Using Common Formative Assessments (CFAs) as a Means to Quantify Perceived Student Changes in IMPACTed Teachers

Gordon J. Aubrecht, II[†], Jennifer Esswein[*], Bill Schmitt[‡], and Jessica Creamer[¶]

*[†]Department of Physics, Ohio State University at Marion, *Education Northwest, ‡Science Center of Inquiry, [¶]Education Specialist*

Abstract. *Middle school teachers are involved in a project to improve their teaching methodology. The project involves content support, lesson construction, and teacher analysis of formative assessments. Common formative assessments (CFAs) were used to analyze and quantify changes in student learning across the four rubric parameters. We analyzed a sample of CFAs administered to middle and high school students across a broad range of science subjects and found evidence for changes in student ideas. For this analysis of CFAs, we established a rubric with four defining parameters: reasoning, clarity, analysis, and correctness. The rubric was revised as a result of our experience and research on this aspect of the project continues*.

In 2008, Aubrecht and Schmitt created the conception of *I*nquiry *M*odel for *P*rofessional *A*ction and *C*ontent-rich *T*eaching (IMPACT) as a means to provide professional development for teachers in school districts classified as "high needs" by the Ohio Department of Education (ODE). The initial IMPACT program has been described elsewhere (Aubrecht & Schmitt, 2010; Aubrecht, 2012). IMPACT is supported by a grant from ODE funded with federal money from the Math and Science Partnership. Esswein and Creamer joined the project in its third year; Esswein is the project's local evaluator, and Creamer, who was a TA in Aubrecht's inquiry classes, initially worked with teachers during the summer institutes and is now the project's Education Specialist. The program was started with middle and high school teachers in the Marion City School District, and 7th and 8th grade teachers in Southwest City School District joined the program in 2014 to see whether IMPACT's successes in Marion could be replicated.

Our primary goal is to help students benefit from better prepared teachers — both pedagogically and in terms of their science content knowledge. IMPACT strives to encourage teachers to honor student

questioning by enhancing their ability to "listen" to their students, and to enable an inquiry-based instructional approach to science content teaching in the classroom. The IMPACT professional development program for teachers involves three parts: content knowledge support; lesson development with pedagogical approach support; and the use of common formative assessments (CFAs) with students.

In order to increase teacher content knowledge and provide inquiry-based pedagogical support for teachers, the IMPACT program has a summer institute workshop as well as daylong grade-level meetings. During the summer institutes teachers meet for seven hours a day, one to two weeks at the end of their school year, and again for one week prior to the start of the next school year. Teachers experience inquiry-based learning — through research-based pedagogical approaches — in the role of the student. During these institutes, teachers are exposed to hands-on, minds-on approaches to science instruction that honor student thinking and reasoning. We want our teachers to learn how to assess their students formatively so they are able to design and differentiate instruction that is specific to the needs of the students rather than a "one-size-fits-all" model of curriculum development. Teachers work together to develop comprehensive, multi-week content units based on the concepts they have experienced themselves. Increased teacher content knowledge and subsequent lesson development is then carried out during the school year in the form of in-service professional development days. The IMPACT team also conducts periodic visits to teachers in their classrooms to expose teachers and students to additional subject content, provide feedback to instructional approaches, provide support for implementing inquiry in the classroom, and for the purposes of completing teacher evaluations as a measure of teacher growth.

The most important thing we eventually added to our program was to encourage teachers to use a common formative assessment (CFA) with their students. We define a CFA as: **Common** → all students at a given grade level receive same questions; **Formative** → written answers build teacher awareness of student thinking through comparison of pre – post results for the purpose of informing and influencing teaching; **Assessment** → content questions illuminate student ideas and how they change from pre – to post-application. All CFA topics are appropriate to the grade level at which they are administered and relevant to Ohio Academic Content Standards.

While a summative assessment tells teachers whether or not students understood the content, formative assessment tells us more about

the reasons ***why*** they did or did not understand the content. It takes most of one full school year to help teachers new to the program understand the idea of formative assessment. When they do finally understand how this type of assessment aids in teacher preparation and reflection of teaching practices, the results are useful for the teachers both professionally and personally. The use of CFAs in the classroom is described in greater detail in the *Methods* section of this paper.

While there are many programs for professional development (PD) available to school districts, most have traditionally been one-day workshops that attempt to change a teacher's pedagogical approach in the timespan of six-and-a-half hours. According to Garet, Porter, Desimone, Birman, and Yoon(2001), however,

> "core features of professional development activities that have significant, positive effects on teachers' self-reported increases in knowledge and skills and changes in classroom practice: (a) focus on content knowledge; (b) opportunities for active learning; and (c) coherence with other learning activities. It is primarily through these core features that the following structural features significantly affect teacher learning: (a) the form of the activity (e.g., workshop vs. study group); (b) collective participation of teachers from the same school, grade, or subject; and (c) the duration of the activity." (p. 916)

As part of their research on effective professional development programs, Garet et al. (2001) also found that professional development is likely to be of higher quality if it is both sustained over time and involves a substantial number of hours (Garet et al., 2001, p. 933). In their conclusion, they assert that to improve professional development the focus on the duration, collective participation, and the core features are essential for it to be deemed effective (Garet et al., 2001, p. 936).

Due in part to the research of Garet and others, the Ohio Department of Education requires grantees to provide at least 120 hours per year of professional development training to teachers involved in an ODE supported program. We estimate that new teachers who attend the summer institutes receive at least 188 hours of professional development per year: 70 hours, summer institutes; 70 hours, grade-level meetings; and

48-96 hours, CFA analyses. This is in addition to the time spent with staff in-class visits and teachers' voluntary attendance at professional development meetings.

As far as we know, the melding IMPACT's three elements (content support, lesson development, and the implementation of formative assessment analyses) together in a single project, which have emerged as we have worked on this project over time, is unique. The focus of this paper is to look more closely at the use and analysis of CFAs in the classroom as a means to support reflective teaching practices and the use of CFAs as a means to quantify perceived changes in several different aspects of students' acquisition of scientific communication abilities and content knowledge based on teacher participation in the IMPACT program.

Methods

Developing the Common Formative Assessments

The first use of CFAs in the classroom was implemented into Marion City Schools during the fourth year of IMPACT. Marion City Schools operate on a quarter-based grading system (as do most districts), so four different CFAs are administered to the students throughout the academic year — one set of content-specific questions per grading period. Each CFA is to be administered before the delivery of subject content (pre-CFA) and after content instruction was complete (post-CFA). The content of the CFAs would correlate with the teachers' required mapping of course subject material in accordance with Ohio's Revised Science Education Standards adopted by the Ohio Department of Education (ODE) (Ohio Department of Education, 2011). The teachers were asked to identify the "Big Ideas" and scientific understandings their students were expected to take away from each of the four grading periods, and the questions were designed accordingly. An example of the multifaceted CFA questioning style is attached as Appendix A and focuses on the concepts of force and energy — two very abstract concepts for middle school students. The CFAs do not ask for definitions of scientific vocabulary — which may be easier for students to answer but do not provide insights into student thinking or even students' true understanding of said definitions. The CFAs, rather, attempt to create a situation encouraging students to display their experience and knowledge of observable scientific phenomena.

Aubrecht and Schmitt wrote the initial formative assessment questions. The questions were open-ended by design in an attempt to elicit varying levels of student understanding and scientific reasoning abilities on a broad range of science content topics (e.g., geology, biology, physics, etc.). Aubrecht and Schmitt created a wide variety of questions for different groups of students, and the questions were discussed by and revised with input from cooperating teachers. Since the inception of the CFAs, questions have been reevaluated by teachers and rewritten to better reveal student reasoning abilities and thought patterns. When South-Western City Schools was added to the project, CFAs were written during their 2012 Summer Institute.

Common Formative Assessments in the Classroom

Teachers were asked to give the pre-CFA to students at minimum two weeks before the expected instruction of CFA subject content was to begin. The insights gained by teachers from student responses to pre-CFAs were to be used formatively: teachers were to use the students' original reasoning, answers, and misconceptions as a platform to help plan and guide instruction of course content throughout the quarter. Presentation of subject matter was to be based on students' original ideas evident in the CFA responses and instruction influenced by lessons created during summer and monthly meetings with a focus on student-centered inquiry-based instruction.

Teachers were then asked to give the post-CFA to students upon completion of subject content instruction. An analysis of student CFA pre-to-post responses (and accompanying report) was to be completed by each teacher. We did not dictate the organization of the reports. Instead, we suggested the teachers analyze the CFAs in the following manner:

> "We would like you to analyze each question separately. We ask you **not to give scores** to students, but consider what they are saying and how they are justifying or reasoning about what they are saying. The purpose of the analysis is to find out what you have learned in your reflections about how your students think about the subject you are teaching - from your students- prior to teaching it (for the pretest) and after you have finished teaching it… along with what this means to your teaching to the current

> class (e.g., 'I plan to reteach X because it was clear there was no progress from the pretest, and the way I plan to do so is …'); or classes of the future (e.g., 'Given this year's experience, next year I plan to do Y because …'). In all cases we are interested in hearing the voices of your students in the analyses. For example, we would like to see selected quotes that demonstrate student thinking, whether accurate or inaccurate, insightful or misled. The greatest value to be gained from these analyses is about how your students are thinking / reasoning rather than just giving a score to the students. *This* is what can help you as a teacher to think about how what you've learned from the students will be used in future classes."

Many teachers in the program analyzed four pre-CFAs and four post-CFAs during the academic year and reflected on their findings. This work was done outside of their official workday, and teachers were compensated for their time. A release approved by Ohio State's Institutional Review Board was signed by each participating teacher; all CFAs, reflections, and reports become the property of IMPACT to be used in scholarly research.

Another Use for Common Formative Assessments

Throughout the history of IMPACT funding, we have been strongly encouraged to develop quantitative measures to show the effects for students as a result of teachers participating in our professional development program, in addition to the anecdotal evidence presented in the *Results* section of this paper. Because we have thousands of CFAs from individual students, we decided that we could look to see if students' ideas grew through instruction and whether students would adduce evidence of their growth in their written responses to the CFAs' open-ended questions. A measure for analyzing the CFAs in a summative manner would need to be developed to quantify growth. The CFAs do not contain a multiple-choice component, which would have been easy to evaluate but would not have served our basic purpose. The IMPACT team decided instead to try to sample characteristics of student responses for cases in which we had both pre-and post-CFA data.

The first task of the IMPACT team was to develop an instrument to quantify student responses. The IMPACT team met to discuss the student characteristics that would be evaluated and a manner for analyzing said characteristics numerically. The team agreed that characteristics to be observed should honor the intentions of IMPACT's program of teacher professional development and reflect the requirements of ODE's *Science Inquiry and Application* Standards (Ohio Department of Education, 2011). The team decided on the following four categories of characteristics and defining parameters for each category:

- **Clarity**: clearness of expression in writing or drawing, 0—nothing, 1, 2—complete sentences, clear drawings
- **Reasoning**: do students reason?, 0—none, 1—superficial, 2—modest, 3—medium, 4—robust
- **Analysis**: do students bring in relevant outside knowledge in building answers?, 0—none, 1, 2—robust
- **Answer**: is the result correct?, 0—no, 1—partially, 2—completely.

Organization of student data was the second step in this process. Student CFAs were organized into batches based on the quarter in which the CFA was given (e.g., Quarter 1, Quarter 2, etc.) and the content subject in which the CFA was given (e.g., 7^{th} grade science, 8^{th} grade science, Biology, Physical Science, etc.). Student CFAs were then coded and matched to insure all data to be evaluated contained pre-and post-responses. Student CFA data were then entered into a digital spreadsheet containing the following separate categories: pretest question 1, pretest question 2, posttest question 1, and posttest question 2. All data were sent to the project evaluator, and a process was devised to determine the team's inter-rater reliability.

Inter-rater reliability was determined by Cohen's Kappa at several levels. Each of the four raters were randomly assigned overlapping students to evaluate on Test 1 and Test 2. For each test, there were four separate parts: 1) Clarity, 2) Reasoning, 3) Analysis, and 4) Correctness. The Kappa statistic has a range of 0.0 – 1.0, where a value of one is perfect alignment.

Results

Common Formative Assessments in the Classroom

Because the intention and design of the CFAs were to be used in a formative manner; initial evidence pertaining to the effects of CFA use in the classroom is primarily anecdotal. Based on selected quotations from teacher reflections with regards to the implementation and analyses of CFAs in the classroom, teachers were able to offer the following insights as to the positive effects common formative assessment had not only on reflective teaching practices but also on their students' ideas and misconceptions before and after instruction:

> One teacher wrote in the analysis of surprise at students' thoughts on the pre-CFA, *"I found that students didn't think there were wild animals in their neighborhood at first. They could only tell me about dogs and cats. I was shocked by this on the pretest. This insight helped me so we could discuss what types of animals might be found in their neighborhood so we could make our food webs."*
>
> Another teacher wrote, *"One insight … was that I found that students struggled with the WHY about how the revolving moon makes moon phases. They wanted to tell me that it was because parts of the moon disappeared somehow during its revolution."*
>
> A teacher wrote of a post-CFA, *"the majority of the students did not have a clear model that they 'owned' regarding atoms."*
>
> A teacher reflected on the analysis process, writing, *"I think the whole thing was a difficult learning experience due to the interference by the administration who really does not understand what is trying to be achieved by a common formative assessment. …"*
>
> *"Through the whole process I could not get it straight that I WAS NOT BEING EVALUATED on my teaching methods, but I*

> *was trying to understand how students actually thought about the topic under study."*
>
> In another reflection, a teacher confessed to amazement *"…at what was being written. Not correct answers and information spit back to me but actual thoughts--good, bad, or indifferent--and ways of thinking about concepts I hadn't taken into consideration."*
>
> A teacher wrote movingly, *"I think the performance record of the students is proof for itself when it comes to the impact the grant has had on our school district. The inquiry-based instruction has really brought about the developmental process that was missing from our district for the first couple [of] years I was here. I have noticed the difference in my classroom considering I have at-risk students who come from very low socioeconomic standards. It has allowed them to not be limited to their home life or economic situation, but instead to flourish academically and prove that they can do science—that they do have self-worth. To me, the impact of the grant has implications far beyond just a classroom. We are trying to change the face of a community in dire straits and we are trying to do this through academic success."*

We acknowledge that the above participating teachers' reflections are subjective. We believe, however, that the more than one hundred and twenty-seven years combined teaching experience of these teachers does qualify them as experts in their field, and their experiences and responses can be considered evidence for the effects of CFAs on teachers' reflective teaching practices and the value CFAs bring to the classroom (Marion City Schools, 2014).

Another Use for Common Formative Assessments

For all following analysis, student responses scored by multiple raters have been averaged in order to count as a single student score for each student. Each question was rated on the four characteristics, and in

order to determine significant improvement from pre-to-post, paired t-tests were performed on each and graphed in Fig. 1 below.

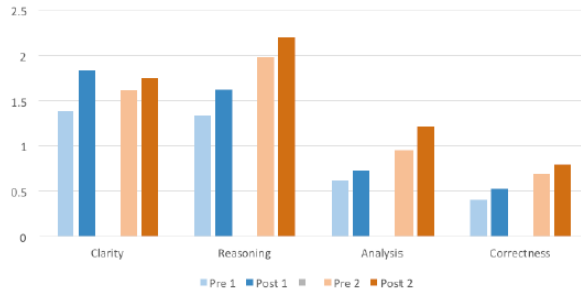

Figure 1. Average pre/post for tests 1 (blues) and 2. (reds)

Tables 1 and 2 indicate the results of the two questions.

Table 1: Paired t-tests showed significant improvement except in the Analysis section.

	Mean gain pre to post	*t value (df)*	*p*
Clarity	0.46	3.61 (24)	> 0.001
Reasoning	0.29	1.53 (24)	0.07
Analysis	0.11	0.94 (24)	0.17
Correctness	0.12	1.40 (24)	0.09

Table 2: Paired t-tests showed significant improvement in all areas at the 0.10 level

	Mean gain pre to post	*t value (df)*	*p*
Clarity	0.14	1.32 (24)	0.10
Reasoning	0.21	1.45 (24)	0.08
Analysis	0.26	2.28 (24)	0.02
Correctness	0.10	0.91 (24)	0.19

In the case of question 1, students showed the most significant improvement in the area of “Clarity”, with “Reasoning” following closely behind. The “Analysis” portion proved to be the most difficult area for

students to show improvement on the CFA question 1. Interestingly, "Analysis" showed the most improvement on question 2 where "Clarity" and "Correctness" was an area of struggle for the students. Unlike question 1, students showed statistically significant improvement in all areas of the question.

This set of CFA data is the first that the team has analyzed at the student level. The group was most aligned (had the highest Kappa statistic) with the "Correctness" portion of the CFA on both questions. The "Reasoning" and "Clarity" portions had the next highest alignment, with "Analysis" falling in last place. With this in mind, the team has discussed two possible options with refining the scoring process: (1) having multiple raters for every test and taking the average; (2) working together to come to a consensus on each student-level test before taking on another set of student responses. In order to get the most reliable data for reporting purposes, the team has chosen the latter option. Further explanation of this refining process is examined in the *Discussion* section of this paper.

Discussion

Rewriting the CFA Scoring Rubric

In order for the team to be able to reach a consensus on the scoring of CFAs, it was determined that the scoring rubric would need clarification as well as refinement. The IMPACT team developed a new rubric with more clearly defined categories, defining parameters of the categories, and agreeing on explicit criteria for scoring student CFA responses on a sliding scale. This rubric was developed with the intent to reflect the core principles of IMPACT as well as address the expected students outcomes for science instruction as described in ODE's State Science Standards (Ohio Department of Education, 2011). An abbreviated version of the new rubric is listed below. The complete modified rubric – including descriptors for sliding scale point values - has been attached to this paper as Appendix B.

- **Communication:** *Is the student able to articulate their response to the posed question(s) in a way that all components of the question are clearly addressed in a coherent fashion, regardless of correctness of answer*
    - 0 – no clarity of expression, 1- partial clarity, 2 – clarity of expression

- **Reasoning:** *Level and progression of logical thought processes (written and/or illustrated) evident in student responses, regardless of correctness of answer*
    - 0 – none, 1 – superficial, 2 – moderate, 3 – medium, 4 – robust
- **Evidence:** *Student uses own knowledge of the world around him/her (e.g., academic and/or life experiences, information given in diagrams, etc.) to support reasoning, regardless of correctness of answer*
    - 0 – none, 1 – partial, 2 – complete
- **Correctness:** *Student's answer to the posed question(s) is accurate and appropriate, given student's grade-level and expected acquisition of knowledge*
    - 0 – not correct, 1 – partially correct, 2 – correct

This rubric is intended to more clearly reflect the changes we hope to observe in students as a result of their teachers participating in our professional development program. Changes were made to two category titles during the refinement process: Clarity → Communication and Analysis → Evidence. The additional explanations of all four categories and defining parameters of point-values for each category were added as a means to clarify the differing levels of student responses we expect to see in CFAs as well as align the scoring of the project team's analyses.

The team was able to agree that in the categories of *Communication, Evidence, and Correctness* - students would be likely to show no to little evidence of ability (0 points), demonstrate some understanding of skill (1 point), or exhibit exceptional aptitude (2 points) for each category. The goal of the CFAs, however, when used in their formative intent, is to allow teachers insights into the *what* and *how* of a student's thinking and reasoning about particular scientific phenomenon. Because our professional development program encourages teachers to use this information, along with an inquiry-based approach to instruction that focuses on student reasoning and explanations, the team expected to see changes in reasoning abilities over time. We based the *Reasoning* category on a 0 – 4 point scale in an attempt to better quantify these changes.

When redefining the first category as *Communication*, the team agreed that the category's emphasis needed to be not only on a coherent response from the student, but also on the student's ability to address the

question(s) posed completely in a scientifically acceptable manner. According to *Ohio's Cognitive Demands for Science,* students should be able to "Communicate with clarity, focus, and organization using rich, investigative scenarios …" (Ohio Department of Education, 2011, p. 10). This expectation from both IMPACT and the state of Ohio is the focus of this category.

Reasoning is the sole category that has a five-point sliding scale (0-4). This category is focused on the logical progression of student thought and a student's ability to justify claims rationally in the CFAs. *Ohio's Cognitive Demands for Science* "Requires students to use scientific inquiry to develop the ability to … [think] critically and logically about the relationship between evidence and explanations …" (Ohio Department of Education, 2011, p. 10). This approach to scientific inquiry is the same methodology we model for teachers in our summer workshops and during the grade-level meetings. This is the approach we expect teachers to implement in their instruction. We expect to see changes in student reasoning abilities as a result.

The team decided to redefine the third category in our rubric as *Evidence.* In the original rubric, *Analysis* was used to describe the student ability to apply outside knowledge in support of their answer. The team decided the definition of "analysis" as we were using it was too vague, and did not completely reflect the changes we hope to observe in student responses. The focus of the *Evidence* category is to measure whether students are able to provide support for their claims. In effect, are students able to "use evidence [and] scientific knowledge to develop their explanations," as defined in ODE's *Scientific Inquiry Learning Cycle* (Ohio Department of Education, 2011, p. 6).

The *Correctness* category is based on a student's ability to use their expected acquisition of knowledge (based on their grade-level and content standards) to answer a question accurately. The Recalling Accurate Science Demand from *Ohio's Cognitive Demands for Science,* "[r]equires students to provide statements about scientifically valid facts, concepts, and relationships" (Ohio Department of Education, 2011, p. 10). While we agree that "having the right answer" is important – and we hope students will be able to articulate scientifically valid information at the culmination of instruction – this is the sole category in our rubric where "correctness of answer" is taken into account.

We believe this newly developed rubric will allow our team to more consistently analyze student CFAs for the purpose of quantifying changes in student learning as a result of teachers participating in our professional development program. The IMPACT team has begun the process of determining inter-rater reliability based on the newly devised rubric.

Limitations

The IMPACT team uses several different measures to quantify changes in teachers' content knowledge and their classroom teaching practices as a result of participation in our program: self-report measures such as RTOP, Lawson's Test of Scientific Reasoning, project team's in-class observations of participating teachers, analysis of CFAs, antidotal evidence from teacher responses, and publically available information such as data from various ODE standardized tests. The results of our previous research indicate evidence of increased teacher-content knowledge, increases in student scores on state standardized tests (Schmitt, Esswein, Aubrecht, & Creamer, 2014), and shifts from teacher-centered to student-centered instructional approaches (Aubrecht, Schmitt, Esswein, & Creamer, 2014).

Preliminary findings regarding the quantifying of CFA scores indicate students exhibit increased scores from pre- to post-test CFAS across all four defining parameters as a result of classroom instruction by teachers participating in IMPACT. There are, however, limitations to these findings:

(1) It is not entirely clear if the increases in student scores on CFAs can be attributed to IMPACTed instructional techniques or simply exposure to course content material. Because we did not visit classrooms prior to the implementation of IMPACT, nor did we administer CFAs in our treatment schools prior to implementing our professional development program, we must rely on anecdotal responses from participating teachers as to the effects IMPACT techniques and the use of CFAs in the classroom have on students.

(2) In the past, we have had no control group for which to measure our IMPACTed students against. In Marion City there is one middle- and one high school building. When IMPACT began, all teachers at each given grade-level in the school district were participants in the program. We could not use a school from a surrounding district as a control group, as circumstances regarding the demographics in Marion City Schools are unique (e.g., 80% of its students on the state's free and reduced lunch plan; high levels of generational poverty, etc.).

(3) Due in part to limited funding (and moreover to the CFAs' intent to be used by teachers as a *formative* assessment) the CFAs are not given in a "closed-test" format. The IMPACT team neither has control over the timing, delivery, and collection of CFAs, nor to the instructions given to students prior to completing CFAs.

Increasing Validity of CFA Scoring Results

In order to gain more evidence of validity when it comes to using CFAs as a measure of student growth, IMPACT has included the addition of a control group in the Southwest City School District for the 2014-2015 school year. This group will serve as a comparison between IMPACTed and non-IMPACTed teachers and the effects on students' responses to CFAs as a result of our professional development program.

Southwest City Schools' students face similar, if less drastic, demographic challenges to those students in Marion City. In the South-Western City School District, there are five different middle school buildings comprising 7th and 8th grade students. Due to individual preferences of administrators and teachers at each of the five buildings, only four of the districts' 8th grade teams of teachers and three of the 7th grade teams of teachers are currently participating in the IMPACT program.

The IMPACT team decided to approach administrators about using the one remaining team of 8th grade teachers' classrooms and the two remaining teams of 7th grade teachers' classrooms as a control group for our study. Control teachers have received no professional development from the IMPACT program (though other forms of PD have been put in place). IMPACT has agreed to provide monetary compensation to the schools in the form of money to be used for supplies and materials for the control teachers' science classrooms.

The control teachers will follow a similar process for distribution of both pre- and posttest CFAs as do the IMPACT teachers: the pre-test will be given in the control schools before content instruction is to begin; the posttest will be given after complete instruction of course content culminates. In the control schools, however, the pre- and posttest are collected by the building administrators immediately upon student completion of each CFA. We ask that control teachers place the CFAs in a manila envelope immediately upon student completion. We do not ask control teachers to analyze the CFAs in any manner.

The Quarter 1 CFA for the 2014-2015 school year will be the first set of data available to IMPACT that provides student samples of both a treatment and control group. An analysis and interpretation of this data and report of our findings is expected later this year.

Conclusion

The use of common formative assessment in the classroom as a way for teachers to gauge student understandings, prepare and alter instructional methods, and foster more reflective teaching practices is a growing practice in the field of education. Furthermore, research indicates that when students are exposed to an open-ended formative assessment format (which is style of questioning IMPACT uses in the creation of CFAs) in an environment that honors student responses (which is the type of environment IMPACT expects its teacher to foster for students) students often "… report greater feelings of competence and self-efficacy in science," (Truth-Nare & Buck, 2011, p. 395).

It is important to note that throughout IMPACT's attempt to devise a way to evaluate CFA responses summatively in order to provide quantitative data for reporting purposes, the teachers involved in our program are expected to use student responses in a formative manner only. The quantifiable results we have obtained from student CFAs are reported to funding agencies, and our numerical data have not been shared with participating teachers. Our goal for IMPACTed teachers is that they to continue to use the CFAs for their original intent — to better serve their students' individual instructional needs by developing an authentic understanding of what their students are thinking and reasoning about in science.

We do believe our attempt to evaluate a common formative assessment summatively in order to measure student growth as a result of teachers participating in our program is as novel as IMPACT's three-pronged approach to professional development. While initial findings indicate the analyzing and scoring of CFAs can be used as a means of quantitative measure for just such evaluation, more research is needed.

Acknowledgement

The IMPACT project has been supported by grants from the Ohio Department of Education C1457-OSCI-09-49 (2008-2009), C1667-MSP-10-410 (2009-2010), EDU01- 0000006141 (2010-2011), EDU01-0000007902

(2011-2012), GRT00029161 (2012-2013), ODE-MSP-10673 (2013-2014), and EDU01-0000013704 (2014-15). Opinions expressed are those of the authors, not the granting agencies.

We appreciate the assistance of Caryn Palatchi in the CFA analysis portion of this work.

References

Aubrecht, G. J., Schmitt, B., Esswein, J., & Creamer, J. (2014). Poster presented at the Mathematics and Science Partnership 2014 meeting: *Using CFAs and observations to measure growth in inquiry-based middle school science teaching.* Washington, DC. : Physics Education Research Group, The Ohio State University.

Aubrecht, G. J. (2012). Changing the way teaching occurs in an American middle school. *Latin American Journal of Physics Education 6,* 23-26.

Aubrecht, G. J., Schmitt, B. (2010). IMPACT 2: Inquiry Model for Professional Action and Content-rich Teaching – A partnership to help reform in middle school teaching and learning. *AURCO Journal 16,* 1-17.

Garet, M. S., Porter, A.C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal 38*, 915-945.

Marion City Schools District. (2014). *Marion City School District teacher tenure report* (MCS Database Report). Marion, OH: Marion City Schools District Service Center.

Ohio Department of Education. (2011). *Ohio's new learning standards: Science Standards.* Retrieved from http://education.ohio.gov/getattachment/Topics/Ohio-s-New-Learning-Standards/Science/Science_Standards.pdf.aspx

Schmitt, B., Esswein, J., Aubrecht, G. J., & Creamer, J. (2014). Poster presented at the Physics Education Research Conference 2014: *Quantifying changes in school teachers' practices.* Minneapolis, MN.

Trauth-Nare, A., & Buck, G. (2011). Using reflective practice to incorporate formative assessment in a middle school science classroom: A participatory action research study. *Education Action Research 19* 3, 379-398.

Personal Biography

Gordon J. Aubrecht, II is emeritus professor of physics at OSU Marion. He graduated from Rutgers University summa cum laude and earned his Ph.D. at Princeton University in theoretical particle physics. He is currently doing physics education research (PER) studying how students understand atoms, nuclei, and the interaction of light and matter, doing professional development, as well as how useful and effective physics by inquiry is. He was awarded the Distinguished Service Citation of the American Association of Physics Teachers in 1994, was presented with the John B. Hart Award for distinguished service from the Southern Ohio Section of the American Association of Physics Teachers in 2002, received the AURCO Distinguished Service Award, the Howard Maxwell Award for Distinguished Service from the Ohio Section of the American Physical Society, and the Louis Nemzer Award from the Ohio State Chapter of the American Association of University Professors for his defense of academic freedom in 2004, and was named Outstanding Referee of the Physical Review and Physical Review Letters in 2014. He is a past president of AURCO. He is a Fellow of American Association for the Advancement of Science, the American Association of Physics Teachers, and the American Physical Society.