**AEA 2016 Paper Presentation**
# Title: Capturing Students' Attitudes and Persistence Towards STEM: A Progression of Evaluation Design
**Imelda Castañeda-Emenaker, EdD, Kathie Maynard, EdD, Cijy Elizabeth Sunny, ABD**

**Abstract**: This paper shares the use of a progression of program evaluation designs in the STEM Bicycle Program's three-implementation years in one, eight, and 18 high schools respectively. This project-based learning program is implemented either as a classroom-embedded or after-school program where students assemble bicycles guided by mentors from the industry. Exploratory and descriptive evaluation designs were applied during the first and the next two years accordingly. Additionally, a STEM Attitudes and Persistence Survey Instrument (SAPSI) was created (adapted from Guzey, Harwell, and Moore, 2014 and Duckworth and Quinn, 2009), piloted, and is being validated to assess students' attitudes and persistence towards STEM with the STEM Bicycle program as the platform.

## Evaluation Approach for a New STEM Project

How does one go about the overall structure of the evaluation for a new STEM Bicycle Project? Understanding the intent of the project team, a progression of evaluation design was implemented. The progression of evaluation design used in this program affirms the usefulness of starting a new and un-known program. The first year used a non-experimental exploratory and developmental evaluation design discovering and learning more about the program. Moving on to the next two years, a descriptive evaluation design was used to better understand the project's program components, services and activities, and determine the appropriate outcomes to measure. These endeavors pave the way for a possible quasi-experimental design (Shadish, et. al., 20020) to determine program effects and outcomes.

It went beyond the intent of the project team or goal of the project but rather extended to questions and search about *What will really be evaluated? Who needs to know the results of evaluation and what will be done with these results? What about the project timeline, budget and other resources?* And of course we would like the evaluation to be valid, reliable and credible!

First, what is STEM? Many or perhaps everyone may be familiar with the term STEM - STEM is an acronym for Science, Technology, Engineering, and Math education. It involves the interdisciplinary aspects of these subject areas in promoting the development of skills and knowledge essential for student success, especially when applied in real-world situations. STEM is rooted in our everyday experiences. For instance, science is in nature, technology is in our phones, engineering is in our cars. Look around and math is everywhere. Those with STEM skills can negotiate challenges, have critical thinking and problem-solving skills, have higher quality of life, and stronger earning potential. The business community, career, and economic projections for the next decade noted that 80% of area jobs in the next decade will require math & science skills; STEM jobs are growing two times faster than non-

STEM jobs; and STEM jobs pay, on average, 70% more than non-STEM jobs. It is no surprise to anyone that STEM careers are helping solve the complex problems facing our community and provide rewarding careers. Thus, the Greater Cincinnati STEM Collaborative (GCSC) focused on STEM, and more specifically, the STEM Bicycle Project.

## **The GCSC and the STEM Bicycle Club Project**

The GCSC was organized to help promote STEM education and the development of workforce within a regional, multi-sectoral, and multi-cultural context of the Midwestern region in the United States. The dynamic aspects of evaluation helped the collaborative to be more in tune with its common goals, agenda, and priorities meant to grow and sustain its efforts for the common good of students in the region and the larger community. Since its inception in 2011, the GCSC has established and is starting to scale after-school programs focused on authentic STEM experiences. Over the years, the GCSC learned about the strength of a diverse cross-functional team of education, non-profit, government, and business organizations in shaping its approach to regional collaboration. These diverse groups are learning together through its emerging learning community and institutionalization of the feedback loop as a systemic process.

GCSC started with five demonstration projects with authentic experiences in STEM, then 11 authentic STEM project experiences, with bigger and bolder plans for the coming years. However, one of GCSC's most important learning, given its limited funding, was to start the program small with one school until positive outcomes are demonstrated by students and then scaled further. The GCSC STEM Bicycle Club was born! As the GCSC (http:// http://greatercincystem.org/stem-bicycle-club/) describes this project:

*The GCSC STEM Bicycle Club is a "heads on, hands on" project that engages middle school students for 10 weeks in the after-school or within classroom environment. Students break down and re-assemble bicycles they get to keep. The STEM Bicycle Club builds student confidence and problem solving skills while reinforcing and bringing relevance to math and science principles taught during the school day. Through the involvement of community mentors and coaches, the club also exposes students to STEM career possibilities.*

**Year 1 (2013-2014) Non-experimental Exploratory and Developmental Evaluation Design Process**

The first year of the STEM Bicycle Club Project started with a non-experimental exploratory developmental approach. The critical program components and roles identified were: The involvement of the triumvirate: GCSC, schools, and community/business using the STEM-focused hands-on experience with curricular integration. The GCSC serves as the backbone organization, the schools are hosts for and implementers of the STEM Bicycle Clubs, and the community/business-industry representatives help with supplies, technical support, and mentoring. Although the GCSC has always been interested in evaluative efforts emphasizing the need for targeted works to promote its eco-system and be culturally relevant toward its goal for collective impact, the focus of the STEM Bicycle Club Project was on effects on students. Remember earlier we asked the questions: *What will really be evaluated? Who needs to know the results of evaluation and what will be done with these results? What about the project timeline, budget, and other resources? And the need for valid, reliable, and credible evaluation?*

In Year 1, the non-experimental, exploratory and developmental evaluation design was conducive in determining what was best to tease out in terms of effects on students. It was good to know that bicycle parts were provided for assembly by the students with the guidance and mentoring of community/business representatives. STEM-related curriculum aligned with the schools' academic standards were prepared for implementation along with the actual bicycle build. Developmental evaluation (Patton, 2011) and capacity building (King, 2007) helped GCSC as it went beyond being a backbone organization to stepping in to provide most needed support for real-time learning and adaptation within complex and emergent situations in this one pilot school as the project emerges. There was enthusiastic but relatively limited funding support in the beginning; thus, limited evaluation budget as well. Given the nature of the hands-on project and the experiences the students have with their mentors, it was deemed that the focus of program measures be on attitudes towards STEM, STEM careers, as well as students' persistence especially with the implementation to the inquiry driven exploratory, argument driven, and hands-on curriculum. Dubois, et.al. (2002) noted that mentoring could significantly affect students especially when strong relationships are formed between mentors and youth. Year 1's evaluation design brought about good understanding of the mentor-mentee experiences and what was emphasized during the mentoring process.

Year 1 exploration led to the adaptation (from Guzey et. al, 2014 & Duckworth and Quinn's, 2009) and piloting of a STEM Attitudes and Persistence Survey Instrument (SAPSI). An initial review of literature was performed on STEM attitudes and interests, attitudes towards science and mathematics, and Bandura's social cognitive and efficacy scales (Bandura, 2001; Bouvier & Connors, 2011; Guzey et al., 2014). Based on Guzey et al., (2014) STEM attitudes survey and Duckworth & Quinn's (2009) grit scale, the evaluator came up with a pool of 28 items. The evaluator borrowed all 8 items of Duckworth & Quinn's (2009) Short Grit Scale. The wordings to some items were refined for better understanding but there was no reduction in the number of items. There were five negatively keyed items and the rest were positively keyed. The negatively keyed items were added to minimize the effect of response bias (Furr & Bacharach, 2014). These negatively keyed items were reverse scored. The responses to the attitudes latent constructs (20 items) were scaled from strongly disagree (1) to strongly agree (5) and the responses to the persistence latent constructs (8 items) were scaled from very much like me (1) to not like me at all (5). The adapted instrument goes along with the exploratory outcomes from the program as well as what was found in literature focusing on the opinion, feelings, beliefs, and likes towards various aspects of STEM (Guzey et.al., 2014) and Duckworth & Quinn's (2009) ideas about persistence. SAPSI (as adapted from Guzey et. al, 2014 & Duckworth & Quinn, 2009) is different from other instruments that are developed in this field because this survey measures both attitudes and persistence of students towards STEM. The evaluator qualitatively operationalized (Furr and Bacharach, 2014) the initial pool of survey items. This qualitative operationalization was based on thematic analysis and alignment with the program evaluation questions. The attitudes construct had five subscales [(1) implications (personal and social), (2) learning STEM (general), (3) learning STEM related skills, (4) confidence in STEM, and (5) interest in STEM careers] and the persistence construct had three subscales [(1) determination, (2) dependability, and (3) Endurance]. Thematic analysis focuses on identifying and describing implicit and explicit ideas (Guest et. al., 2012). Since the factors or scales in the SAPSI were qualitatively operationalized, it was not clear if the factors were correlated or uncorrelated with each other. For content validity, the program team members provided feedback on the survey items. Sunny (2015) included details of the SAPSI development in her paper. Year 1 (2013-2014) project implementation involved 20 hours of building of low-cost bicycles with middle school students while building STEM skills and increasing exposure to STEM skills. With GCSC's leadership, one school, together with at least three community/business partners, were involved during the first year. Results of the pilot survey included a range of 10% to 19% increases within the sub-scales in the pre-post survey results for the 17 students involved in the project. It was not possible to do a paired comparison with the survey data at that time. Areas of growth for the project and the evaluation included the need for STEM-related curriculum to enhance the students' hands-on experience and the need to be conscious about the consistent use of the pre-post project ID for a more robust analysis of results (Castañeda, et al, 2014). The school teachers have been involved in helping students remember their project IDs unique to their schools and each student.

**Year 2 (2014-2015) Descriptive Evaluation Design Execution**

Progressing to a descriptive evaluation design during the second year of the program paves the way to better understanding of the program components and improvements needed in terms of the mode of implementation in schools, the role of mentors, and male/female differences in attitudes and persistence in STEM. In Year 2, there were eight different public schools with 94 middle and high school students from Ohio and Kentucky that were involved in the project. The choice of participating schools was purposive (Creswell, 2012). According to the evaluation services report (Castañeda-Emenaker et al., 2015), selection depended on existing school partners and/or schools within the business' vicinity. Businesses provided financial and/or in-kind support as mentors. Another criterion for selection was schools in "high poverty" situations. The collaborative adopted an open policy for program implementation since this was considered as another pilot test. Four years through its existence, GCSC has 11 active funders, 26 K-12 partner schools, five higher educational institutions, 18 business partners, and 14 identified community organizations were actively involved in 2014-2015. These were evident through documentations, interviews, focused groups, and the pre-post STEM Attitudes and Persistence Survey Instrument (SAPSI) surveys used along with the program descriptive evaluation design.

Continued validation of the SAPSI (as adapted from Guzey et. al, 2014 and Duckworth and Quinn, 2009) was pursued during the second year despite challenges in the results from the previous year. Another qualitative operationalization was conducted with the pre-test survey items. Additionally, teachers at the participating middle schools and high schools and project team members also assessed the instrument for content validity. Sunny (2015) described the instrument validation further. Exploratory Factor Analysis (EFA) with oblique rotation - promax (Furr & Bacharach, 2014) was applied to evaluate the internal structure of the SAPSI generating two constructs: attitudes and persistence. Confirmatory Factor Analysis CFA) was not used because of the small sample. The attitudes construct retained the five subscales [(1) implications (personal and social), (2) learning STEM (general), (3) learning STEM related skills, (4) confidence in STEM, and (5) interest in STEM careers]. The persistence construct ended up with two subscales [(1) determination and (2) dependability/ endurance]. Cronbach's alphas for the overall pre-survey and post-survey and for the individual constructs and sub-scales indicated moderate to high (range of .387 to .864) internal consistency indicating that the instrument is useful in assessing the attitudes and persistence of middle and high school students towards STEM, although validation of the SAPSI is not yet complete. The item language was simplified further and one persistence item was dropped. There are three negatively worded items. This time all 28 revised and adapted SAPSI items were scaled from strongly disagree (1) to strongly agree (5).

The design was primarily descriptive although the report to the project team was presented in terms of the formative and summative project components for the year (Castañeda, et. al, 2015). The project team published the project results on its website (http://greatercincystem.org) focusing on specific items that were of interest to the funders and other stakeholders: 24% increased interest in STEM careers; 11% increased interest to work in teams; 10% improved understanding of the value to learn new technologies and math. Areas of growth for evaluation remained to be the issues about consistent use of project IDs despite the teachers' involvement in providing the project IDs; thus the challenge about data cleaning to ensure the right match for robust students' results based on the pre-post surveys. One hundred fourteen students were involved but only 94 pre-post responses had been matched. The project team strategized with GCSCs central office providing a unified source for project IDs in the following year.

**Year 3 (2015-2016) Continued Descriptive Evaluation Design Execution**

Continuous descriptive evaluation design was executed in Year 3 improving the understanding of the program characteristics. Reporting the results was presented as formative/summative results because of the project team expectations. Again, program documentations and especially the pre-post SAPSI surveys were used as program measures. Knowing that it is important to have a valid and reliable instrument, continuous SAPSI validation was pursued during the third year. Eighteen schools across Cincinnati, Kentucky, and Indiana with 220 students were involved with the STEM Bicycle Club Project in Year 3.

Rasch analysis (Linacre, 2016) was used in Year 3 using 214 student pre-test responses to check on the psychometric properties of SAPSI and confirm its validity and reliability. (Note: six of the respondents specifically noted that they do not want their responses included in the analysis). Taking the instrument as unidimensional, the Rasch separation output generated a person separation (PS) index of 2.24 and a person reliability (PR) of .83. For the 28 measured items, the item separation (IS) index is 7.00 and the item reliability is .98. Cronbach Alpha is .89. Rasch analysis was also conducted using the a priori two-construct dimension of (1) attitudes and (2) persistence. For "attitudes", the Rasch separation output generated a person separation (PS) index of 2.23 and a person reliability (PR) of .84. For the 20 measured items, the item separation (IS) index is 6.07 and the item reliability is .97. Cronbach Alpha is .91. For "persistence", the Rasch separation output generated a person separation (PS) index of 1.58 and a person reliability (PR) of .71. For the 8 measured items, the item separation (IS) index is 8.60 and the item reliability is .99. Cronbach Alpha is .68. Per Linacre (2016), apart from evaluating the reliability of both the person measures and items used in the instrument, the item separation and the person separation will indicate how well the items are able to differentiate different respondents and how the items are able to differentiate by the group of respondents accordingly (Boone et.al, 2014). An item separation greater than three (IS > 3) means that the sample is able to confirm the construct validity (or item difficulty hierarchy) of the instrument (Linacre, 2012). Based on the guideline for person separation index (Wright and Masters, 1982), a person separation index of 1.5 is an acceptable separation; an index of 2.00 means a good level of separation. The project team improved its confidence with the instrument being used as a measure for student attitudes and persistence about STEM. There were no significant pre-post changes in students' STEM attitudes and persistence but there were some changes registered by five out of 18 schools especially in areas about learning with science and math, engineering and technology, as well as interests in STEM careers. Further studies about the nuances in the differences of the contexts of these schools would be reviewed to reconcile the differences. Areas of growth for evaluation remained to be the project IDs.

## Year 4 (2016-2017) Evaluation with Quasi-experimental Design

On to Year 4 (2016-2017) with 15 schools already vetted for the 2016-2017 spring implementation. The generous support of private and business funders and the encouraging results from the students' surveys, made possible this expansion in Year 4. GCSC is continuing fundraising between now and November. "Round 2" club selections of more schools will be announced December 1. Clubs are selected based on criteria published as part of the application process: Inclusion and Access, Commitment, Sustainability, and Sponsor Priority. Meanwhile, other projects sponsored by the GCSC (such as the 3-D printing Projects), which benefited from the evaluation approaches implemented in the STEM Bicycle Project will be using the SAPSI instrument during their spring 2016-2017 implementation. Although the SAPSI instrument still needs further validation, the demonstrated instrument reliability made possible the decision to use this instrument for a possible quasi-experimental design in Year 4 to determine real program effects of the STEM Bicycle Club Project. Similarly, a quasi-experimental approach may be used for the 3-D printing project using the SAPSI.

The evaluators' experience in using a progression of evaluation design and the adaptation and validation of the STEM Attitudes and Persistence Survey (SAPSI) affirm the need for the AEA evaluation guiding principles (AEA, 2004) of systematic inquiry, integrity, and competence. Additionally, these allowed for the promotion of the program evaluation standards (Yarbaugh, et. al., 2010) of utility, feasibility, propriety, accuracy, and accountability standards as results are monitored and used for program improvements and accountability to GCSC's growing funders and supporters.

## **References:**


American Evaluation Association (AEA) (2004). *Guiding principles for evaluation*. American Evaluation Association. Accessed on 15 March, 2016. Retrieved from http://www.eval.org/publications/guidingprinciples.

Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual review of psychology*, *52*(1), 1-26.

Boone, W. J., Staver, J. R., & Yale, M. S. (2014). Rasch analysis in the human sciences. NY: Springer Science+Business Media Dordrecht. doi:10.1007/978-94-007-6857-4

Bouvier, S., & Connors, K. (2011). *Increasing student interest in science, technology, engineering, and math (STEM): Massachusetts STEM pipeline fund programs using promising practice.* Report prepared for the Massachusetts Department of Higher Education, Boston, MA.

Castañeda-Emenaker, I., Sunny, C.E., Morrison, A., Dariotis, J.K. (2015). *The GCSC STEM Bicycle Club Report.* Unpublished report. Cincinnati, OH: Evaluation Services Center, University of Cincinnati.

Castañeda-Emenaker, I., Jordan, J. & Steedly, K. (2014). *Evaluation of the Greater Cincinnati STEM Collaborative (GCSC) project.* Unpublished report. Cincinnati, OH: Evaluation Services Center, University of Cincinnati.

Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Boston: Pearson Education.

DuBois, D. L., Holloway, B. E., Valentine, J.C., & Cooper, H. (2002). Effectiveness of mentoring programs for youth: A meta-analytic review. *American Journal of Community Psychology*, Vol. 30, No. 2, April 2002 (°C 2002) University of Missouri at Columbia

Duckworth, A.L. & Quinn, P. D. (2009) 'Development and Validation of the Short Grit Scale (Grit-S)', Journal of Personality Assessment, 91: 2, 166 — 174

Furr, R. M., & Bacharach, V. R. (2014). *Psychometrics: An introduction*. Thousand Oaks, CA: Sage.

Guest, G., MacQueen, K. M., & Namey, E. E. (2012). *Applied thematic analysis*. London: Sage.

Guzey, S. S., Harwell, M., & Moore, T. (2014). Development of an instrument to assess attitudes toward science, technology, engineering, and mathematics (STEM). *School Science and Mathematics*, *114*(6), 271-279.

King, J. A. (2007). *Developing* evaluation capacity through process use. In J.B. cousins (Ed.) *Process use in theory, research, and practice. New Directions for Evaluation*, 116, 45-59.

Linacre, J.M. (2016). Winsteps® Rasch measurement computer program. Beaverton, Oregon: Winsteps.com

Patton, M.Q. (2011). *Developmental evaluation*. NY: The Guilford Press.

Shadish, W. R., Cook, T.D., Campbell, D.T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston, MA: Houghton Mifflin.

Sunny, C. E. (2015). *STEM attitudes and persistence instrument evaluation.* Unpublished report. Cincinnati, OH: University of Cincinnati.

Wright, B. D. & Masters, G. (1982). *Rating scale analysis*. Chicago: Mesa Press.

Yarbaugh, D. B., Caruthers, F.A., Shulha, L.M., Hopson, R. K. (2010). *The program evaluation standards: A guide for evaluators and evaluator users*. (3rd ed.). Thousand Oaks, CA: Sage.

**Contact Information:**

Imelda Castaneda-Emenaker, EdD
castania@ucmail.uc.edu
(513) 556-3816

Kathie Maynard, EdD
maynarjk@ucmail.uc.edu
(513) 556-2023

Cijy Elizabeth Sunny, MA, ABD
sunnyce@mail.uc.edu