Understanding How Integrated Computational Thinking, Engineering Design, and Mathematics Can Help Students Solve Scientific and Technical Problems in Mathematics and Career Technical Education (INITIATE)

NSF Award #1741784



**Relevant Summative Evaluation Report for Data in SAJSE paper**

**August 2017 – June 2021**

Prepared by Acumen Research and Evaluation, LLC



**Research and Evaluation, LLC**

**2044 Paran Road**

**Cookeville, TN 38506**

**419-973-7405**

 **“Accuracy of observation is the equivalent**

 **of accuracy of thinking.”---Wallace Stevens**

|  |
| --- |
|  |
|  |

Executive Summary

Three Cohorts of secondary school teachers participated in the INITIATE professional development program. Combined, a total of 23 mathematics teachers, 9 career technology education teachers, and 3 other teachers (teacher’s aids, special education) participated. The mathematics teachers, collectively, taught Algebra 1, Geometry, Trigonometry and Pre-Calculus, Calculus I and II, and Statistics. The Career Technology Education (CTE) teachers taught Electronics (3 levels) and Manufacturing Technology. Teachers came from Toledo Public Schools and Washington Local Schools (a suburb of Toledo).

To assess evaluation questions that examined the effect INITIATE had on teachers, evaluators conducted direct observations of teachers teaching what they considered an inquiry-based lesson prior to participation in the INITIATE Summer Institute and a project-based lesson developed as a result of participating in the Summer Institute in the following academic year. Direct observations were made for Cohorts 1 and 2. Using the Electronic Quality of Inquiry Protocol (EQUIP), observations were rated on four factors of inquiry-based instruction: Instructional, Discourse, Assessment, and Curriculum. These scores were compared with post-Institute observation scores to examine growth and possible influence of the INITIATE project. Teachers showed small to moderate gains across each of the four factors, sometimes moving from a Developing Inquiry to the Proficient Inquiry category overall.

A test of computational thinking was given to the teachers from all three Cohorts on the first and last days of the SI as well as approximately eight months after the Institute. A scoring rubric was developed by a team comprised of INITIATE content experts and Acumen measurement experts. It was developed based upon computational thinking best practices. Interrater reliability was established between three members of Acumen and INITIATE. These three raters then scored all of the completed tests. Results showed statistically significant gain and a large effect size gain (Hedge’s g) in teachers’ knowledge of computational thinking after participating in the Summer Institute.

1. Teaching Practice

For Cohorts 1 and 2, Acumen visited each of the participating teachers’ classrooms in the spring prior to participation to observe a typical inquiry-based math or computer science lesson. The three special education support teachers from Cohort 2 were not observed as they did not teach lessons. Subsequent observations were made in the academic year following the teachers’ completion of the Summer Institute (SI). The purpose of the post-Institute observations was to observe teachers delivering one or both of the lessons they developed during the SI. The evaluation also examined degree to which teachers implemented the INITIATE model as designed. Observations were scored using the Electronic Quality of Inquiry Protocol (EQUIP) (Inquiry In Motion, Clemson University). Cohort 3, which was not observed directly, is discussed in a separate section.

The EQUIP rubric measures four factors associated with inquiry instruction and based upon NGSS—instruction factors, discourse factors, assessment factors, and curriculum factors. Within these four factors are 19 indicators. Scores from pre- and post-Institute observations were compared to determine if there were areas of improvement between the observations. The rating scale includes four levels—pre-inquiry, developing, proficient, and exemplary. Researchers have found a strong positive relationship between proficient scores on the EQUIP and student achievement (Marshall & Horton, 2011).

The EQUIP observation categories were converted to a four-point ordinal scale (pre-inquiry became a 1, developing became a 2, etc.) to better analyze the data. Previous evaluations (e.g., NURTURES—University of Toledo NSF MSP) showed the EQUIP observations measure our desired trait (i.e. inquiry facets of instruction).

All six Cohort 1 teachers were observed twice (i.e., before participating in the institute and after the institute). Of the 14 Cohort 2 content teachers, 12 were observed both pre and post SI. Total number of participants providing a pre and post SI observation score was 18. On average, the 18 classes observed were comprised of 46% female students, 57% minority students, and a class size of 17 (class size ranged from 10 to 24). The following sections summarize the Cohort 1 and 2 teachers’ performances across the four different inquiry-based factors measured by the EQUIP instrument.

*Instructional*

Table 1 reports the ratings for teachers’ performance on the five constructs measured under the Instructional factor. Prior to participation in INITIATE the teachers were already performing at an acceptable level (Proficient) on this factor. Overall, the median rating showed an increase in the post-Institute observations on the “Instructional Strategies,” the “Order of Instruction,” and the “Teacher Role,” indicators. For “Instructional Strategies,” teachers moved from primarily lecturing to cover content to using student-led investigations to promote strong conceptual understanding. Regarding “Order of Instruction,” teachers encouraged and expected students to explain concepts rather just the teacher explaining everything. Finally, teachers moved away from being the center of the lesson to acting more like a facilitator for students to learn in their own way.

Across all teachers’ ratings for Instructional constructs, Pre-Inquiry was selected only twice, compared to the nine times this category was selected on the pre-Institute observations. The Exemplary Inquiry rating was selected twelve times on the post observation, compared to four occurrences pre-Institute, with one teacher receiving Exemplary designation on three out of the five Instructional factor indicators.

*Table 1. Cohort 1 & 2 Comparison of observations scores for Instructional Factors*

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Instructional Strategies |  | Order of Instruction |  | Teacher Role |  | Student Role |  | Knowledge Acquisition |
|  |  |  |  |  |  |  |  |  |  |  |
| Pre-Institute | Mode | 3 |  | 2 |  | 3 |  | 3 |  | 3 |
| Median | 2 |  | 2 |  | 3 |  | 3 |  | 3 |
|  |  |  |  |  |  |  |  |  |  |  |
| Post-Institute | Mode | 4 |  | 3 |  | 4 |  | 3 |  | 3 |
| Median | 4 |  | 3 |  | 4 |  | 3 |  | 3 |

*Discourse*

Table 2 provides the ratings for teachers’ performance on the five indicators measured under the Discourse factor. The teachers showed overall improvement on the “Complexity of Questions,” “Questioning Ecology,” and “Classroom Interactions” indicators, moving from the Developing Inquiry category level to the Proficient Inquiry level on each scale. These suggest that teachers had increased success in engaging students in more open-ended discussions and did a better job of challenging students to explain and justify their answers. Combined, these three indicators reflected that teachers had increased success in engaging students in more open-ended discussions and did a better job of challenging students to explain and justify their answers.

Across all teachers’ ratings for Discourse constructs, Pre-Inquiry was selected five times on the pre-observation, occurring three times for one teacher (a second observation of this teacher showed notable increases on this construct). The Exemplary Inquiry rating was selected four times, occurring in all of the factors except “Questioning Level.” The overall ratings for Discourse showed teachers to have achieved a level of Proficient for the lessons observed.

*Table 2. Comparison of Cohort 1& 2 observations scores for Discourse Factors*

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Questioning Level |  | Complexity of Questions |  | Questioning Ecology |  | Communication Pattern |  | Classroom Interactions |
|  |  |  |  |  |  |  |  |  |  |  |
| Pre-Institute | Mode | 3 |  | 2 |  | 2 |  | 2 |  | 2 |
| Median | 3 |  | 2 |  | 2 |  | 2 |  | 2 |
|  |  |  |  |  |  |  |  |  |  |  |
| Post-Institute | Mode | 3 |  | 3 |  | 3 |  | 3 |  | 3 |
| Median | 3 |  | 3 |  | 3 |  | 3 |  | 3 |

*Assessment*

The ratings for teachers’ performance on the five indicators measured under the Assessment factor are shown in Table 3. Teachers showed increases on all indicators on the post observations indicating that they used assessment to enhance critical thinking by encouraging students to reflect on their learning at an understanding or authentic level. During post-observations teachers also solicited information to assess student understanding and took the next steps of using that information to adjust the instruction.

Across all teachers’ ratings for the Assessment factor, Pre-Inquiry was selected 13 times, but four of those occurring exclusively for one teacher (the same observation discussed in the Discourse section above). The Exemplary Inquiry rating was selected 17 times on the post-observation compared to only 3 times pre-Institute. One teacher received Exemplary rating on all five factors.

*Table 3. Comparison of Cohort 1 observations scores for Assessment Factors*

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Prior Knowledge |  | Conceptual Development |  | Student Reflection |  | Assessment Type |  | Role of Assessing |
|  |  |  |  |  |  |  |  |  |  |  |
| Pre-Institute | Mode | 2 |  | 3 |  | 2 |  | 2 |  | 2 |
| Median | 2 |  | 3 |  | 2 |  | 2 |  | 2 |
|  |  |  |  |  |  |  |  |  |  |  |
| Post-Institute | Mode | 3 |  | 4 |  | 3 |  | 3 |  | 4 |
| Median | 3 |  | 4 |  | 3 |  | 3 |  | 3 |

*Curriculum*

Table 4 reports the ratings for teachers’ performance on the four indicators measured under the Curriculum factor. While teachers improved on all indicators, the most dramatic gains were in “Learner Centrality” and “Organizing and Recording Information.” The increase in “Learner Centrality” suggests that teachers became more flexible in allowing individual students to explore and learn at their own pace as well as design and carry out their own investigations. The “Organizing” indicator reflects that teachers gave students more independence in how they organized and recorded their data and findings (as opposed to prescribing a set way). The result is a more experiential learning experience for the students.

*Table 4. Comparison of Cohort 1 & 2 observations scores for Curriculum Factors*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Content Depth |  | Learner Centrality |  | Content -Investigation Integration |  | Organizing and Recording Information |
|  |  |  |  |  |  |  |  |  |
| Pre-Institute | Mode | 3 |  | 2 |  | 3 |  | 1 |
| Median | 3 |  | 2 |  | 3 |  | 2 |
|  |  |  |  |  |  |  |  |  |
| Post-Institute | Mode | 4 |  | 4 |  | 4 |  | 3 |
| Median | 4 |  | 4 |  | 4 |  | 3 |

Across all teachers’ ratings for Curriculum constructs, Pre-Inquiry was selected just once compared to six times during pre-Institute observations. The Exemplary Inquiry rating was indicated 17 times, while it was not selected at all in pre-Institute observations. The Exemplary rating was observed most frequently in the “Integration of Content and Lesson” indicator, suggesting the teachers integrated the content of the lesson with the student investigation seamlessly.

*Overall comparison of pre- to post-Institute observations*

Prior to participation in the INITIATE Summer Institute, teachers in general displayed inquiry-based instruction at the Developing to Proficient Inquiry level with the Instructional factor showing the greatest degree of proficiency. When the teachers were observed again after their participation in the SI, there were more occurrences of Exemplary ratings across each of the four factors as well as improvements in both the mode and median scores. Overall, the teachers improved the most on the Assessment and Curriculum factors.

1. Teacher Computational Thinking Assessment

Teachers completed the Computational Thinking assessment (CT) on the first and last days of the SI, and again approximately 8 months after the SI. The Teacher CT assessment was designed by project staff and evaluators to gauge participant knowledge and understanding of computational thinking concepts and how to implement them in the classroom.

*CT assessment*

The Teacher CT assessment consists of 13 items. The first four items ask teachers to define a CT concept (algorithm design, pattern recognition, abstraction, and decomposition) and provide something a student might do to exhibit understanding of that concept. Responses are scored on a 4-point scale that rewards accurate definition of concept that clearly links with an appropriate student behavior. The next six items explore a fictitious student homework assignment where the respondent needs to identify and justify the CT concept involved with different aspects of the assignment. These responses are dichotomized into correct/incorrect identification of the element and sufficient/insufficient justification. The final three questions require respondents to detail how they would integrate CT into specific and general math and CTE classroom environments. Again, answers are scored on a 4-point scale that rewards respondents for depth, clarity, and comprehensiveness of their response. Thus, respondents could score a total of 0 (very little to no CT knowledge) up to 27 (very high level of CT knowledge).

Krippendorff’s alpha was computed in the first year to verify inter-rater reliability. The reliability index between three raters was .83 with a range of .79 to .87. Subsequent years used two of the three raters for this assessment.

*Assessment of teacher CT*

To examine if the CT assessment measured our desired trait, i.e., teacher knowledge of computational thinking concepts, raw scores obtained were converted to the logit scale. Logits are standardized units in the Rasch model. Rasch modeling was used to transform the ordinal values to an interval scale to allow for parametric comparison. The Rasch model was selected because it is a unidimensional measurement model. Included in Rasch measurements are dimensional tests that reveal if more than one trait or attribute is being measured. The Rasch interval measures of teacher CT were used to compare teacher performance on the assessment over time. Table 5 shows the average measures for Cohort before entering the INITIATE program and just following participation in the SI. For each Cohort, scores increased on the post participation assessment.

*Table 5. Average CT measures*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|   | Mean | n | Std. Deviation | Std. Error Mean |
| Cohort 1 |
| Pre-SI | 12.2 | 6 | 3.3 | 1.3 |
| Post-SI\_1 | 15.6 | 6 | 3.0 | 1.2 |
| Cohort 2 |
| Pre-SI | 10.9 | 13 | 4.4 | 1.2 |
| Post-SI | 16.9 | 13 | 4.9 | 1.4 |
| Cohort 3 |
| Pre-SI | 15.1 | 11 | 3.4 | 1.0 |
| Post-SI | 17.9 | 11 | 3.4 | 1.0 |

When combining the 30 participants with eligible pre- and post-SI CT scores from all three cohorts, Table 6 shows that the overall increase displayed between the start and the end of the two-week long summer institute was statistically significant. Further, Hedge’s g of 1.04 indicates a large effect size; namely, the pre-SI CT scores are one standard deviation lower than the post-SI scores.

*Table 6. Comparison of Teacher CT knowledge gained during Summer Institute*



1. Conclusions

Procedural suggestions recommended by the Cohort 1 teachers were reviewed by INITIATE leadership and, as a result, changes were made to the 2019 Summer Institute and the academic year follow up. Anecdotal comments from Acumen observers as well as INITIATE staff and teacher participants verified that the changes improved the program delivery. In 2020, the SI was modified to an online format. Teacher feedback confirmed that these changes still provided high quality professional development. Table 7 summarizes findings.

Table 7 summarizes evaluation findings with the evaluation questions.

*Table7. Evaluation questions*

|  |  |  |
| --- | --- | --- |
| Evaluation question | Findings | Conclusion |
| 1. Does INITIATE teacher PD improve teacher understanding of CT standards and PBL?  | Analysis of the teacher CT test shows statistically significant gains in CT knowledge after participation in the program. Direct observations of teaching pre- and post-participation show small to moderate gains in the four factors of inquiry-based instruction as measured by the EQUIP. No conclusions can be drawn about Cohort 3 as direct observation was not possible and lesson plan examination showed no change.  | Yes |
| 2. Do teachers integrate CT into their mathematics and computer science teaching? | EQUIP findings support a positive integration of CT into teaching.  | Yes |

References

Marshall, J. C., & Horton, R. M. (2011). The relationship of teacher‐facilitated, inquiry‐based instruction to student higher‐order thinking. School Science and Mathematics, 111(3), 93-101.

