Coordinator Name
Paul Bauer

Planning Team
Paul Bauer
Randy Barnes

Description of the Program (2012-2013)
One of the purposes of the Program Evaluation department is to provide quantitative analysis and interpretation of data to assist a variety of programs within SSD. Student outcomes have become increasingly important to SSD not only for continuous improvement of instruction, but for accreditation as well. This effort to link student outcomes to other predictive factors is essential to improve student learning. The results of this analysis will have implications for policy and programs.

Description of How the Program's Services are Developed and Delivered
The Program Evaluation department has filed disaggregated data reports detailing Missouri Assessment Program (MAP) results by individual groups for several years. In the past only mean scores for groups were reported, however, there is a need for deeper analysis using quantitative techniques which can guide program development and instructional intervention.   In order to provide the SSD Board of Education with timely reporting of Missouri School Improvement Program (MSIP) results and have time to provide deeper analysis, the program evaluation cycle for disaggregation of data was divided into two parts. The first report was approved by the Board of Education in Fall 2013 and provided disaggregation of MAP data by groups. The second report would be a deeper analysis provided later in the year. This report is the first to provide that deeper analysis as an individual program evaluation report. This analysis includes only students taking Grade Level Assessments or End of Course Exams and excludes scores from the Missouri Assessment Program-Alternate (MAP-A) This is because the MAP-A is a different types of assessment and including its results would invalidate the analysis. When possible, CSIP goal measures were incorporated to align the program evaluation to SSD continuous improvement.

Key Program Stakeholder Groups
☒ Students
☐ Parents
☒ Staff
☒ Administrators

☐ Board of Education
☐ Taxpayers
☐ Other (Specify.)

Student and/or Stakeholder Needs Addressed by the Program
This report has implications for interventions to improve student learning.

Overall Goals of the Program	Expected Measurable Outcomes
Goal 1: Using linear regression*, evaluate correlation of STAR assessments as a predictive measure for MAP assessments in support of CSIP strategy 1.1.3 "Develop and review predictive measures for require state assessments in all academic content areas."	1.1 Correlation results of STAR Math to MAP Mathematics assessment. 1.2 Correlation results of STAR Reading to MAP English Language Arts assessment.
Goal 2: Eliminate achievement gaps between groups based on Gender, Ethnicity and Socio-Economic status.	2.1 Gender will not be a significant predictor of MAP proficiency as determined by logistic regression*. 2.2 Ethnicity will not be a significant predictor of MAP proficiency as determined by logistic regression. 2.3 Socio-economic status will not be a significant predictor of MAP proficiency as determined by logistic regression.
Goal 3: Cohort of SSD students taking MAP from 2011 to 2013 will demonstrate growth.	3.1 2011 – 2013 cohort will show growth on MAP assessment of English Language Arts from 2011 to 2013. 3.2 2011 – 2013 cohort will show growth on MAP assessment of Mathematics from 2011 to 2013.

Evaluation Questions
- What is the status of the program's progress toward achieving the goals?
- What do students and other stakeholders consider to be the strengths and weaknesses of the program?
- What do staff consider to be the strengths and weaknesses of the program?
- How does the program's actual implementation compare with the program's design?
- How should priorities be changed to put more focus on achieving the goals?
- How should goals be changed? Any added or removed?

Data Collection Methods
☐ Surveys and questionnaires
☐ Interviews
☐ Document reviews
☐ Observations
☐ Focus groups
☐ Case studies
☒ Assessments
☒ Other Quantitative Analysis

*Operational Definitions
Linear Regression - Linear regression uses the values from an existing data set consisting of measurements of the values of two variables, X and Y, to develop a model that is useful for predicting the value of the dependent variable, Y for given values of X. For example, predicting that students who like to read (Variable X) perform better on a test (Variable Y).
Logistic Regression – For this report logistic regression is used to refer to the problem in which the dependent variable is binary, for example, in the case that a student is either proficient or not proficient. For example, predicting that girls are 1.3 times more likely than boys to score "Proficient" on a test.
MAP – Missouri Assessment Program.
MSIP – Missouri School Improvement Plan. Each school district receives an MSIP report each year to determine accreditation status.

Disaggregated Data Analysis
2013-2014 Program Evaluation Report   Special School District

Evaluation Results

What is the status of the program's progress toward achieving the goals?

Goal 1: Using linear regression*, evaluate correlation of STAR assessments as a predictive measure for MAP assessments in support of CSIP strategy 1.1.3 "Develop and review predictive measures for required state assessments in all academic content areas."

Measurable Objective 1:	1.1 Correlation results of STAR Math to MAP Mathematics assessment.

Results: Met. The STAR Math scale score is a significant predictor of MAP Math scale score.

Pearson's correlation coefficient R. A Pearson's correlation coefficient of .7 is considered very strong. The Pearson's coefficient for correlation of STAR Math to MAP Math was .745 for 2012 and .883 for 2013.

Adjusted R Square. The adjusted R Square for 2012 was .549 indicating that 54.9% of the variance in MAP scores is explained by STAR scores. For 2013 77.5% of the variance is explained by the STAR math score. This statistic can be interpreted as indicating that the skills the STAR assessment measures account for 77.5% of the skills the MAP scores measure.

The correlation for each year is statistically significant at p<.0005.

Regression of STAR Reading Scale Score to ELA MAP Scale score.		
	Pearson's R	Adjusted R Square
2012	0.745	.549
2013	0.883	0.775

Math 2012						
Model Summary[b]						
Model	(Pearson's) R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson	
1	.745[a]	.555	.549	28.145	2.177	
ANOVA[a]						
Model		Sum of Squares	df	Mean Square	F	Sig. (p value)
1	Regression	73254.406	1	73254.406	92.477	.000[b]

Math 2013						
Model Summary[b]						
Model	(Pearson's) R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson	
1	.883[a]	.779	.775	23.943	2.404	
ANOVA[a]						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	95180.492	1	95180.492	166.032	.000[b]

Measurable Objective 2:	1.2 Correlation results of STAR Reading to MAP English Language Arts assessment.

Results: Met. The STAR Reading Scale score is a significant predictor of MAP ELA Scale score. A Pearson's correlation coefficient of .7 is considered very strong. The Pearson's coefficient for correlation of STAR Reading Scale Score to MAP ELA Scale Score was over .8 for both 2012 and 2013.

The adjusted R square indicates the percent of variance accounted for by the independent variable. For both years the percent of variance accounted for by the STAR scale score was over 65%. This statistic can be interpreted as indicating that the skills that the STAR score measures account for 65% of the skills the MAP scores measure. It points to the interpretation that reading ability itself accounts for more than half of the English Language Arts MAP score.

The correlation for each year is statistically significant at $p<.0005$.

Regression of STAR Reading Scale Score to ELA MAP Scale score.		
	Pearson's R	Adjusted R Square
2012	0.809	.649
2013	0.862	0.738

English Language Arts 2012						
Model Summary[b]						
Model	(Pearson's) R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson	
1	.809[a]	.654	.649	21.013	2.004	
ANOVA[a]						
Model		Sum of Squares	df	Mean Square	F	Sig (p value).
1	Regression	59334.474	1	59334.474	134.375	.000[b]

English Language Arts 2013						
Model Summary						
Model	(Pearson's) R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson	
1	.862a	0.744	0.738	18.515	1.705	
ANOVA[a]						
Model		Sum of Squares	df	Mean Square	F	Sig (p value).
1	Regression	47730.216	1	47730.216	139.230	.000[b]

Goal 2: Eliminate achievement gaps between groups based on Gender, Ethnicity and Socio-Economic status.

Measurable Objective 1:	2.1 Gender will not be a significant predictor of MAP proficiency as determined by logistic regression*.
Results: Met: Gender was not found to be a significant predictor of MAP proficiency.	

Measurable Objective 2:	2.2 Ethnicity will not be a significant predictor of MAP proficiency as determined by logistic regression.
Results: Not Met: Ethnicity was found to be a significant predictor of MAP proficiency at p<.05. The odds ratio could be expressed as "Asian/White students are 2.1 times as likely to be proficient as non-Asian/Whites. " **	

Measurable Objective 3:	2.3 Socio-economic status will not be a significant predictor of MAP proficiency as determined by logistic regression.
Results: Met: Socio-economic status was not found to be a significant predictor of MAP proficiency.	

Variables in the Equation									
						Sig.	Exp(B)(like R square	95% C.I.for EXP(B)	
		B	S.E.	Wald	df	(p value)		Lower	Upper
Step 1[a]	GenderRecode(1)	-.415	.275	2.274	1	.132	.661	.385	1.132
	FreeReduced(1)	.037	.308	.014	1	.905	1.037	.567	1.899
	Ethnicity(1)	.743	.351	4.491	1	.034	2.103	1.057	4.183
	Constant	-.726	.342	4.511	1	.034	.484		
a. Variable(s) entered on step 1: GenderRecode, FreeReduced, Ethnicity.									
Of the three independent variables tested on Proficiency, only Ethnicity was significant. The results indicate that Asian/White students are 2.1 times as likely to be proficient as non Asian/Whites.									

Goal 3: Cohort of SSD students taking MAP from 2011 to 2013 will demonstrate growth.

Measurable Objective 1:	3.1 2011 – 2013 cohort will show growth on MAP assessment of English Language Arts from 2011 to 2013.

Results: Met: 42 students had English Language Arts MAP scores in each of the three years. Friedman's Test of repeated measures was used to determine if test scores were significantly higher across the three year period. The results indicate that these 42 students scored significantly higher on 2012 and 2013 MAP test than their 2011 scores. However, the comparison of 2012 to 2013 scores was not significant. The average score for 2013 was lower than for 2012.

Table 1 Results of Friedman's test for ELA MAP scores from 2011 to 2013 indicate that there is a significant difference between scores.

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distributions of CAMAP11, CAMAP12 and CAMAP13 are the same.	Related-Samples Friedman's Two-Way Analysis of Variance by Ranks	.000	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Table 2 Comparison between years shows that there is significant difference from 2011 to 2012 and from 2011 to 2013, but not from 2012 to 2013

Comparison	Adjusted Significance
2011 to 2012	0.031
2011 to 2013	<0.005
2012 to 2013	0.341

Measurable Objective 2:	3.2 2011 – 2013 cohort will show growth on MAP assessment of Mathematics from 2011 to 2013.

Results: Met: 40 students had Math MAP scores in each year from 2011 to 2013. Friedman's test shows significant difference from 2011 to 2012 and from 2011 to 2013, but not from 2012 to 2013.

Table 3 Results of Friedman's test for Math scores shows there is a significant difference between scores over the period.

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distributions of MathMAP11, MathMAP12 and MathMAP13 are the same.	Related-Samples Friedman's Two-Way Analysis of Variance by Ranks	.000	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Table 4 Comparison between years shows significant growth from 2011 to 2012 and from 2011 to 2013, but not from 2012 to 2013.

Comparison	Adjusted Significance
2011 to 2012	0.005
2011 to 2013	<0.0005
2012 to 2013	0.438

What do key staff and stakeholders consider to be the strengths and opportunities for improvement /weaknesses of the program?

Strengths
- *STAR scale scores are highly correlated to MAP scale scores for both Reading and Math.*
- *Differences in scores between groups based on gender and socio-economic status are not significant.*
- *Cohort of students with tests in 2011 through 2013 showed significant growth from 2011 to 2012 and to 2013 for both Math and English Language Arts.*

Opportunities/Weaknesses
- *Differences in scores based on ethnicity were significantly different. Asian or white students are 2.1 times as likely as non-Asian or White students to score proficient or advanced*
- *Students in the cohort with tests from 2011 through 2013 did not show significant growth from 2012 to 2013 for both Math and English Language Arts.*

How well aligned are the program's priorities and processes with the goals of the program?

> The analysis of disaggregated data is well aligned with the goals of SSD.

Deployment Level of Program Services: The program services are in the early stages of deployment in most areas or schools.

Should priorities be changed to put more focus on achieving the goals? ☒ Yes ☐ No
If Yes describe change in priorities.

> More emphasis should be placed on data collection and timely analysis to provide insights for interventions and process improvement.

Should goals be changed, added or removed? ☐ Yes ☒ No

Evaluation Implications

General Recommendation Resulting from the Evaluation
Select from the following possible recommendations resulting from the evaluation:
☐ Continue the program as is. It is meeting or exceeding all expected outcomes.
☒ Expand the program, replicating effective components.
☐ Streamline, refine, or consolidate elements of the program.
☐ Redesign the program.
☐ Reevaluate the purpose and/or goals of the program.
☐ Discontinue ineffective or nonessential program components.
☐ Discontinue the program.
☐ Other (Specify.)

Action Plans

1. Analyze data regarding the achievement gap based on ethnicity to identify possible avenues of intervention.
2. Expand cohort analysis to determine if there are patterns to growth as indicated by MAP scores.

Cost and Funding Source The costs are included in the budget.

*Operational Definitions
Linear Regression - Linear regression uses the values from an existing data set consisting of measurements of the values of two variables, X and Y, to develop a model that is useful for predicting the value of the dependent variable, Y for given values of X. For example, predicting a student's final exam score based on the student's quiz scores during the semester.
Logistic Regression – For this report logistic regression is used to refer to the problem in which the dependent variable is binary, for example, in the case that a student is either proficient or not proficient.
MAP – Missouri Assessment Program.
MSIP – Missouri School Improvement Plan. Each school district receives an MSIP report each year to determine accreditation status.
p level refers to the probability that the results of the study happened by chance. Technically, p level is used to reject the null hypothesis that there is no effect or no relationship. Practically, the lower the p value is the more likely the difference or relationship is to be significant. The standard of $p<.05$ is accepted as worthy of reporting. Lower values for p, such as $p<.0005$ are significant at a much higher level. $p<.0005$ would mean that there is less than a 0.05% chance that the results were obtained by chance.

** SSD student population is primarily African American or White. Recent trends in research report Asian and White students together and all others in another category. The advantage is that all students can be reported clearly and simply.