THE GOVERNOR'S OFFICE of STUDENT ACHIEVEMENT

APPLIED LEARNING STUDENT QUESTIONNAIRE: *META-ANALYSIS*

Meta-Analysis Findings:
Fall 2012-Spring 2015

Table of Contents

Prepared by:

sengelman@sagefoxgroup.com

SageFox CONSULTING GROUP
30 Boltwood Walk • Amherst, Mass. 01002
413.256.6169 • www.sagefoxgroup.com

June 2015

tmcklin@sagefoxgroup.com

Executive Summary

Objective and Methods

The objective of the current report is to assess the longitudinal effects that 10 Innovation Fund projects had on students' attitudes across three years of data collection: 2012-2013; 2013-2014; and 2014-2015. Using data collection from the Applied Learning Student Questionnaire (ALSQ) from Fall 2012 to Spring 2015, this report seeks to address the following primary evaluation questions: *What is the summary effect size of the combined programs across the lifespan of data collection? Across the lifespan of ALSQ data collection, does project efficacy improve over time?*

To address the primary evaluation questions, a meta-analysis was conducted to estimate the total impact of the programs on students' attitudes across time. A meta-analysis is a summary or synthesis of research findings. It looks at all of the individual studies conducted and summarizes the effect sizes. An effect size is an index for describing the magnitude of an intervention's effect on outcomes. The advantage of reporting effect sizes is that it is recognized by researchers in a variety of disciplines as an easily interpretable way to quantify the effect of an intervention (Coe, 2002).[1] Effect sizes were computed using Cohen's *d* and are intended to measure the practical importance of a significant finding. Cohen (1988)[2] classified effect sizes as:

Small	≤ .20
Medium	> .20 & <.80
Large	≥ .80

Thus, this meta-analysis synthesizes effect sizes across time to arrive at an average effect size or a summary effect. To conduct the meta-analysis, effect sizes were first calculated at each point of data collection. Next, the effect sizes were weighted to account for sample size and variance.[3] That is, before combining effects, effect sizes were weighted to avoid undue influence of studies with small sample sizes. Then, an average of the weighted effect sizes is calculated. The process of synthesizing results of studies into an overall effect is called meta-analysis. All calculations were conducted using the Comprehensive Meta-Analysis (CMA) software.

In addition to computing summary effect sizes across all programs, the current report also employed a post-hoc subgroup analysis using the CMA software. A between-group heterogeneity statistic ($Q_{BETWEEN}$) was computed to test for statistical differences in the weighted effect sizes for two subsets of the population: 2-year programs vs. 3 year programs. That is, the post-hoc subgroup analysis allows us to assess how *time* or program duration differentially impact the summary effect. This analysis allows us to answer the following secondary evaluation question: *Do programs that participated in data collection across all three years show a higher summary effect size than projects that participated for 2 years or less?*

All data used in the above-mentioned analyses were derived from the ALSQ administration across six time points (T): T1: December 2012; T2: May 2013; T3: December 2013; T4: May 2014; T5: December 2014; T6: May 2015. The ALSQ was administered as a retrospective pretest survey whereby students are asked to rate their attitudes before the program and after the program. The ALSQ is comprised of the following survey constructs:

1. **Intrinsic Motivation**: motivation stemming from goals of mastery, learning and challenge. Example, "It is important for me to learn what is being taught in this program."

[1] Coe, R. (2000). *What is an effect size? A brief introduction*. Unpublished manuscript, Durham University, UK.
[2] Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed). Hillsdale, NJ: Lawrence Earlbaum Associates.
[3] Meta-analysis assigns greater weight to larger sample studies and lesser weight to smaller sample studies to produce a weighted average effect size. Thus, the standard error of effect size is used to weight effect sizes when combining studies so that large studies are considered more important than small studies in the analysis.

2. **Self-management/Self-Regulation:** effortful and persistent behaviors that are used to guide, monitor, and direct the success of one's learning and performance. Example, "I turn all my assignments in on time."
3. **Intent to Persist:** aspirations, plans, and goals to pursue additional education and a career in STEM. Example, "I intend to get a college degree in STEM (Science, Technology, Engineering, and Math)."
4. **Problem Solving:** inquiry-based learning environment that provides higher-order cognitive tasks and real-world applications. Example, "I work out explanations on my own."
5. **Implementation Activities:** hands-on activities designed to increase exposure to STEM topics and real-world applications. Example, "We learn what scientists/technicians/engineers/ mathematicians or other STEM professionals do."

Participants

The ALSQ was administered at six time points. In total, 6,751 surveys were administered from December 2012 to May 2015:

Table 1. ALSQ Administration Time Points

Time Point		n
Time 1 (T1)	*December 2012*	847
Time 2 (T2)	*May 2013*	962
Time 3 (T3)	*December 2013*	1,611
Time 4 (T4)	*May 2014*	1,350
Time 5 (T5)	*December 2014*	1,041
Time 6 (T6)	*May 2015*	970
Total		**6,751**

In total, ten programs utilized the ALSQ to collect data regarding students' attitudes. Seven programs participated in data collection for 2 years and three programs participated in data collection for 3 years. See Table 2. The programs also varied in terms of the number of student participants. For example, the programs in the following schools/districts surveyed more than 1,000 students across their lifespan: Drew Charter, Barrow County, and Carroll County. By contrast, Morehouse College, Murray County, and Georgia Tech administered the survey to less than 200 students across time.

Table 2. ALSQ Administration Time Points

Years	Program	T1: Dec. 2012	T2: May 2013	T3: Dec. 2013	T4: May 2014	T5: Dec. 2014	T6: May 2015	Overall
2	Drew Charter School- Partners of Innovation	283	273	388	304			**1,248**
2	Morehouse College Daily and Residential Summer Programs[1]		40		43			**83**
2	21st Century STEM Collaborations in Barrow County	381	312	422	358			**1,473**
2	Murray County STEM Academy	44	37	75	72			**198**
2	21st Century Academy of Environmental Studies- Rockdale County			169	179	313	301	**962**
2	Real STEM- Georgia Southern University			54	32	212	215	**513**
2	21st Century STEM Problem-Solving Skills- Georgia Tech			88	27	11	4	**130**
3	STEP Academy- Gwinnett County	62	109	129	118	230	140	**788**
3	Tift County Mechatronics Partnership	36	31	67	66	76	71	**347**
3	STEM for Life Carroll County Schools	41	160	219	151	199	239	**1,009**
	Total	**847**	**962**	**1,611**	**1,350**	**1,041**	**970**	**6,751**

[1]*ALSQ was administered in August 2013 (T2) and August 2014 (T4).*

Results & Discussion

The results are organized by evaluation questions (EQ) and discussed accordingly.

EQ 1: What is the summary effect size of the combined programs across the lifespan of data collection?

Aggregating data across all time points, Table 3 summarizes students' responses to the ALSQ survey constructs. **It is clear that the programs statistically significantly increased students' *intrinsic motivation, self-management/self-regulation skills* and *intent to persist* from 'before' to 'now.'** The summary effect sizes reveal a medium magnitude of the intervention on *intrinsic motivation, self-management/regulation* and *intent to persist*. The largest impact (*d*=0.70) was on students' *intrinsic motivation*. This suggests that the programs were particularly effective at boosting students' motivations to learn STEM and master the material being taught. Despite these promising gains in attitudes, it is important to note that the 'now' scores across the following constructs did *not* reach or exceed the optimal average of 4.00 (1, Strongly Disagree to 5, Strongly Agree): *Intent to persist, problem solving*, and *implementation activities*. See Figure 1.

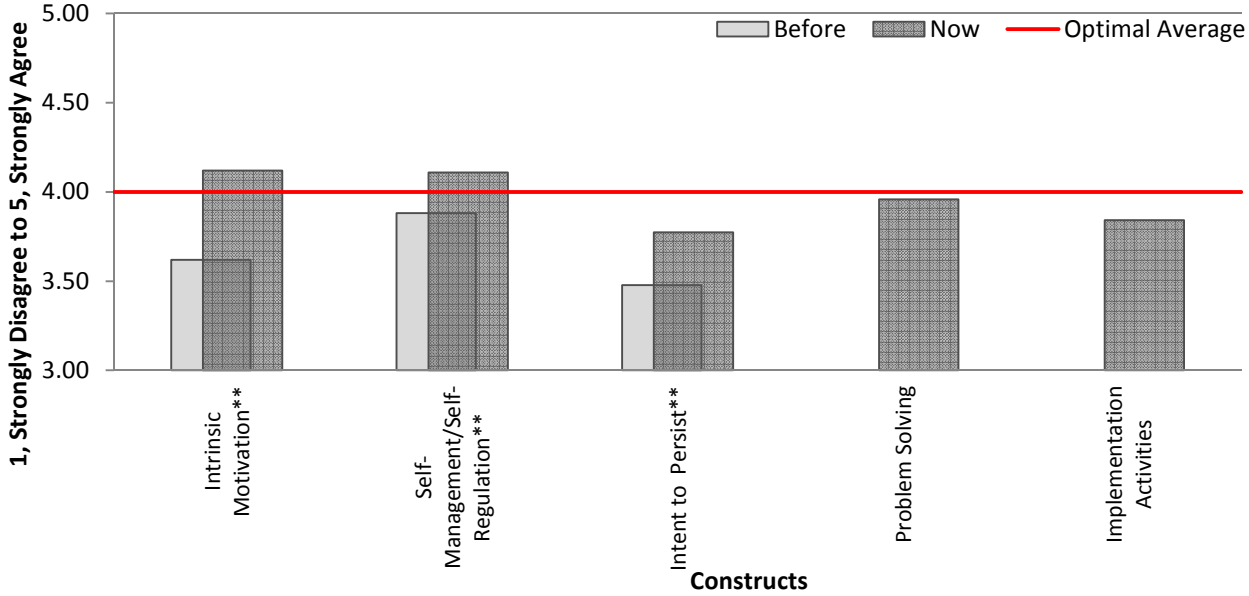
Together, the data suggest that the programs, across three years of data collection, had a positive and medium impact on students' attitudes towards STEM. However, students' intentions to persist in STEM have not reached or exceeded optimal levels. Additionally, enhancing the inquiry-based learning environment may be needed as students' problem solving skills and perceptions of the implementation activities fell slightly below optimal levels.

Table 3. Omnibus Results (n=6,751)

	Before		Now		Paired samples t-test	Summary Effect Size
	Mean	Assessment	Mean	Assessment		
Intrinsic Motivation	3.62	Attention ✓	4.12	Good ☺	p<0.001**	0.70 (M)
Self-Management/Regulation	3.88	Attention ✓	4.11	Good ☺	p<0.001**	0.44 (M)
Intent to Persist	3.48	Action !	3.77	Attention ✓	p<0.001**	0.43 (M)
Problem Solving		--	3.96	Attention ✓	--	--
Implementation Activities		--	3.84	Attention ✓	--	--

Note. Scale: 1, Strongly Disagree to 5, Strongly Agree. Assessment: Good= Above 4.0; Attention= Below 4.0; Action= Below 3.5. **p<.001, *p<.01, †p<.05. Effect size: Large (L): ≥ .80; Medium (M): > .20 & <.80; Small (S): ≤ .20. See Tables 7 – 11 for more detailed information.

Figure 1. Omnibus Results

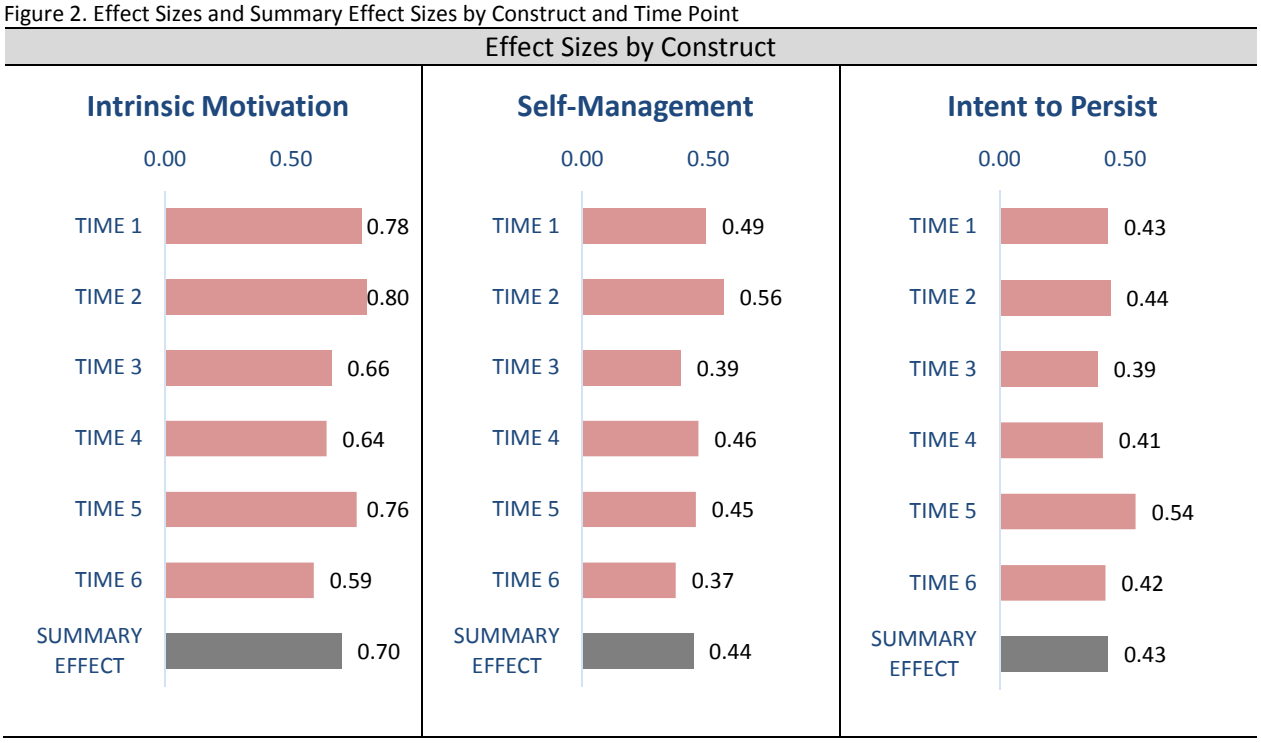


Note. **p<0.001, *p<0.01, †p<0.05. Scale was truncated to enhance visual clarity.

EQ 2: Across the lifespan of ALSQ data collection, does program efficacy improve over time?

Disaggregating the findings by time point, Figure 2 displays the effect sizes for each construct at each time point. For instance, a medium to large effect size was found across Time 1 and Time 2 (2012-2013) for *intrinsic motivation.* However, at Time 3 and Time 4, the effect size abated slightly to *d*=0.66 and *d*=0.64, respectively. By Time 5, the effect size bounced back to d=0.76. However, the effect size was at its lowest (*d*=0.59) at Time 6.

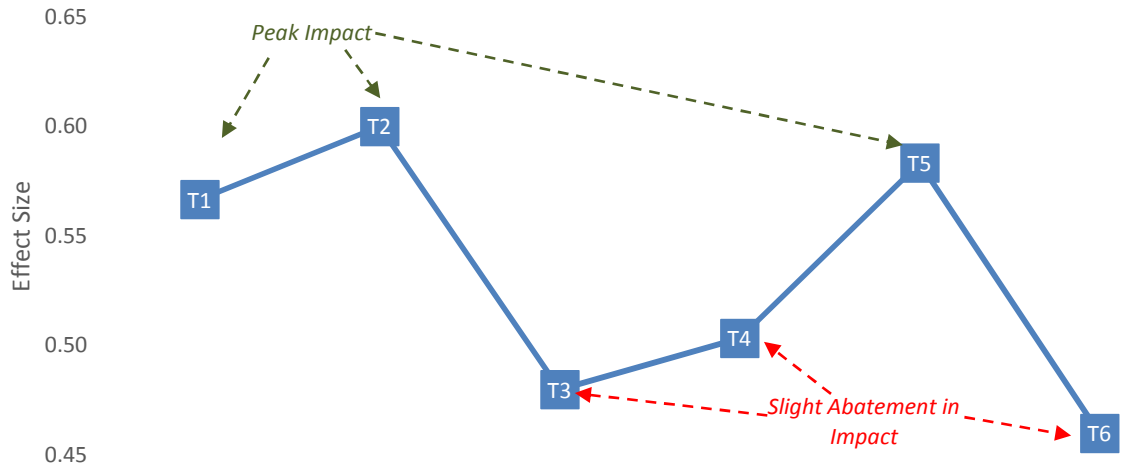
Figure 2. Effect Sizes and Summary Effect Sizes by Construct and Time Point



Note. Effect size: Large (L): ≥ .80; Medium (M): > .20 & <.80; Small (S): ≤ .20. See Tables 16 – 19 for more detailed information.

Examining trends in effect sizes across all constructs, it appears that the programs were at their peak impact at Time 1 (December 2012), Time 2 (May 2013), and Time 5 (December 2014). A slight decrement in impact is observed at Time 3 (December 2013), Time 4 (May 2014), and Time 6 (May 2015). Together, this indicates that **programs were at their peak efficacy during the first year of data collection (2012-2013**); however, **efficacy abated slightly during the second year of data collection (2013-2014) and the latter half of the third year**.

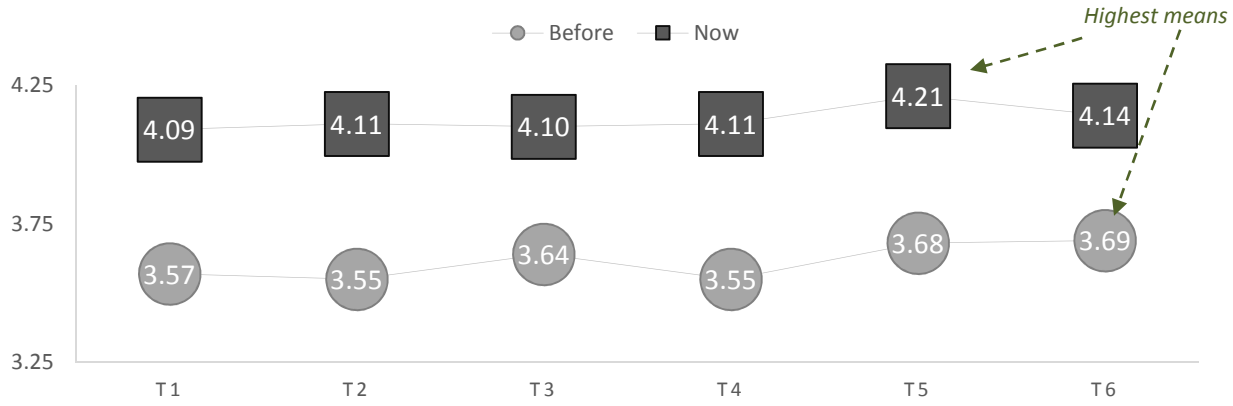
Figure 3. Longitudinal Trends in Program Impact (Effect Size)



Note. Data displayed represent average effect sizes per time point (T) across all three constructs: *Intrinsic motivation*, *self-management/regulation*, and *intent to persist*. Figure is intended to capture general longitudinal trends across constructs.

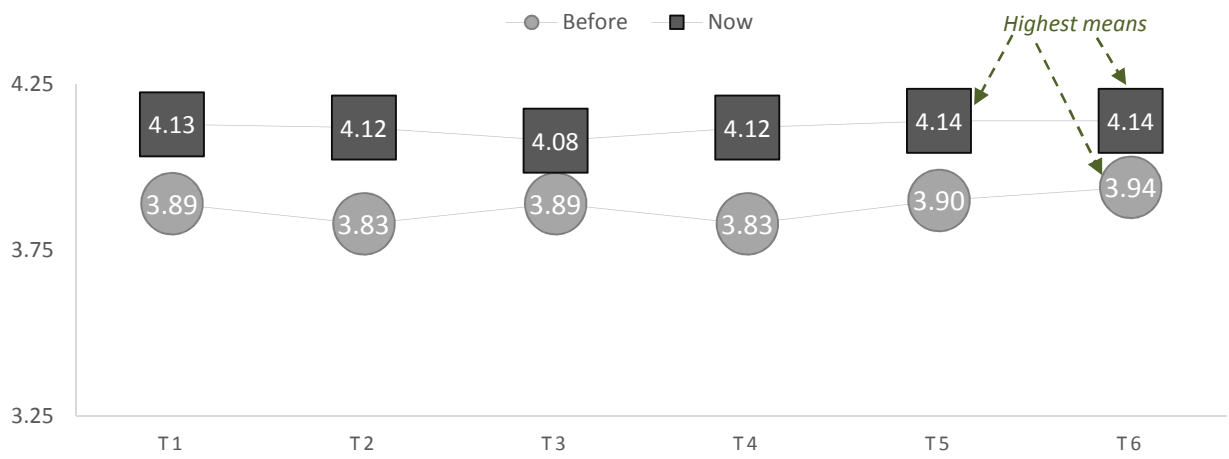
In addition to examining effect sizes across time, it is also of importance to explore longitudinal trends in mean scores for each construct. For instance, it is possible that mean scores 'before' and 'now' increased across time. Figures 4 -8 capture aggregated means for each construct at each time point. Across all constructs, **the highest means** (both "before" and "now") **are observed at Time 5** (December 2014) **and Time 6** (May 2015). This suggests that **students' attitudes have reached their highest point during the third year of data collection.**

Figure 4. Intrinsic Motivation: Means per Time Point



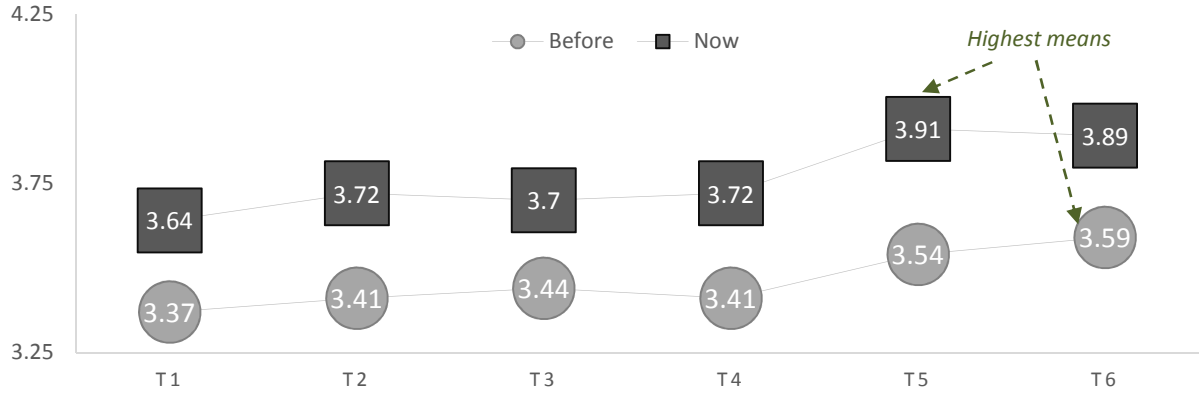
Note. Scale: 1, Strongly Disagree to 5, Strongly Agree.

Figure 5. Self-Motivation/Regulation: Means per Time Point



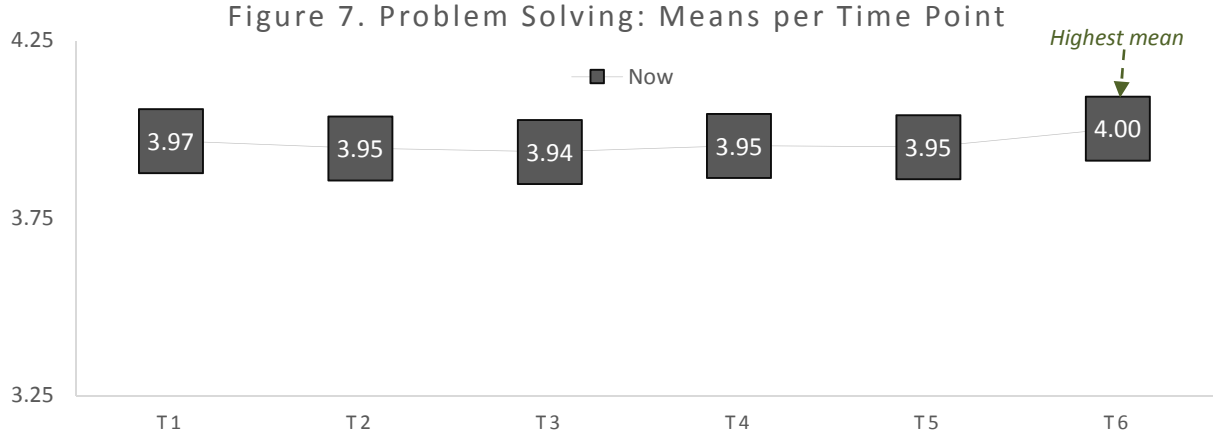
Note. Scale: 1, Strongly Disagree to 5, Strongly Agree.

Figure 6. Intent to Persist: Means per Time Point



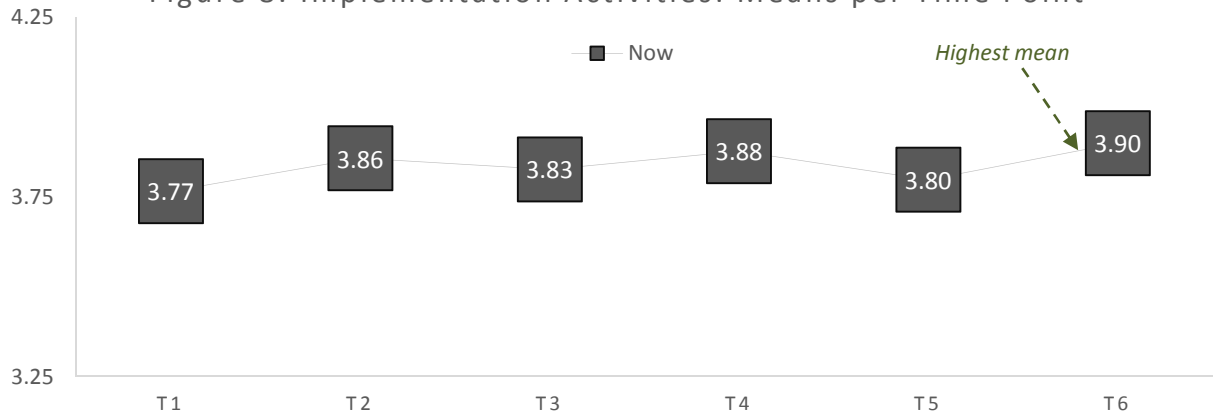
Note. Scale: 1, Strongly Disagree to 5, Strongly Agree.

Figure 7. Problem Solving: Means per Time Point



Note. Scale: 1, Strongly Disagree to 5, Strongly Agree. 'Before' data was not assessed on the ALSQ.

Figure 8. Implementation Activities: Means per Time Point



Note. Scale: 1, Strongly Disagree to 5, Strongly Agree. 'Before' data was not assessed on the ALSQ.

EQ 3: Do programs that participated in data collection across all three years show a higher summary effect size than projects that participated for two years or less?

Table 4 indicates that both 2 year and 3 year programs participating in ALSQ data collection show statistically significant increases from 'before' to 'now.' Comparing summary effect sizes between 2 year and 3 year programs, the results of the post-hoc subgroup analysis reveal that the latter had statistically significantly higher summary effect sizes than the former across all constructs. That is, the summary effect size differences between the two subsets of programs are statistically significant: Intrinsic Motivation; $Q_{BETWEEN}$= 24.87, p<.01; Self-Management/Self-Regulation: $Q_{BETWEEN}$= 31.56, p<.01; Intent to Persist: $Q_{BETWEEN}$ = 42.93, p<.010.[4] Thus, we can confidently conclude that **programs that participated in data collection for three years were more impactful than programs that participated for two years.**

Table 4. Omnibus Results by Program Type (2 year Programs vs. 3 year Programs)

	2 Year Programs (n=4,607)				3 Year Programs (n=2,144)			
	Before	Now	Paired Samples t-test	Summary Effect size	Before	Now	Paired Samples t-test	Summary Effect size
Intrinsic Motivation	3.64 Attention ✓	4.08 Good ☺	p<0.001**	**0.66 (M)**	3.57 Attention ✓	4.19 Good ☺	p<0.001**	**0.81 (L)**
Self-Manag./Reg.	3.94 Attention ✓	4.11 Good ☺	p<0.001**	**0.39 (M)**	3.75 Attention ✓	4.10 Good ☺	p<0.001**	**0.55 (M)**
Intent to Persist	3.50 Attention ✓	3.72 Attention ✓	p<0.001**	**0.37 (M)**	3.43 Action !	3.88 Attention ✓	p<0.001**	**0.55 (M)**
Problem Solving	--	3.96 Attention ✓	--	--	--	3.95 Attention ✓	--	--
Implementation Activities	--	3.82 Attention ✓	--	--	--	3.89 Attention ✓	--	--
Average				*0.47 (M)*				*0.64 (M)*

Note. Scale: 1, Strongly Disagree to 5, Strongly Agree. Assessment: Good= Above 4.0; Attention= Below 4.0; Action= Below 3.5. **p<.001, *p<.01, †p<.05. Effect size: Large (L): ≥ .80; Medium (M): > .20 & <.80; Small (S): ≤ .20.

[4] For additional information regarding post-hoc subgroup analyses using the CMA software, please refer to: Borenstein, M., Hedges, L, Higgins, J., & Rothstein, H. (2009). *Introduction to Meta-Analysis*. John Wiley & Sons, Ltd. See Chapter 19 for information regarding the mathematical computation of the *Q*-statistic.

Further analyses of program types (2 year vs 3 year programs) suggest that there is **statistically significant variability or dispersion *within* groups**.[5] See Table 5. This indicates that within the seven studies that participated in data collection for 2 years, there is a significant amount of dispersion. The three programs that participated in data collection for 3 years also show a significant amount of dispersion, albeit less than the 2 year programs. Together, we can conclude that, within groups, the efficacy of the programs vary considerably in such a manner that cannot be attributed to chance alone. For example, Table 6 indicates that, within 2 year programs, Morehouse College shows a medium to large summary effect size whereas Murray County shows a small summary effect size. This is evidence of a wide dispersion of effect sizes within 2 year programs. See Figure 9.

Table 5. Variability within Program Type (2 year vs 3 year programs)

	2 Year Programs (n=4,607)	3 Year Programs (n=2,144)
	Q_{WITHIN}	Q_{WITHIN}
Intrinsic Motivation	95.89**	28.11**
Self-Manag./Reg.	54.36**	1.54
Intent to Persist	33.23**	37.88**

Note. **p<.001, *p<.01, †p<.05.

Table 6. Summary Effect Sizes by Program and Construct

Years	Program	Summary Effect Sizes				
		Intrinsic Motivation	Self-Manag./ Regulation	Intent to Persist	*Overall Average*	Rank
2	Drew Charter School- Partners of Innovation	0.70 (M)	0.40 (M)	0.34 (M)	*0.48 (M)*	7
2	Morehouse College Daily and Residential Summer Programs[1]	0.79 (M)	1.01 (L)	0.49 (M)	*0.76 (M)*	2
2	21st Century STEM Collaborations in Barrow County	0.78 (M)	0.46 (M)	0.41 (M)	*0.55 (M)*	6
2	Murray County STEM Academy	0.25 (M)	0.11 (S)	0.13 (S)	*0.16 (S)*	10 (lowest)
2	21st Century Academy of Environmental Studies- Rockdale County	0.52 (M)	0.29 (M)	0.35 (M)	*0.39 (M)*	8
2	Real STEM- Georgia Southern University	0.82 (L)	0.47 (M)	0.53 (M)	*0.61 (M)*	4
2	21st Century STEM Problem-Solving Skills- Georgia Tech	0.24 (M)	0.29 (M)	0.17 (S)	*0.23 (M)*	9
3	STEP Academy- Gwinnett County	0.84 (L)	0.59 (M)	0.55 (M)	*0.66 (M)*	3
3	Tift County Mechatronics Partnership	1.10 (L)	0.69 (M)	0.91 (L)	*0.90 (L)*	1 (highest)
3	STEM for Life Carroll County Schools	0.70 (M)	0.53 (M)	0.47 (M)	*0.57 (M)*	5

Note. Summary effect size: Large (L): ≥ .80; Medium (M): > .20 & <.80; Small (S): ≤ .20. Color Coding: Large= Green; Medium=Orange; Small=Red. The "Overall Average" provides an average summary effect size across all three constructs.

[5] For additional information regarding post-hoc subgroup analyses using the CMA software, please refer to: Borenstein, M., Hedges, L, Higgins, J., & Rothstein, H. (2009). *Introduction to Meta-Analysis*. John Wiley & Sons, Ltd. See Chapter 19 for information regarding the mathematical computation of the *Q*-statistic.

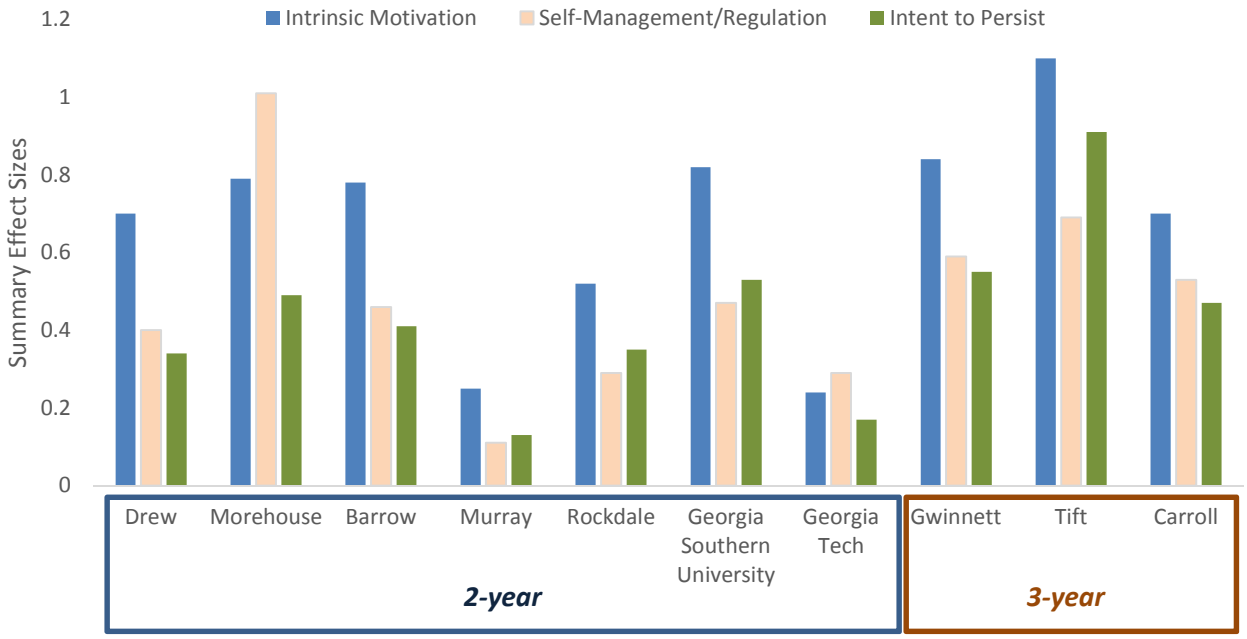
Figure 9 and Table 6 indicate that the **top 3 programs** with the largest impact (summary effect size) are:

- Tift County Mechatronics Partnership
- Morehouse College Daily and Residential Summer Programs
- STEP Academy- Gwinnett County

The three above-mentioned programs generated a medium to large impact on students' attitudes. By contrast, the **bottom 3 programs** that generated the least impact (summary effect size) are:

- Murray County STEM Academy
- 21st Century STEM Problem-Solving Skills- Georgia Tech
- 21st Century Academy of Environmental Studies- Rockdale County

Figure 9. Summary Effect Sizes by Program and Construct



Note. Summary Effect size: Large (L): ≥ .80; Medium (M): > .20 & <.80; Small (S): ≤ .20.

Summary of Findings

The current report synthesizes the results of ten programs that administered the ALSQ to students from Fall 2012 to Spring 2015. Overall, the results of the meta-analysis suggest that the effect sizes ranged from 0.43 to 0.70. This suggests a medium effect size; thus, we can state that the programs are collectively effective at improving students' attitudes towards STEM and generate a medium impact. Despite these positive findings, the programs appeared to be more efficacious in year 1 (2012-2013) than in year 2 (2013-2014) or the latter half of year 3 (2014-2015). Additionally, programs that participated in data collection for 3 years show a statistically significantly larger summary effect size than programs that participated in data collection for 2 years. This suggests that 3 year programs are more impactful on students' attitudes than 2 year programs. Areas for future improvement include 1) enhancing students' intentions to persist in STEM education, 2) ensuring that each program is being implemented with fidelity, and 3) monitoring the quality of program implementation across sites to minimize variability in program efficacy. Additional evaluation efforts may be needed in further exploring reasons for variability in program efficacy among 2 year programs, in particular, and examining gender and race/ethnic differences in program impact.

Appendix A. Detailed Omnibus Results

Table 7. Intrinsic Motivation

Intrinsic Motivation		n	Mean[1]		Paired Samples t-test[2]		1 (Strongly Disagree)	2 (Disagree)	3 (Neutral)	4 (Agree)	5 (Strongly Agree)
1. I prefer class work that is challenging so I can learn new things.	Before	6737	3.27		p<0.001**		8%	14%	36%	27%	15%
	Now	6673	3.83				4%	5%	24%	38%	29%
2. It is important to me to learn what is being taught in this program.	Before	6724	3.88		p<0.001**		3%	5%	24%	37%	31%
	Now	6658	4.32				2%	2%	11%	33%	52%
3. I like what I am learning in this program.	Before	6680	3.60		p<0.001**		4%	8%	33%	33%	21%
	Now	6631	4.10				3%	3%	17%	35%	42%
4. I think I will be able to use what I learn in this program in other classes.	Before	6672	3.62		p<0.001**		4%	10%	29%	35%	22%
	Now	6624	4.16				2%	4%	15%	35%	44%
5. Even when I do poorly on a test, I try to learn from my mistakes.	Before	6715	3.94		p<0.001**		3%	6%	20%	36%	35%
	Now	6664	4.39				1%	2%	9%	32%	55%
6. I think that what I am learning in this program is useful for me to know.	Before	6660	3.70		p<0.001**		4%	8%	29%	34%	25%
	Now	6611	4.20				2%	3%	14%	33%	47%
7. I think that what we are learning in this program is interesting.	Before	6663	3.49		p<0.001**		6%	10%	34%	30%	20%
	Now	6636	3.99				3%	5%	19%	33%	39%
8. Understanding STEM (Science, Technology, Engineering, and Math) is important to me.	Before	6705	3.62		p<0.001**		5%	9%	31%	29%	26%
	Now	6658	4.14				3%	3%	17%	32%	45%
9. I enjoy STEM (Science, Technology, Engineering, and Math) in general.	Before	6684	3.46		p<0.001**		7%	11%	33%	27%	22%
	Now	6652	3.95				4%	5%	20%	32%	39%

Note. [1] Reference lines are set at 3.5 and 4. [2]Please note that only students with matched Pre and Post data were assessed for significance. Desired statistically significant changes are highlighted in green and undesired statistically significant changes are highlighted in red **p<0.001, *p<0.01, †p<0.05. Highest percentages are highlighted in gray.

Table 8. Self-Regulation/Self-Motivation

Self-Regulation/Self-Motivation		n	Mean[1]		Paired Samples t-test[2]		1 (Strongly Disagree)	2 (Disagree)	3 (Neutral)	4 (Agree)	5 (Strongly Agree)
10. I turn all my assignments in on time.	Before	6723		3.59	p<0.001**		3%	11%	31%	32%	23%
	Now	6646		3.90			2%	6%	24%	37%	31%
11. I miss class often. (n)	Before	6682		1.67	p<0.001**		63%	19%	10%	5%	3%
	Now	6627		1.63			67%	16%	8%	5%	4%
12. I am often late for class. (n)	Before	6602		1.68	p<0.001**		61%	21%	11%	5%	3%
	Now	6566		1.66			64%	19%	9%	5%	4%
13. I set aside time to do my homework and study.	Before	6699		3.35	p<0.001**		7%	13%	34%	30%	16%
	Now	6655		3.72			5%	7%	27%	34%	27%
14. When I say I'm going to do something, I do it.	Before	6708		3.77	p<0.001**		2%	6%	30%	36%	26%
	Now	6669		4.06			1%	3%	21%	37%	37%
15. I am a hard worker.	Before	6690		4.01	p<0.001**		2%	4%	21%	35%	37%
	Now	6649		4.29			1%	2%	14%	33%	50%
16. I finish whatever I begin.	Before	6673		3.79	p<0.001**		2%	7%	29%	34%	28%
	Now	6659		4.08			1%	3%	21%	36%	39%

Note. [1] Reference lines are set at 3.5 and 4. [2] Please note that only student with matched Pre and Post data were assessed for significance. Desired statistically significant changes are highlighted in green and undesired statistically significant changes are highlighted in red. **p<0.001, *p<0.01, †p<0.05. (n) negatively worded statement. Highest percentages are highlighted in gray.

Table 9. Intent to Persist

Intent to Persist		n	Mean[1]		Paired Samples t-test[2]		1 (Strongly Disagree)	2 (Disagree)	3 (Neutral)	4 (Agree)	5 (Strongly Agree)
17. I am considering a career in STEM (Science, Technology, Engineering, and Math).	Before	6712		3.14	p<0.001**		14%	17%	30%	19%	20%
	Now	6679		3.51			11%	11%	24%	22%	31%
18. I intend to get a college degree in STEM (Science, Technology, Engineering, and Math).	Before	6702		3.26	p<0.001**		11%	15%	31%	21%	22%
	Now	6665		3.60			9%	10%	25%	23%	33%
19. I can see myself working in STEM (Science, Technology, Engineering, and Math).	Before	6691		3.16	p<0.001**		14%	16%	30%	21%	19%
	Now	6660		3.50			11%	11%	24%	25%	29%
20. Someday, I would like to have a career in STEM (Science, Technology, Engineering, and Math).	Before	6686		3.13	p<0.001**		14%	16%	31%	19%	19%
	Now	6622		3.47			11%	12%	25%	22%	29%
21. I intend to graduate from high school.	Before	6683		4.69	p<0.001**		2%	1%	6%	9%	83%
	Now	6670		4.79			1%	1%	4%	6%	88%

Note. [1] Reference lines are set at 3.5 and 4. [2]Please note that only students with matched Pre and Post data were assessed for significance. Desired statistically significant changes are highlighted in green and undesired statistically significant changes are highlighted in red **p<0.001, *p<0.01, †p<0.05. Highest percentages are highlighted in gray.

Table 10. Problem Solving, Now Only

Problem Solving	n	Mean[1]	Assessment		1 (Strongly Disagree)	2 (Disagree)	3 (Neutral)	4 (Agree)	5 (Strongly Agree)
22. In this program, my teacher(s) tells me how to improve my work.	6621	4.11	Good ☺		3%	3%	16%	36%	42%
23. In this program, my teacher(s) lets us choose our own topics or projects to investigate.	6556	3.46	Action !		7%	11%	32%	28%	22%
24. In this program, I work out explanations on my own.	6690	3.76	Attention ✓		1%	4%	32%	43%	20%
25. In this program, I have opportunities to explain my ideas.	6682	3.89	Attention ✓		2%	5%	23%	41%	29%
26. In this program, we plan and do our own projects and/or experiments.	6673	3.74	Attention ✓		4%	8%	27%	35%	27%
27. In this program, we work on real-world problems.	6684	3.94	Attention ✓		3%	5%	21%	36%	34%
28. In this program, we have class discussions.	6671	4.18	Good ☺		2%	3%	14%	35%	45%
29. In this program, we investigate to see if our ideas are right.	6657	4.02	Good ☺		2%	3%	20%	39%	35%
30. In this program, we need to be able to think and ask questions.	6656	4.26	Good ☺		2%	2%	13%	36%	47%
31. In this program, we are expected to understand and explain ideas.	6671	4.22	Good ☺		2%	2%	15%	38%	44%

Note. [1] Reference lines are set at 3.5 and 4. Assessment: Good=Above 4.0; Attention=Below 4.0; Action=Below 3.5. Highest percentages are highlighted in gray.

Table 11. Implementation Activities, Now Only

Implementation Activities	n	Mean[1]	Assessment		1 (Strongly Disagree)	2 (Disagree)	3 (Neutral)	4 (Agree)	5 (Strongly Agree)
32. In this program, my teacher(s) takes notice of students' ideas.	6615	3.87	Attention ✓		4%	5%	23%	36%	32%
33. In this program, my teacher(s) shows us how new information relates to what we have already learned.	6541	4.11	Good ☺		2%	3%	16%	38%	40%
34. In this program, we learn what scientists/ technicians/ engineers/ mathematicians or other STEM professionals do.	6655	3.76	Attention ✓		4%	7%	24%	36%	28%
35. In this program, we do our work in groups.	6645	3.82	Attention ✓		2%	4%	30%	35%	28%
36. In this program, we interact with scientists/ technicians/ engineers/ mathematicians or other STEM professionals.	6644	3.65	Attention ✓		6%	9%	25%	33%	27%

Note. [1] Reference lines are set at 3.5 and 4. Assessment: Good=Above 4.0; Attention=Below 4.0; Action=Below 3.5. Highest percentages are highlighted in gray.

Table 12. Educational Plans

What is the highest level of education you plan to achieve?	Before		Now		*Change*[1]	
	n	%	n	%	*n*	*%*
High School	1044	16%	538	8%	-506	-7.70%
2-year college	843	13%	578	9%	-265	-3.98%
4-year college	1890	29%	1380	21%	-510	-7.62%
Graduate School	1398	22%	1734	27%	+336	+5.48%
Professional School	1315	20%	2189	34%	+874	+13.84%
Total	**6491**	**100%**	**6419**	**100%**		
Average[2]	**2.97**		**3.35**		**p<0.001**(significant)[3]**	

Note. [1] Change from Before to Now. Increases are highlighted in green; decreases are highlighted in red.

[2]To compute averages, the following codes were applied: High School (1), 2-year college (2), 4-year college (3), Graduate School (4), Professional School (4).

[3]Paired samples t-test, p-value: **p<0.001, *p<0.01, †p<0.05.

Table 13. Demographics

Gender	n	%
Female	3167	48%
Male	3478	52%
Total	**6646**	**100%**

Ethnicity	n	%
Asian	243	4%
Black	2618	39%
Hispanic	624	9%
Native American	43	1%
White	2428	37%
Multiracial	491	7%
Other	202	3%
Total	**6649**	**100%**

Grade	n	%
6th	1177	18%
7th	1410	21%
8th	1637	25%
9th	348	5%
10th	433	7%
11th	696	11%
12th	924	14%
Other	0	0%
Total	**6625**	**100%**

Table 14. Participation

How long have you participated in this program?		n	%
Dosage	0 semesters	185	3%
	1 semester	2525	38%
	2 semesters	1711	26%
	3 semesters	403	6%
	4 or more semesters	1009	15%
	Summer Only	95	1%
	Don't Know	713	11%
	Total	**6641**	**100%**

Did you participate in this program during the summer?		n	%
Summer Participation	No	4896	74%
	Yes	1178	18%
	Don't Know	553	8%
	Total	**6628**	**100%**

Table 15. Program Rating

Program Rating: How would you rate this program?	n	Mean[1]	Assessment		Very Poor (1)	Poor (2)	Average (3)	Good (4)	Excellent (5)
	6633	4.16	Good ☺		3%	2%	16%	35%	44%

Note. [1] Reference lines are set at 3.5 and 4. Assessment: Good=Above 4.0; Attention=Below 4.0; Action=Below 3.5. Highest percentages are highlighted in gray.

Appendix B. Omnibus Results by Time Point

Table 16. Results by Time Point, Intrinsic Motivation

Intrinsic Motivation						
	Before		Now		Paired samples t-test	Effect size
	Mean	Assessment	Mean	Assessment		
T1 (n=847)	3.57	Attention ✓	4.09	Good ☺	p<0.001**	0.78 (M)
T2 (n=962)	3.55	Attention ✓	4.11	Good ☺	p<0.001**	0.80 (L)
T3 (n=1611)	3.64	Attention ✓	4.10	Good ☺	p<0.001**	0.66 (M)
T4 (n=1350)	3.55	Attention ✓	4.11	Good ☺	p<0.001**	0.64 (M)
T5 (n=1041)	3.68	Attention ✓	4.21	Good ☺	p<0.001**	0.76 (M)
T6 (n=970)	3.69	Attention ✓	4.14	Good ☺	p<0.001**	0.59 (M)
Omnibus	**3.62**	**Attention ✓**	**4.12**	**Good ☺**	**p<0.001****	**0.70 (M)**

Note. Scale: 1, Strongly Disagree to 5, Strongly Agree. Assessment: Good= Above 4.0; Attention= Below 4.0; Action= Below 3.5. **p<.001, *p<.01, †p<.05. Effect size: Large (L): ≥ .80; Medium (M): > .20 & <.80; Small (S): ≤ .20.

Table 17. Results by Time Point, Self-Management/Self-Regulation

Self-Management/Self-Regulation						
	Before		Now		Paired samples t-test	Effect size
	Mean	Assessment	Mean	Assessment		
T1 (n=847)	3.89	Attention ✓	4.13	Good ☺	p<0.001**	0.49 (M)
T2 (n=962)	3.83	Attention ✓	4.12	Good ☺	p<0.001**	0.56 (M)
T3 (n=1611)	3.89	Attention ✓	4.08	Good ☺	p<0.001**	0.39 (M)
T4 (n=1350)	3.83	Attention ✓	4.12	Good ☺	p<0.001**	0.46 (M)
T5 (n=1041)	3.90	Attention ✓	4.14	Good ☺	p<0.001**	0.45 (M)
T6 (n=970)	3.94	Attention ✓	4.14	Good ☺	p<0.001**	0.37 (M)
Omnibus	**3.88**	**Attention ✓**	**4.11**	**Good ☺**	**p<0.001****	**0.44 (M)**

Note. Scale: 1, Strongly Disagree to 5, Strongly Agree. Assessment: Good= Above 4.0; Attention= Below 4.0; Action= Below 3.5. **p<.001, *p<.01, †p<.05. Effect size: Large (L): ≥ .80; Medium (M): > .20 & <.80; Small (S): ≤ .20.

Table 18. Results by Time Point, Intent to Persist

Intent to Persist						
	Before		Now		Paired samples t-test	Effect size
	Mean	Assessment	Mean	Assessment		
T1 (n=847)	3.37	Action !	3.64	Attention ✓	p<0.001**	0.43 (M)
T2 (n=962)	3.41	Action !	3.72	Attention ✓	p<0.001**	0.44 (M)
T3 (n=1611)	3.44	Action !	3.70	Attention ✓	p<0.001**	0.39 (M)
T4 (n=1350)	3.41	Action !	3.72	Attention ✓	p<0.001**	0.41 (M)
T5 (n=1041)	3.54	Attention ✓	3.91	Attention ✓	p<0.001**	0.54 (M)
T6 (n=970)	3.59	Attention ✓	3.89	Attention ✓	p<0.001**	0.42 (M)
Omnibus	**3.48**	Action !	**3.77**	**Attention ✓**	**p<0.001****	**0.43 (M)**

Note. Scale: 1, Strongly Disagree to 5, Strongly Agree. Assessment: Good= Above 4.0; Attention= Below 4.0; Action= Below 3.5. **p<.001, *p<.01, †p<.05. Effect size: Large (L): ≥ .80; Medium (M): > .20 & <.80; Small (S): ≤ .20.

Table 19. Results by Time Point, Problem Solving and Implementation Activities

	Problem Solving		Implementation Activities	
	Now		Now	
	Mean	Assessment	Mean	Assessment
T1 (n=41)	3.97	Attention ✓	3.77	Attention ✓
T2 (n=160)	3.95	Attention ✓	3.86	Attention ✓
T3 (n=219)	3.94	Attention ✓	3.83	Attention ✓
T4 (n=151)	3.95	Attention ✓	3.88	Attention ✓
T5 (n=199)	3.95	Attention ✓	3.80	Attention ✓
T6 (n=239)	4.00	Good ☺	3.90	Attention ✓
Omnibus	**3.96**	**Attention ✓**	**3.84**	**Attention ✓**

Note. Items related to the above-mentioned constructs were included as 'now' only items. Scale: 1, Strongly Disagree to 5, Strongly Agree. Assessment: Good= Above 4.0; Attention= Below 4.0; Action= Below 3.5.

Appendix C. Construct Reliabilities

Table 20. Construct Reliabilities (Omnibus n=6,751)

Constructs		Cronbach's alpha	*Reliability Interpretation*
Intrinsic Motivation (9-items)	Before	0.881	*Very good*
	Now	0.903	*Excellent*
Self-Management/Self-Regulation (7-items)	Before	0.762	*Good*
	Now	0.748	*Good*
Intent to Persist (5-items)	Before	0.875	*Very good*
	Now	0.887	*Very good*
Problem Solving (10-items)	Now	0.879	*Very good*
Implementation Activities (5-items)	Now	0.831	*Very good*

Cronbach's Alpha Reliability Key: Cronbach's alpha is a measure of the internal consistency of items in a construct. This statistic ranges from 0 to 1.00; the higher the value the better. An alpha of .80 or higher is considered to have achieved very good measurement reliability; an alpha of .65 is considered acceptable (Field, 2009).

Reliability	Interpretation
.90 and above	Excellent reliability; at the level of the best measures
.80 - .90	Very good
.70 - .80	Good; in the range of most. There are probably a few items which could be improved.
.60 - .70	Somewhat low. This measure needs to be supplemented by other measures (e.g., more surveys) to determine outcomes. There are probably some items which could be improved.
.50 - .60	Suggests need for revision of measure, unless it is quite short (ten or fewer items). The test definitely needs to be supplemented by other measures (e.g., more tests).
.50 or below	Questionable reliability. This measure should not contribute heavily to the outcomes and needs revision.

From: J. C. Nunnally, Psychometric Theory. New York: McGraw-Hill, 1967, pp. 172-235.

Reference:

Field, A. (2009). *Discovering Statistics Using SPSS, 3rd Edition*. Thousand Oaks, CA: Sage Publications.