

# Supplemental Educational Services Program Learning Gains Evaluation Plan

**Florida Gulf Coast University**  
College of Education



June 29, 2013

Nicholas S.Favazzo

**Contents**

Introduction ..... 3

Evaluation Process ..... 4

*Instruments* ..... 4

*Pretest-Posttest Comparison* ..... 4

*Matched Groups*..... 4

*Plan* ..... 5

Fidelity of Implementation ..... 6

Evaluation Results ..... 7

*Interpretation of evaluation findings* ..... 7

*Accountability and Transparency*..... 9

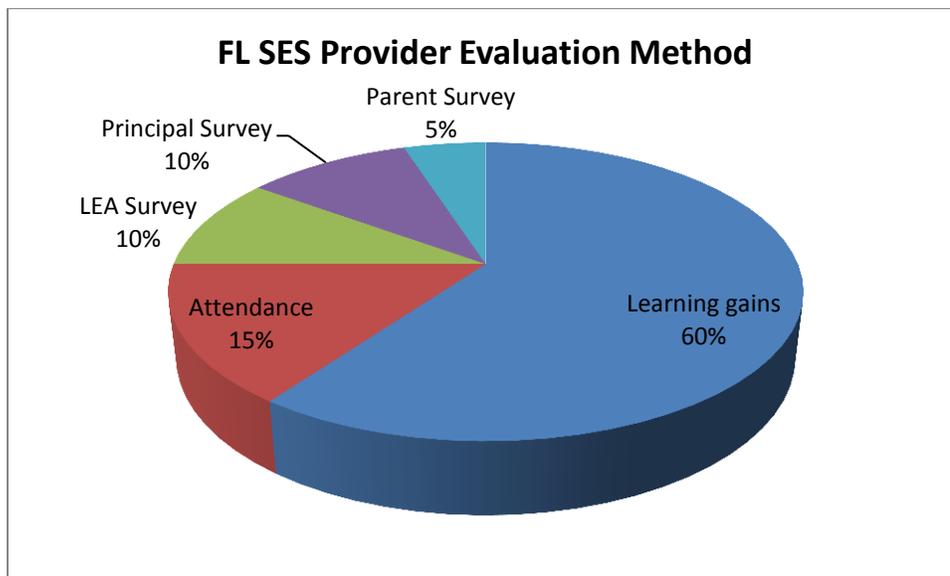
Conclusion..... 10

References ..... 11

## Introduction

This evaluation paper is being completed as a project for EDF 6461: *Introduction to Program Evaluation* at Florida Gulf Coast University in response to an acknowledged need. The SES Program subsidizes supplemental education, or tutoring, services to underperforming students that possess a suboptimal socioeconomic status. While logic exists for Program continuation, provider efficacy determination via rigorous evaluation is needed to properly attribute the source of student learning.

Evaluation of Supplemental Educational Services (SES) Program providers has been problematic. In 2010, the State Board of Education established Florida Administrative Code 6A-1.0391 in attempt to elucidate the evaluation process. A 500 point scale was established that uses learning gains and others factors, as indicated in the chart below. Despite this, numerous discrepancies continue to have effects on learning gains- which comprises the majority of the current evaluation. This thwarts accurate evaluation. Recognition of these disparities is necessary for true evaluation of SES Program providers' performance.



This evaluation paper establishes an alternative evaluation plan that is inclusive of procedures accounting for current discrepancies. In doing so, a more objective and accurate evaluation of learning gains will better answer the question, “Did the provider increase student achievement in reading/language arts or mathematics?”

## Evaluation Process

### *Instruments*

1. *Knowledge assessment instruments.* Uniform tools appropriate for the grade level and subject(s) to be evaluated (mathematics and language arts) can be used to assess students' initial knowledge and subsequent change.
2. *Demographic data.* An anonymous list of demographic data from each student within the same classroom, school, and system of any student receiving SES Program services is required.
3. *Time of learning records.* In addition to standard attendance logs, the time spent daily learning each of the subject(s) to be evaluated (mathematics and language arts), at the appropriate grade level, should be recorded.

### *Pretest-Posttest Comparison*

The evaluation methodology to determine learning gains will solely be quantitative by comparison of actual scores of knowledge assessment instrument for the appropriate grade level in language arts and math. Ross, Potter, & Harmon (2006) assert that actual scores, as opposed to general classifications, result in more rigorous analysis. While pretest and posttest comparison of actual scores will effectively determine learning gains, provenance remains abstruse. Evaluation of matched groups and time of learning records can serve to eradicate this ambiguity.

### *Matched Groups*

One must be cognizant of the need to establish matched groups since comparison of equivalent individuals is mandatory for establishing reliability and validity during the evaluation. More fundamental to the current issue, matched groups allow accurate source attribution for learning gains. For the purpose of this evaluation, we seek to establish parity with regard to demographics and instructional hours per applicable topic.

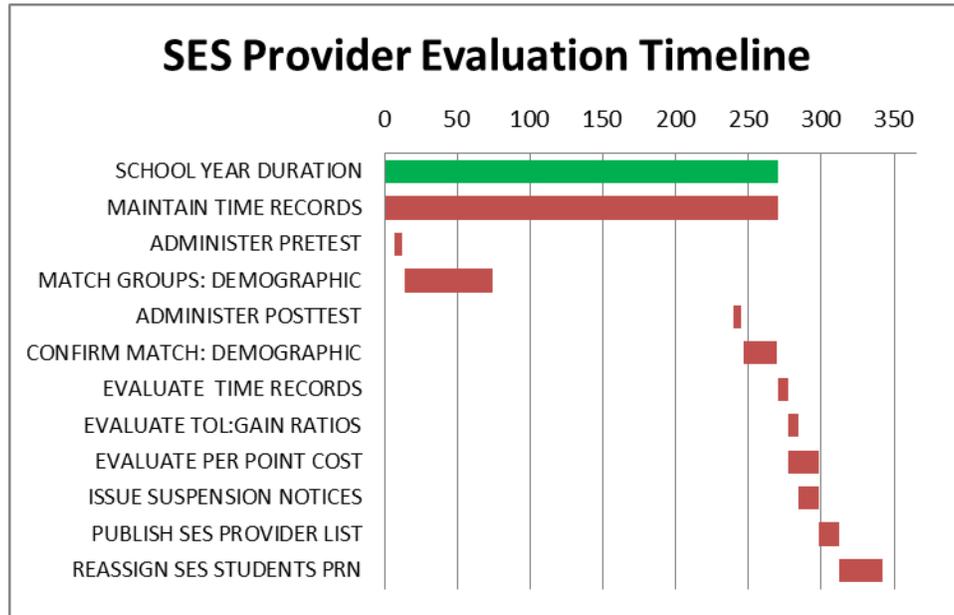
Demographic comparison is understandable. Age, whether assessed separately or not, contributes to mental ability via continued neurological development. Gender differences continue to be confirmed in self-assessment ability (Ammons & Brooks, 2013), mathematics (Jackson, Brummel, Pollet, & Greer, 2013), and I would presume numerous other areas. Socioeconomic status may limit or augment student access to educational technology and infrastructure in the home setting – as well as communication amongst parents and educators. Additional factors to consider include English language learner (ELL) status and learning disabilities that may exist.

“Like students” must also account for exposure to instruction in the evaluated topic. Here, we can turn to hours of learning records to ensure that gains were not simply due to increased instructional time in a particular classroom, school, or system. Since this is a possibility, equal *time of learning* records should be maintained by students' teachers and SES providers. Since teachers may dedicate proportionately more time to underperforming students, these records should be maintained for all students within the school system to allow for matched group comparisons.

The most reliable match for comparison would be against a demographically similar non-SES student that received equivalent hours of learning from the same teacher. Where this is not possible, comparison against similar students with equivalent hours of learning from within the school or elsewhere within the school district is possible, but an increasing compromise in reliability may occur with each step outside the student's own classroom.

*Plan*

The following timeline establishes correlations of key evaluation components in relation to the school year. It should be noted that flexibility does exist with regard to the actual number of days for each task. However, with limited exception, the general order of operations should remain consistent.



## Fidelity of Implementation

We want to ensure fidelity during implementation of the evaluation. This requires standardization- not only of the instruments- but also in the handling of such instruments. Timing, delivery, collection, scoring, and recording of evaluation instruments must be consistent for students, from both the matched (non-SES) and control (SES) groups. For SES evaluation, this involves:

- **Tool:** A uniform tool that is determined by professionals to be reliable and valid for measurement of applicable grade level content. State adopted standardized tests work well. Separate assessments must exist for mathematics and language arts.
  - **Pretest Timing:** During the first week of the school year. Tutoring may occur after the previous annual assessment (i.e. during the summer break), so pretests must be delivered prior to any SES or classroom instruction that will be evaluated.
  - **Posttest Timing:** During the last month of the school year. By the last month of the school year, learning gains (or losses) should be evident.
  - **Delivery:** The administration should follow an established protocol. Time limits and access to outside resources (i.e. text book or calculator) must be consistent. While state adopted standardized test procedures typically account for such variables, self-developed tools must also be delivered in a clearly articulated, equitable manner.
  - **Collection:** Collection should follow an established protocol to disallow variance. Once the assignment is complete, further student access should be prohibited.
  - **Scoring:** A clearly articulated, uniform grading system must be in place. Use of a common rubric or electronic grading would minimize the occurrence of subjectivity in grading. This is undoubtedly easier for mathematics or vocabulary, but technology exists to assess reading and speech abilities in a uniform manner.
  - **Recording:** Actual test scores should be recorded in an equal manner. Again, use of technology is likely for maintenance of such a comprehensive database. Where scores are manually recorded, the ability to access and alter them must be significantly limited.
- **Tool:** A uniform tool for recording time spent learning.
  - **Timing:** Daily. Accuracy is likely to be increased if records are completed each day.
  - **Collection:** Collection should follow an established protocol to disallow variance and alteration. Time spent learning each applicable content area (mathematics and language arts) should be recorded by both teachers and SES providers to the nearest one minute interval. Once the record is entered, further access should be prohibited.
  - **Recording:** Time of learning records should be maintained in a uniform manner. Again, use of technology is likely for maintenance of such a comprehensive database. Where records are manually maintained, access must be significantly limited.

Fidelity of implementation also entails authentic participation of all who are involved in the evaluation process. While subjectivity is recognized with introduction of human variables, efforts should be made to ensure that students complete assessments independently and honestly. Accountability systems must promote objective handling of all evaluation tools by educators and SES providers.

### Evaluation Results

The results of this more rigorous approach to evaluation will be an ability to accurately attribute learning gains to the appropriate source. Interpretation of evaluation findings will reliably determine whether learning gains were due to interventions offered by the student’s school or SES Provider.

#### Interpretation of evaluation findings

Actual scores from demographically equivalent students with comparable exposure to applicable content, as determined from learning assessment tool pretest data and time of learning records, establish a benchmark for each individual student. Wherever possible, more than one student can be used for comparison and each can be assigned a reliability factor based on correlation.

*Example: Assume that student 001 is the student of an SES Provider, SES999, for mathematics tutoring. Evaluating the school demographic data and hours of learning records provided by the student, we establish the following matched group of students 173, 292, and 450. Cumulatively, the matched group has a match of 101.63%. This indicates a slight advantage due to age and initial ability, as evidenced by pretest scores, but also compensates for a slight disadvantage of the matched group due to less time of learning devoted to mathematics.*

SES 999	STUDENT		AGE		GENDER		ECONOMIC		ELL		DISABLED		CLASS: MATH TOL		MATH PRETEST	
	Number	Match %	Months	Match %	M/ F	Match %	Class	Match %	Y/ N	Match %	Y/N	Match %	Minutes	Match %	Score	Match %
CONTROL	S001	100.00%	84	100.00%	M	100.00%	1	100.00%	N	100.00%	N	100.00%	5400	100.00%	72	100%
	S173	100.96%	85	101.19%	M	100.00%	1	100.00%	N	100.00%	N	100.00%	4800	88.89%	84	116.67%
	S292	100.76%	83	98.81%	M	100.00%	1	100.00%	N	100.00%	N	100.00%	5000	92.59%	82	113.89%
	S450	103.16%	87	103.57%	M	100.00%	1	100.00%	N	100.00%	N	100.00%	5800	107.41%	80	111.11%
	AVG	101.63%		101.19%		100.00%		100.00%		100.00%		100.00%		96.30%	82.00	113.89%

Prepared with this reliable data as a means for comparison, one must then assess actual scores from learning assessment post-test to determine if learning gains have occurred. Whether or not learning gains have occurred, proper attribution is still necessary.

Data from time of learning records must be evaluated to determine the efficacy of the SES provider’s time of learning versus classroom time of learning in order to properly acknowledge the source of any determined gain or loss.

Learning gains may result from a disproportionately high amount of time spent in the classroom on the particular topic, and thus would be attributable to the teacher. Alternatively, the additional time spent with the SES provider may be the cause of learning gain. Comparison of time of learning records should illuminate trends revealing the correct basis for performance.

Example continued: Student 001 has an obvious learning gain according to learning assessment tool actual scores. In fact, the gain exceeds two of the three subjects in the matched group. Yet, further evaluation is warranted. In comparing the “lower performing” matched students, we observe two key points. First, scores are considerably higher in the matched group, indicating an increased ability. More importantly, students with both less and more gains had significantly less total time of learning. The only statistically significant correlation is that more time spent in the classroom on this subject resulted in higher learning gains. Thus, the gains are likely attributable to the time of learning in the classroom.

SES 999	STUDENT		MATH POSTTEST		MATH GAIN		SES: MATH TOL		ALL: MATH TOL		CLASS: MATH TOL	
	Number	Match %	Score	Match %	Score	Match %	Minutes	Match %	Minutes	Match %	Minutes	Match %
CONTROL	S001	100.00%	80	100%	8	100%	1000	100.00%	6400	100%	5400	100.00%
MATCHED	S173	100.96%	90	112.50%	6	75.0%	0		4800	75.00%	4800	88.89%
	S292	100.76%	89	111.25%	7	87.5%	0		5000	78.13%	5000	92.59%
	S450	103.16%	90	112.50%	10	125.0%	0		5800	90.63%	5800	107.41%

This, in itself, does not equitably assign a value to the services provided. If SES provider instruction results in a student receiving 50% more annual time in a particular subject than matched group students, an associated learning gain of 150% is anticipated. Establishing time of learning: learning gain (TOL:GAIN) ratios is a simple way to understand how much learning time is typically associated with a gain of 1 for matched students, which can then be used to correlate value.

Example continued: The matched group has an average TOL:GAIN ratio of 650 with standard deviation of 66.144. Therefore, between 583.856 and 716.144 minutes should result in a gain of 1 for like students. Using the reported 1000 minutes, SES intervention should result in a learning gain of 1.396-1.713.

SES 999	STUDENT		MATH GAIN		CLASS: MATH TOL		SES: MATH TOL		ALL: MATH TOL	
	Number	Score	Minutes	Ratio	Minutes	Ratio	Minutes	Ratio		
CONTROL	S001	8	5400	675	1000	6400	800			
MATCHED	S173	6	4800	600	0	4800	600			
	S292	7	5000	625	0	5000	625			
	S450	10	5800	725	0	5800	725			
	AVERAGE		7.667	5200	650	0	5200			
	STD. DEV		2.082	529.150	66.144		529.150	66.144		

MATCHED TOL RANGE		
1SD	AVG	1SD
583.856	650	716.144
VALUE OF SES MATH TOL		
1.71275	1.538462	1.396368

Using this approach, we also have an alternative, quantitatively significant, means to determine SES provider ability. Based on the control class TOL:GAIN ratio, if a student does not fall within the matched standard deviation range, it is sensible to rule that the SES provider efforts are unsuccessful.

*Example continued: The control group (SES student 001) has a class TOL:GAIN ratio of 675. Using the matched standard deviation of 66.144, between 608.856 and 741.144 minutes should result in a gain of 1 for this student. Using the reported 1000 minutes, SES intervention should result in a learning gain of 1.396-1.713.*

CONTROL TOL RANGE		
1SD	AVG	1SD
608.856	675	741.144
VALUE OF SES MATH TOL		
1.642424	1.481481	1.349266

Value can then be assigned from these figures. 1000 minutes of SES time equals 16.67 hours. At \$50 hourly, this would represent that \$833.33 is being spent for a projected gain of 1.48 points. Expressed differently, \$563.06 per 1 point gain!

**Accountability and Transparency**

If 50% or more of the students using a particular SES provider have learning gains less than 1 standard deviation, respective to documented time of learning, that SES provider should be placed on probation until subsequent progress of students can be evaluated. All students should be reassigned to other SES providers for the following year.

If the previously underperforming students demonstrate gains during the subsequent year, a pragmatic conclusion would be that efforts of the provider are futile and further business with that SES provider should be forbidden. If, however, the students again have gains less than 1 standard deviation during the subsequent year, the SES provide may be reinstated as a qualified vendor.

Serving as a means of transparency for the public, as well as an objective representation of value, SES providers should have their associated cost per 1 point gain published on an annual basis. This may result in motivation for underperforming SES providers to increase efforts or decrease cost of service. While the former would be preferred, either outcome would benefit stakeholders in the Program.

## Conclusion

It is recognized that additional extraneous factors, such as access to educational technology, may play a role in one's learning ability. However, the evaluation plan suggested above is believed to objectively and equitably assign value to SES providers through standardized test comparison of demographically matched groups.

I am optimistic that the increased transparency of statistically significant evaluation findings will promote self-assessment and modification of suboptimal curricula to mirror that of better performing, more desired, SES providers. If substandard providers disregard best practices, this process is likely to eventually eliminate their participation by determination of substandard performance.

## References

Ammons, J. L., & Brooks, C. M. (2013). Gender and comparative evaluations in student groups. *Academy of Educational Leadership Journal, 17*, 39+.

Jackson, A. T., Brummel, B. J., Pollet, C. L., & Greer, D. D. (2013). An evaluation of interactive tablets in elementary mathematics education. *Educational Technology Research and Development, 61*(2), 311-332.

Ross, S. M., Potter, A., & Harmon, J. (2006). Evaluating supplemental education service providers: Suggested strategies for states. 2<sup>nd</sup> ed. Center for innovation and improvement.