{0j} + \beta_{1j}*(Grade_{ij}) + \beta_{2j}*(ATRISKCO_{ij}) + \beta_{3j}*(Female_{ij}) + \beta_{4j}*(Asian_{ij}) + \beta_{5j}*(Black_{ij}) + \beta_{6j}*(Hispanic_{ij}) + \beta_{7j}*(AmericanIndian_{ij}) + \beta_{8j}*(Mixed_{ij}) + \beta_{9j}*(PacificIslander_{ij}) + \beta_{10j}*(Math\_z\text{-}score\_pretest\ 13_{ij}) + r_{ij}$

Level-2 Model

$\beta_{0j} = \gamma_{00} + \gamma_{01}*(Treatment_j) + \gamma_{02}*(Math\_z\text{-}score\_pretest13\_ME_j) + u_{0j}$
$\beta_{1j} = \gamma_{10}$
$\beta_{2j} = \gamma_{20}$
$\beta_{3j} = \gamma_{30}$
$\beta_{4j} = \gamma_{40}$
$\beta_{5j} = \gamma_{50}$
$\beta_{6j} = \gamma_{60}$
$\beta_{7j} = \gamma_{70}$
$\beta_{8j} = \gamma_{80}$
$\beta_{9j} = \gamma_{90}$
$\beta_{10j} = \gamma_{100}$

D. The model used to estimate this impact is shown below

Level-1 Model

$Language\_z\text{-}score\_posttest14_{ij} = \beta_{0j} + \beta_{1j}*(Grade_{ij}) + \beta_{2j}*(ATRISKCO_{ij}) + \beta_{3j}*(Female_{ij}) + \beta_{4j}*(Asian_{ij}) + \beta_{5j}*(Black_{ij}) + \beta_{6j}*(Hispanic_{ij}) + \beta_{7j}*(AmericanIndian_{ij}) + \beta_{8j}*(Mixed_{ij}) + \beta_{9j}*(PacificIslander_{ij}) + \beta_{10j}*(Language\_z\text{-}score\_pretest\ 13_{ij}) + r_{ij}$

Level-2 Model

$\beta_{0j} = \gamma_{00} + \gamma_{01}*(Treatment_j) + u_{0j}$

$\beta_{1j} = \gamma_{10}$

$\beta_{2j} = \gamma_{20}$

$\beta_{3j} = \gamma_{30}$

$\beta_{4j} = \gamma_{40}$

$\beta_{5j} = \gamma_{50}$

$\beta_{6j} = \gamma_{60}$

$\beta_{7j} = \gamma_{70}$

$\beta_{8j} = \gamma_{80}$

$\beta_{9j} = \gamma_{90}$

$\beta_{10j} = \gamma_{100}$

E. The model used to estimate this impact is shown below

Level-1 Model

$Math\_z\text{-}score\_posttest14_{ij} = \beta_{0j} + \beta_{1j}*(ATRISKCO_{ij}) + \beta_{2j}*(Asian_{ij}) + \beta_{3j}*(Black_{ij}) + \beta_{4j}*(Hispanic_{ij}) + \beta_{5j}*(\text{AmericanIndian}_{ij}) + \beta_{6j}*(Mixed_{ij}) + \beta_{7j}*(PacificIslander13_{ij}) + \beta_{8j}*(Math\_z\text{-}score\_pretest13_{ij}) + r_{ij}$

Level-2 Model

$\beta_{0j} = \gamma_{00} + \gamma_{01}*(Treatment_j) + u_{0j}$

$\beta_{1j} = \gamma_{10}$

$\beta_{2j} = \gamma_{20}$

$\beta_{3j} = \gamma_{30}$

$\beta_{4j} = \gamma_{40}$

$\beta_{5j} = \gamma_{50}$

$\beta_{6j} = \gamma_{60}$

$\beta_{7j} = \gamma_{70}$

$\beta_{8j} = \gamma_{80}$

Table 2: Details

- Column A:
  - Enter the contrast ID number – This ID is used to link to the contrast IDs listed in Section 1, and to other tables in subsequent section of this document.
- Column B:
  - *Optional*: Enter the contrast name(s) corresponding to the contrast ID numbers entered in column A.
- Column C:
  - Enter the name of the post-test measure
- Columns D, E:
  - If applicable, enter the numbers of treatment and comparison clusters that contributed data to the analysis.
- Columns F, G:
  - If applicable, enter the numbers of treatment and comparison students that contributed data to the analysis.
- Columns H, I:
  - Enter the student-level standard deviations of the post-test measure, if available, otherwise, enter the standard deviation of cluster means, if those are the only data available.
    - For pre-post contrast (no comparison group), enter the post-test standard deviation in column H, enter the pre-test standard deviation in column I.
- Column J:
  - Enter a code (a footnote) for the source of the standard deviations reported in columns F and G
    - Codes for SD sources:
      - "A" = student-level SDs calculated from study data (full sample)
        - A.1 = student level SDs calculated from Study data (full sample), and z-scored prior to analysis (so that SD =1)
      - "B" = student-level SDs calculated from study data, specific to the subgroup analyzed in the current contrast
      - "C" = student-level SDs obtained from full, state population for relevant grade(s) and year(s)
      - "D" = student-level SDs obtained from published source (list source, e.g., test manual)
      - "E" = standard deviation of cluster means calculated from study data

        - "F", "G",… etc. = User defined codes. If you enter a user-defined code, you will need to provide a description of the code here that will explain what the code means.

- Column K:
  - *Optional*: Enter the control group mean of the post-test measure (unadjusted mean is optimal)
    - For pre-post contrast (no comparison group), enter the pre-test mean here.
- Column L:
  - Enter the impact estimate
- Column M:
  - *Optional*: Enter the standardized effect size (impact estimate divided by pooled standard deviation)
- Column N:
  - Enter the standard error of the impact estimate
- Column O:
  - Enter the two-tailed p-value for the test of whether the impact estimate is significantly different than zero. (enter exact p-value rather than "p < .05", etc)
- Column P:
  - Enter a code (a footnote) to describe the impact model used for this contrast. Examples are shown below
    - "A" = The model used to estimate this impact is shown in Section X.X of the design summary that is on file with the AR team (Document name = " xxx", document date = "mm/dd/yy")
    - "B" = The model used to estimate this impact is shown below
      - Describe here
    - "D", "E", …
      - Describe here
- Column Q:
  - Enter the degrees of freedom from the model used to calculate the impact estimate.
- Column R [more for TA team/evaluator but not part of Survey]:
  - *Optional*: Enter document name & page #/printout that findings are from
- Column S [more for TA team/evaluator but not part of Survey]:
  - *Optional*: Indicate whether data are for a student-level or a cluster-level inference

Table 3: Base Sample Sizes (Randomized Designs Only)
Base sample size tables are only relevant for randomized designs. For designs with

- Randomization of clusters and joiners included
  - Complete only table 3A.
- Randomization of clusters and no joiners included
  - Complete table 3A and 3B
- Randomization of students
  - Complete table 3B only.

Table 3.A: Base Sample Sizes for Studies with Randomization of Clusters

Table 3.A: Base Sample Sizes of Clusters (Schools)

		Contrasts (all contrasts from cluster randomized designs listed in Contrast List in Table 1 -- (add more columns as needed)							
		Contrast ID		Contrast ID		Enter ID		Enter ID	
		Contrast Name		Contrast Name		Contrast Name		Contrast Name	
Row #		T	C	T	C	T	C	T	T
1	# of schools Randomized	4	4	4	4	2	2	2	2
2	# of schools in Impact Analysis	4	4	4	4	2	2	2	2
3	Difference (Row1-Row2) [if difference = 0, skip to row 8 and enter "0" again. No need to indicate Reason for Loss]	0	0	0	0	0	0	0	0
4	Joiners included (Yes or No)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Reason for Loss (Evaluator lists reason for loss – add as many rows as needed[a]									
5									
6									
7									
8	Total Loss (Sum of "reasons for loss" rows - This number should equal "Difference" shown in Row 3)	0	0	0	0	0	0	0	0

[a] Most Reasons for Loss will be considered endogenous; if evaluator considers a reason for loss to be exogenous, enter sufficient information to explain why

Table 3.B: Base Sample Sizes of Students for Studies a) with Randomization of Students or b) with Randomization of Clusters and No Joiners

Table 3.B: Base Sample Sizes of Students

		Contrasts (all contrasts from cluster randomized designs listed in Contrast List in Table 1 -- (add more columns as needed)							
		Contrast ID		Contrast ID		Enter ID		Enter ID	
		Contrast Name		Contrast Name		Contrast Name		Contrast Name	
Row #		T	C	T	C	T	C	T	T
1	# of students randomized[a]								
2	# of students in impact analysis[a]								
3	Difference (Row1-Row2) [if difference = 0, skip to row 8 and enter "0" again. No need to indicate Reason for Loss]								
4	Joiners included (Yes or No)	No	No	No	No	No	No	No	No
Reason for Loss (Evaluator lists reason for loss – add as many rows as needed[b]									
5									
6									
7									
8	Total Loss (Sum of "reasons for loss" rows - This number should equal "Difference" shown in Row 3)								

[a] In studies with randomization of clusters and no joiners, the numbers in this table should reflect only the counts of students in non-attrited clusters.

[b] Most Reasons for Loss will be considered endogenous; if evaluator considers a reason for loss to be exogenous, enter sufficient information to explain why

Table 4: Baseline Equivalence of Treatment and Control Groups

Use Table 4.A for reporting data that will be used to assess the baseline equivalence of students. In these studies the analysis sample is defined as including only students with valid (non-missing) pre-test and post-test measures. Each student that contributes a post-test measure to the impact analysis also contributes a pre-test (baseline) measure to the baseline balance assessment. Use Table 4.B for reporting data that will be used to assess the baseline equivalence of clusters. In these studies, the students that are used to calculated pre-test (baseline) means are not the exact same students as those used to calculate post-test means.

Reviewers will use the data reported in these tables to calculate standardized differences between treatment and control groups at baseline.

Table 4.A: Baseline Equivalence of Students

Table 4.A: Baseline Equivalence of Students

A	B	C	D	E	F	G	H	I	J	K	L	M	N
Contrast ID #	Contrast Name (Optional)	Pre-test Measure Name	Treatment Group N	Comparison Group N	Unadjusted Treatment Group SD	Unadjusted Comparison Group SD	Standard Deviation Source (Code)	Unadjusted Comparison Group Mean (Optional)	Treatment – Comparison Difference	Standardized T-C Difference (Optional)	Pre-test shown in this row was used as a control in the impact model for this contrast ? (Y/N)	Code for T-C Difference Calculation	Source of Data (Optional)
	ELA Middle	CRCT	3839	4020	.994	1.003	.9837	-.042	.0487	.0495	Y	A	
	Math Middle	CRCT	3811	3992	.9938	1.001	.9975	-.068	.1022	.1025	Y	A	

A. The model used to estimate the T-C difference is shown below:

Level-1 Model

$$Y_{ij} = \beta_{0j} + r_{ij}$$

Level-2 Model

$$\beta_{0j} = \gamma_{00} + \gamma_{01}*(Treatment_j) + u_{0j}$$

where

Y_{ij} = is the baseline, pre-test measure of the i^{th} student in the j^{th} school.

Treatment = 1in treatment group, =0 if control.

r_{ij}= random effect representing the difference between student ij's score and the predicted mean score for school j, r_{ij}~ ND (0, σ_2).

$u_{0j=}$ the random error component for the deviation of the intercept of a school from the overall intercept.

Table 4.A: Details

- Column A:
  - Enter the contrast ID number
- Column B:
  - *Optional*: Enter the contrast name(s) corresponding to the contrast ID numbers entered in column A.
- Column C:
  - Enter the name of the pre-test measure
- Columns D, E:
  - Enter the numbers of treatment and comparison students that contributed data to the analysis.
- Columns F, G:
  - Enter the student-level standard deviations of the pre-test measure.
- Column H:
  - Enter the code for the source of the standard deviations reported in columns F and G
    - Codes for SD sources:
    - "A" = SDs calculated from study data (full sample)
      - A.1 = student level SDs calculated from Study data (full sample), and z-scored prior to analysis (so that SD =1)
    - "B" = SDs calculated from study data, specific to the subgroup analyzed in the current contrast
    - "C" = SDs obtained from full, state population for relevant grade(s) and year(s)
    - "D" = SDs obtained from published source (list source, e.g., test manual)
    - "E", "F", "G",… etc. = User defined codes. If you enter a user-defined code, you will need to provide a footnote to the table that will explain what the codes mean.
- Column I:
  - *Optional:* Enter the unadjusted comparison group mean of the pre-test measure.
    - If the impact analysis for this contrast weighted data in the analysis (e.g., sampling weights), present the weighted Comparison group mean here.
- Column J:
  - Enter the difference between treatment and Comparison group pre-test means.
    - See guidance on *"difference between treatment and Comparison group means."*
- Column K:
  - *Optional*: Enter the standardized T-C difference in the pre-test measure, where the standardization is calculated by dividing the T- C difference in Column J by the pooled standard deviation of T and C groups. The AR team will also calculate this standardized difference to assess baseline equivalence.

- Column L:
  - Indicate whether or not this pre-test measure shown in this row was used as a statistical control variable in the analytic model used to estimate the impact for this contrast (Yes/No).
- Column M:
  - Enter a code (a footnote) to describe how the T-C difference shown in Column J was calculated. Example codes follow
    - "A" = The method used to calculated the T-C difference shown in column J was described in Section X.X of the design summary that is on file with the AR team (Document name = " xxx", document date = "mm/dd/yy")
    - "B" = The T-C difference shown in column J was calculated as simple difference of unadjusted means, as described in "Method 1" of "i3 findings Reporting Shells_09222014.docx?
    - "C" = The T-C difference shown in column J was calculated by applying a modified form the impact model for this contrast, where the modified model had the pre-test as the dependent variable and did not include other covariate controls , as described in "Method 2" of "i3 findings Reporting Shells_09222014.docx?
    - ""D", "E", …
      - Describe here
- Column N: [more for TA team/evaluator but not part of Survey]:
  - *Optional*: Enter document name & page #/printout that findings are from

Guidance on *difference between treatment and Comparison group means.*

Either of the following ways of calculating the difference between treatment and Comparison group pre-tests means is acceptable.

Method 1: Calculate means in treatment and Comparison groups, and calculate the difference as the treatment group mean minus the Comparison group mean:

Let

$\overline{Y}_{pre.Treatment}$ be the mean of the pre-test score calculated using treatment group data.

$\overline{Y}_{pre.Comparison}$ be the mean of the pre-test score calculated using comparison group data.

Then,

Treatment –Comparison Difference $= \overline{Y}_{pre.Treatment} - \overline{Y}_{pre.Comparison}$

Method 2: Calculate the treatment minus Comparison group difference using a statistical model that has the same structural components as the statistical model that is used to to estimate impacts on the outcome variable for the same contrast. Method 2 may be appropriate when the impact model has a multi-level structure, or includes terms for randomization blocks or matching blocks, and where these structural features of the impact model have the effect of giving differential weight to the outcome data. In these

cases, in order to reflect the differential weighting of outcome data imposed by the impact model, the TA team recommends fitting the structural components of impact model to the pre-test measure to estimate a treatment minus comparison difference in the pre-test measure that is weighted in the same manner that the outcome data are weighted in the impact analysis. An example follows.

We explain this recommended approach and a recommended approach for reporting results of the baseline balance testing via the following example. Suppose the model that will be used to estimate impact for a particular contrast has the two-level structure of students nested in schools, as shown below:

Level 1:

$$Y_{ij} = \beta_{0j} + \beta_{1j}(Y_{pre.ij}) + \sum_{k=1}^{K} \beta_{k+1}(x_{ijk}) + \varepsilon_{ij}$$

Level 2:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(Trt) + \mu_{0j}$$
$$\beta_{1j} = \gamma_{10}$$
...
$$\beta_{1+k,j} = \gamma_{1+k,0}$$

where

Y_{ij} = an outcome measure of the i^{th} student in the j^{th} school.

$Y_{pre.ij}$ = is the baseline, pre-test measure of the i^{th} student in the j^{th} school.

Trt = 1in treatment group, =0 if Comparison.

x_k = represent k=1…K student-level covariates (e.g., demographics) included in the model.

ε_{ij} = represents residual error. Assumed to be distributed normal with mean 0 and variance σ^2.

μ_{0j} = is a random intercept term for the 4th school. Assumed to be distributed normal with mean 0 and variance τ^2, and independent of ε_{ij}.

Fitting the above model to the data would produce an estimate $\hat{\gamma}_{01}$, that represents the estimated impact of treatment.

We recommend the model for testing for baseline equivalence have the form:

Level 1:

$$Y_{pre.ij} = \beta_{0j} + \varepsilon_{ij}$$

Level 2:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(Trt) + \mu_{0j}$$

where the terms are the same as those described previously for the impact model. This model has the same structural components as the impact model (in this case random intercepts for schools), but does not include the other covariate x's. Fitting the above model to the data would produce an estimate $\hat{\gamma}_{01}$, which represents the estimated treatment-Comparison difference in the pre-test measure. This estimate ($\hat{\gamma}_{01}$) could be entered into Column I of the table shown above.

Table 4.B: Baseline Equivalence of Clusters

Table 4.B: Baseline Equivalence of Clusters

A	B	C	D	E	F	G	H	I	J	K	L	M	N
Contrast ID #	Contrast Name (Optional)	Pre-test Measure Name	Treatment Group N of Clusters	Comparison Group N of Clusters	Unadjusted Treatment Group SD	Unadjusted Comparison Group SD	Standard Deviation Source (Code)	Unadjusted Comparison Group Mean (Optional)	Treatment – Comparison Difference	Standardized T-C Difference (Optional)	Pre-test shown in this row was used as a control in the impact model for this contrast? (Y/N)	Code for T-C Difference Calculation	Source of Data (Optional)

Table 4.B: Details

- Column A:
  - Enter the contrast ID number
- Column B:
  - *Optional*: Enter the contrast name(s) corresponding to the contrast ID numbers entered in column A.
- Column C:
  - Enter the name of the pre-test measure
- Columns D, E:
  - Enter the numbers of treatment and comparison cluster that contributed data to the analysis.
- Columns F, G:
  - Enter the student-level standard deviations of the pre-test measure, if available, otherwise, enter the standard deviation of cluster means, if those are the only data available.
- Column H:
  - Enter the code for the source of the standard deviations reported in columns F and G
    - Codes for SD sources:
      - "A" = student-level SDs calculated from study data (full sample)
        - A.1 = student level SDs calculated from Study data (full sample), and z-scored prior to analysis (so that SD =1)
      - "B" = student-level SDs calculated from study data, specific to the subgroup analyzed in the current contrast
      - "C" = student-level SDs obtained from full, state population for relevant grade(s) and year(s)
      - "D" = student-level SDs obtained from published source (list source, e.g., test manual)
      - "E" = standard deviation of cluster means calculated from study data
      - "F", "G",… etc. = User defined codes. If you enter a user-defined code, you will need to provide a footnote to the table that will explain what the codes mean.
- Column I:
  - *Optional*: Enter the Comparison group mean of the pre-test measure.
    - If the impact analysis for this contrast weighted data in the analysis (e.g., sampling weights), present the weighted Comparison group mean here.
- Column J:
  - Enter the difference between treatment and Comparison group pre-test means.

        - See guidance on *"difference between treatment and Comparison group means"* (in the previous section).

- Column K:
  - *Optional*: Enter the standardized T-C difference in the pre-test measure, where the standardization is calculated by dividing the T- C difference in Column J by the pooled standard deviation of T and C groups. The AR team will also calculate this standardized difference to assess baseline equivalence.
- Column L:
  - Indicate whether or not this pre-test measure shown in this row was used as a statistical control variable in the analytic model used to estimate the impact for this contrast (Yes/No).
- Column M:
  - Enter a code (a footnote) to describe how the T-C difference shown in Column J was calculated. Example codes follow
    - "A" = The method used to calculated the T-C difference shown in column J was described in Section X.X of the design summary that is on file with the AR team (Document name = " xxx", document date = "mm/dd/yy")
    - "B" = The T-C difference shown in column J was calculated as simple difference of unadjusted means, as described in "Method 1" of "i3 findings Reporting Shells_09222014.docx?
    - "C" = The T-C difference shown in column J was calculated by applying a modified form the impact model for this contrast, where the modified model had the pre-test as the dependent variable and did not include other covariate controls , as described in "Method 2" of "i3 findings Reporting Shells_09222014.docx?
    - ""D", "E", …
      - Describe here
- Column N: [more for TA team/evaluator but not part of Survey]:
  - *Optional:* Enter document name & page #/printout that findings are from

Table 5: Fidelity of Implementation of Intervention
Table 5.1: Key Components of Intervention [All Grant Types]

Table 5.1: Description of Key Components

Planned Intervention Activity: All *key components* measured across years of implementation	List of Key Indicators For Each Key Component
Integrated learning platform and data management system LMS.	1) 75% of key project personnel satisfied with LMS functionality. ***Key component is Recommendation Engine* (1=yes, 0= no)
	2) More than 60% of ELA and math teachers satisfied with their access to the LMS (1=yes, 0= no)
	3) More than 60% of ELA and math teachers satisfied with the LMS's access to data (1=yes, 0= no)

Planned Intervention Activity: All *key components* measured across years of implementation	List of Key Indicators For Each Key Component
	4) More than 60% of ELA and math teachers satisfied with variety of learning resource repository items that are aligned to standards (1=yes, 0= no)
Develop Professional Learning Support Teams: Train the Trainer	1.) Appropriate number trained as trainers for each school. (1=yes, 0= no)
	2.) Training addresses topics needed to use LMS system (3 most training sessions, 2 some training sessions, 1 few training sessions)
	3.) Training resources/ materials appropriately support learning how to use the LMS -- usefulness, quality, clarity-- (3 most resources, etc., 2 some resources, etc., 1 few resources)
	4.) Lengths of training sessions are appropriate for content. (3 most, 2 some, 1 few)
	5.) At least 75% of Support Team Personnel received training on a timeline that allowed time for them to develop skill using LMS in order to assist other teachers. (1=yes, 0= no)
	6.) Support Team Personnel have the knowledge they need to begin assisting other teachers. (3=75% to 100%, 2=50% to 74%, 1=25% to 49%, 0=less than 25%)
Provide Teacher Professional Development	1.) ELA and math teachers and SpEd inclusion teachers in the school receive training. (3 =75% to 100% of subject-level teachers per school level, 2 =50% to 74%, 1=less than 50% of subject-level teachers)
	2.) Training materials and delivery addresses topics needed to use LMS system (3 most, 2 some, 1 few)
	3.) Training resources/ materials appropriately support learning how to use the LMS (usefulness, quality, clarity) (3 most, 2 some, 1 few)
	4.) Teachers agreement that training addresses topics needed to use LMS system (3 =80% to 100% of subject-level teachers per school level, 2 =60% to 79%, 1=40% to 59%, 0= less than 40%)
	5.) Teachers agreement that training resources/ materials appropriately support learning how to use the LMS (usefulness, quality, clarity) (3 =80% to 100% of subject-level teachers per school level, 2 =60% to 79%, 1=40% to 59%, 0= less than 40%)
	6.) Teachers agreement that training sessions provided at appropriate times in order for teachers to facilitate using LMS. (3 =80% to 100%, 2 =60% to 79%, 1=40% to 59%, 0= less than 40%)
	7.) Teacher agreement that lengths of training sessions are appropriate for content. (3 =80% to 100%, 2 =60% to 79%, 1=40% to 59%, 0= less than 40%)
	8.) Teachers are satisfied with support and training for LMS (3 =80% to 100%, 2 =60% to 79%, 1=40% to 59%, 0= less than 40%)
School Level Support	1.) School has an assigned technology specialist trained on LMS whose workload and schedule allows time to support teachers in using LMS.(3=high, 2=medium, 1=low)
	2.) School has at least one ELA teacher and one math teacher who serve as peer leaders in using LMS. The teachers' workloads and schedules allow time to support teachers in using LMS. (2=both a math and ELA teacher, 1= Peer leader for only 1 subject, 0= no peer leaders in place)
	3.) Teacher agreement on survey items that administration is providing the support needed for teachers to personalize instruction using the LMS. (3=high 75% to 100%); 2=medium 50% to 74%); 1=low less than 50%);
	4.) School has put into practice policies and procedures that promote all students having access to resources even when a student's access to technology resources from home is limited. (3=high, 2=medium, 1=low)
	5.) Students identified as at risk have individual learning plans. (3=high 75% to 100% of at risk students have plan); 2=medium 50% to 74%); 1=low less than 50%)

Planned Intervention Activity: All *key components* measured across years of implementation	List of Key Indicators For Each Key Component
	6.) School provides opportunities for parents to learn to use LMS (2=multiple opportunities, 1=only one opportunity, 0=no instruction for parents in how to use LMS)
	7.) Students know how to use the LMS (3=high 75% to 100%); 2=medium 50% to 74%); 1=low less than 50%);
District Level Support	1.) FCS personnel assigned and trained to support and maintain LMS related to technology, data, and reporting. (3-high, 2-medium, 1-low)
	2.) i3 Team provides training and on-going support for trainers in the schools. (3-high, 2-medium, 1-low)
	3.) Administrators and teachers satisfaction with level of support provided by district for LMS. (3-high, 2-medium, 1-low)
	4.) Student agreement that tools and technologies available to them are adequate to support learning. (3=high 75% to 100%); 2=medium 50% to 74%); 1=low less than 50%)
	5.) Administrators and teachers satisfied with adopted method for identifying student learning preferences. (3=high 75% to 100%); 2=medium 50% to 74%); 1=low less than 50%)

Table 5.2: Fidelity of Implementation of Intervention(s) by Year [All Grant Types]

Findings from Evaluator Study of Implementation: IMPLEMENTATION YEAR 1

Enter calendar year: _____ (e.g., 2010-11; Sept. 2011-June 2012; Summer 2012)

Intervention Components: Copy from list above	Implementation measure (total number of measurable indicators representing each component	Sample Size at the Sample Level (# of schools, districts, etc)	Representativeness of sample: Measured on All (A), Some (S), or None (N) of the units representing the intervention group in the impact analyses[b]	Component Level Threshold for Fidelity of Implementation for the Unit that is the Basis for the Sample-Level	Evaluator's Criteria for "Implemented with Fidelity" at Sample Level	Component Level Fidelity Score for the Entire Sample	Implemented with Fidelity? (Yes, No, N/A)
Planned Intervention Activities [i.e., key components]							
Mediators[a] [if choosing to report findings for fidelity of Intervention - OPTIONAL]							

[a] Mediators refer to measurement of short-term outcomes, e.g., changes in student/teacher/parent behaviors or attitudes that are assumed to have a direct connection to long-term outcomes. If these mediators have been measured in both the treatment and comparison groups **and** are being reported as part of an exploratory contrast, they should not be entered here. If these mediators have been measured in the treatment group only or in both the treatment and the comparison group **and** are not being reported as part of an exploratory contrast, the findings for just the treatment group would be reported here. If evaluators believe they have findings on mediators to report, they should consult with their TA Liaison or the AR helpdesk to confirm that the findings are appropriate for this section of the Survey.

[b] *All*: If the intervention group in the impact analysis includes 10 schools and fidelity measurement includes these 10 schools, the evaluator would enter "A" indicating that All of the schools in the impact analysis are represented in the fidelity findings. *Some*: If the intervention group in the impact analysis includes teachers in grades K to 3 but fidelity is measured only for teachers in Kindergarten, the evaluator would enter "S" indicating that Some of the teachers in the impact analysis are represented in the fidelity findings. *None*: If the intervention group in the impact analysis includes grades 7 - 9 but fidelity is measured only for grades 5-6, the evaluator would enter "N" indicating that None of the grades in the impact analysis are represented in the fidelity findings.

Findings from Evaluator Study of Implementation: IMPLEMENTATION YEAR 2

Enter calendar year: _____ (e.g., 2010-11; Sept. 2011-June 2012; Summer 2012)

Intervention Components: Copy from list above	Implementation measure (total number of measurable indicators representing each component	Sample Size at the Sample Level (# of schools, districts, etc)	Representativeness of sample: Measured on All (A), Some (S), or None (N) of the units representing the intervention group in the impact analyses[b]	Component Level Threshold for Fidelity of Implementation for the Unit that is the Basis for the Sample-Level	Evaluator's Criteria for "Implemented with Fidelity" at Sample Level	Component Level Fidelity Score for the Entire Sample	Implemented with Fidelity? (Yes, No, N/A)
Planned Intervention Activities [i.e., key components]							
Mediators[a] [if choosing to report findings for fidelity of Intervention - OPTIONAL]							

[a] Mediators refer to measurement of short-term outcomes, e.g., changes in student/teacher/parent behaviors or attitudes that are assumed to have a direct connection to long-term outcomes. If these mediators have been measured in both the treatment and comparison groups **and** are being reported as part of an exploratory contrast, they should not be entered here. If these mediators have been measured in the treatment group only or in both the treatment and the comparison group **and** are not being reported as part of an exploratory contrast, the findings for just the treatment group would be reported here. If evaluators believe they have findings on mediators to report, they should consult with their TA Liaison or the AR helpdesk to confirm that the findings are appropriate for this section of the Survey.

[b] *All*: If the intervention group in the impact analysis includes 10 schools and fidelity measurement includes these 10 schools, the evaluator would enter "A" indicating that All of the schools in the impact analysis are represented in the fidelity findings. *Some*: If the intervention group in the impact analysis includes teachers in grades K to 3 but fidelity is measured only for teachers in Kindergarten, the evaluator would enter "S" indicating that Some of the teachers in the impact analysis are represented in the fidelity findings. *None*: If the intervention group in the impact analysis includes grades 7 - 9 but fidelity is measured only for grades 5-6, the evaluator would enter "N" indicating that None of the grades in the impact analysis are represented in the fidelity findings.

Findings from Evaluator Study of Implementation: IMPLEMENTATION YEAR 3

Enter calendar year: _____ (e.g., 2010-11; Sept. 2011-June 2012; Summer 2012)

Intervention Components: Copy from list above	Implementation measure (total number of measurable indicators representing each component	Sample Size at the Sample Level (# of schools, districts, etc)	Representativeness of sample: Measured on All (A), Some (S), or None (N) of the units representing the intervention group in the impact analyses[b]	Component Level Threshold for Fidelity of Implementation for the Unit that is the Basis for the Sample-Level	Evaluator's Criteria for "Implemented with Fidelity" at Sample Level	Component Level Fidelity Score for the Entire Sample	Implemented with Fidelity? (Yes, No, N/A)
Planned Intervention Activities [i.e., key components]							
Mediators[a] [if choosing to report findings for fidelity of Intervention - OPTIONAL]							

[a] Mediators refer to measurement of short-term outcomes, e.g., changes in student/teacher/parent behaviors or attitudes that are assumed to have a direct connection to long-term outcomes. If these mediators have been measured in both the treatment and comparison groups **and** are being reported as part of an exploratory contrast, they should not be entered here. If these mediators have been measured in the treatment group only or in both the treatment and the comparison group **and** are not being reported as part of an exploratory contrast, the findings for just the treatment group would be reported here. If evaluators believe they have findings on mediators to report, they should consult with their TA Liaison or the AR helpdesk to confirm that the findings are appropriate for this section of the Survey.

[b] *All*: If the intervention group in the impact analysis includes 10 schools and fidelity measurement includes these 10 schools, the evaluator would enter "A" indicating that All of the schools in the impact analysis are represented in the fidelity findings. *Some*: If the intervention group in the impact analysis includes teachers in grades K to 3 but fidelity is measured only for teachers in Kindergarten, the evaluator would enter "S" indicating that Some of the teachers in the impact analysis are represented in the fidelity findings. *None*: If the intervention group in the impact analysis includes grades 7 - 9 but fidelity is measured only for grades 5-6, the evaluator would enter "N" indicating that None of the grades in the impact analysis are represented in the fidelity findings.

Findings from Evaluator Study of Implementation: IMPLEMENTATION YEAR 4

Enter calendar year: __Summer 2013-June 2014____ (e.g., 2010-11; Sept. 2011-June 2012; Summer 2012)

Intervention Components: Copy from list above	Implementation measure (total number of measurable indicators representing each component	Sample Size at the Sample Level (# of schools, districts, etc)	Representativeness of sample: Measured on All (A), Some (S), or None (N) of the units representing the intervention group in the impact analyses[b]	Component Level Threshold for Fidelity of Implementation for the Unit that is the Basis for the Sample-Level	Evaluator's Criteria for "Implemented with Fidelity" at Sample Level	Component Level Fidelity Score for the Entire Sample	Implemented with Fidelity? (Yes, No, N/A)
Planned Intervention Activities [i.e., key components]							
Integrated learning platform and data management system LMS	4	6 schools	A	High implementing school = 4 (passes all 4 indicators)	100% of schools:	0	No
Develop Professional Learning Support Teams: Train the Trainer	6	6 schools	A	High implementing school has a score >9	67% of schools (4 of 6)	100% (6 schools)	Yes
Provide Teacher Professional Development	8	6 schools	A	High implementing school has a score >15	67% of schools (4 of 6)	100% (6 schools)	Yes
School Level Support	7	6 schools	A	High implementing school has a score >11	67% of schools (4 of 6)	100% (6 schools)	Yes
District Level Support	5	6 schools	A	High implementing school has a score >9	67% of schools (4 of 6)	100% (6 schools)	Yes

Findings from Evaluator Study of Implementation: IMPLEMENTATION YEAR 5

Enter calendar year: _Summer 2014-June 2105_____ (e.g., 2010-11; Sept. 2011-June 2012; Summer 2012)

Intervention Components: Copy from list above	Implementation measure (total number of measurable indicators representing each component	Sample Size at the Sample Level (# of schools, districts, etc)	Representativeness of sample: Measured on All (A), Some (S), or None (N) of the units representing the intervention group in the impact analyses[b]	Component Level Threshold for Fidelity of Implementation for the Unit that is the Basis for the Sample-Level	Evaluator's Criteria for "Implemented with Fidelity" at Sample Level	Component Level Fidelity Score for the Entire Sample	Implemented with Fidelity? (Yes, No, N/A)
Planned Intervention Activities [i.e., key components]							
Integrated learning platform and data management system LMS	4	*12 schools	A	High implementing school = 4 (passes all 4 indicators)	100% of schools:	0	No
Develop Professional Learning Support Teams: Train the Trainer	6	*12 schools	A	High implementing school has a score >9	67% of schools (8 of 12)	100% (12 schools)	Yes
Provide Teacher Professional Development	8	*12 schools	A	High implementing school has a score >15	67% of schools (8 of 12)	100% (12 schools)	Yes
School Level Support	7	*12 schools	A	High implementing school has a score >11	67% of schools (8 of 12)	100% (12 schools)	Yes
District Level Support	5	*12 schools	A	High implementing school has a score >9	67% of schools (8 of 12)	100% (12 schools)	Yes

*NOTE: 6 schools which were formally the control group became part of the treatment group in 2014-15. The sample size increased from 6 schools in 2013-14 to 12 schools in 2014-15.

Add similar tables for any additional years of implementation findings

Table 5.3: Key Components Supporting Scale-Up [Scale-Up Grants Only]

Key Activities/Components Related to Scale-Up
Example: Creation on regional teacher and principal training centers
Example: Further developed recruitment plan

Table 5.4: Scale-Up Goals [Scale-Up Grants Only]

Scale-up Goal(s) Note: These are quantified and measurable	Scale-up Results by Grant End	Goal Met by Grant End? (Yes, No)	Scale-up Results by Year [Optional]				
			Y1	Y2	Y3	Y4	Y5
Example: 500 additional teachers trained	*500*	*Yes*	*N/A*	*200*	*300*	*400*	*500*
Example: 175 Additional principals trained	*175*	*Yes*	*N/A*	*100*	*175*	*N/A*	*N/A*
Example: Contract with and implement program in 1,000 K-12 schools	*750*	*No*	*N/A*	*450*	*500*	*600*	*750*