Physics Education Research Group

THE OHIO STATE UNIVERSITY

Can middle school students reasoning abilities be enhanced?

Gordon Aubrecht[1], Jennifer Esswein[2], Jessica Creamer[3], and Bill Schmitt[4]

[1] Department of Physics, Ohio State University, Columbus and Marion, Ohio

[2] External Evaluator, Portland, Oregon

[3] Education Specialist, Upper Sandusky, Ohio

[4] The Science Center of Inquiry, Fountain Hills, Arizona

Abstract: Students of teacher participants in a program supported by MSP state funds in a high needs urban district are compared to students of teachers in the same district who are not participants. The program emphasizes reasoning and evidentiary thinking. Staff-generated common formative assessments were rated using a rubric. We find no significant difference between students of treatment and control teachers in the preassessment, but we do find significant differences in many postassessments. Limitations of the study will be discussed.

Research supported in part by grants from the Ohio Department of Education under the Mathematics and Science Partnership grants to the states.

The in-service teacher professional development program, IMPACT, consists of three distinct and inter-leaved parts: content support during summer institutes and school year grade-level meetings; teacher unit creation based on the IMPACT staff-created comprehensive content units as templates; and common formative assessments (CFAs). [1]

The IMPACT summer portion involves the teachers meeting for one or two weeks just after their school year ends and for a week just before school begins. Teachers experience inquiry-based learning as students both to increase content knowedge and to experience inquiry-based pedagogical approaches as students. The implementation does this through summer content workshops delivering content through research-based pedagogical approaches, and through periodic visits to teachers teaching in their classrooms. The latter offers the opportunity for the staff to expose teachers to additional content as well as affording their feedback.

QUANTITATIVE RESULTS

The aim of IMPACT is to transform teachers' teaching by enhancing their ability to "listen" to their students, to teach instructors how to formatively assess so they are able to design instruction specific to the needs of their students rather than one-size-fits-all lessons. Hence, the CFAs. [2-4]

The CFAs were originally created by the staff, and were open-ended so that students would be encouraged to write (there has been an increased emphasis in Ohio on reading and writing in recent years in response to perceptions of low success statewide). We first asked teachers what their "big ideas" were from each quarter. We restricted the number of questions to two or three. We then distributed the questions to teachers to see whether the teachers agreed that the questions assessed important issues from that quarter. We fixed the perceived difficulties. Subsequent to giving the pre-CFAs to students and teachers' discussions and analyses, we had the teachers make further changes that seemed necessary. [2,4]

Because we made changes between pre- and post-CFAs in many quarters in both grades, some compari-sons were not possible. In addition, sometimes the versions distributed to the control teahers were old versions (i.e., inconsistent with the teachers' revised versions). Just one eighth-grade CFA and three seventh-grade CFAs were identical between pre- and post- and treatment and control teachers. Paired pre- and post- lists were assembled of students who took both CFAs and the external evaluator randomly chose 100 students from treatment and control groups for analysis. The students' numbers were given to the education specialist who assessed the students' writing according to the rubric (Clarity: clearness of expression in writing or drawing; Reasoning: do students reason?; Analysis: do students bring in relevant outside knowledge in building answers?; and Correctness: is the result correct?; details shown in Table 1). Subsequently, we examined effect sizes and the external evaluator applied standard statistical tests (paired t-tests) to these data.

Control teachers changed their instruction focus because they read their students' answers to the questions—and using the students answers in their instruction. In other words, we were actually giving treatment to the control teachers! Because the CFAs constitute one-third of our treatment, we are actually not separating treatment and control teachers. [4]

Table 1: CFA rubric

COMMUNICATION

Are the students 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?

NO CLARITY OF EXPRESSION, 0 pts.	PARTIAL CLARITY, 1 pt.	CLARITY OF EXPRESSION, 2 pts.
Student response cannot be understood with regards to the question being answered. Sentence structure/penmanship/illustrations are too poor for coherent communication (e.g., "I don't know (IDK)", "yes", "no" etc.)	Student response is somewhat understood. Sentence structure/penmanship /illustrations are sufficient to convey some of the student's ideas; student does not address all components of posed question in a coherent manner, regardless of correctness of answer	Student response can be clearly understood. Sentence structure/penmanship/illustrations are able to convey student's ideas coherently and all components of posed question(s) are completely answered, regardless of correctness of answer

REASONING

Level and progression of logical thought processes (written and/or illustrated) are evident in student response, regardless of correctness of answer

NONE, 0 pts.	SUPERFICIAL, 1 pt	MODERATE, 2 pts.	MEDIUM, 3 pts.	ROBUST, 4 pts.
No reasoning is evident in student response; simple and/or incomplete answers (e.g., "yes", "no", "I don't know", etc.)	One or two word answers with little evidence of logical thought progression; no justification of ideas/explanations/claims	Some evidence of logical progression of thought demonstrated in writing and/or drawings; Student infers justifications of ideas/explanations/claims	Greater evidence of logical progression of thought in student response; Student justifies most ideas/ explanations/claims	Logical progression of thought evident throughout student response; Student justifies all ideas/ explanations/claims

EVIDENCE

Students use their own knowledge of the world around them (e.g., academic and/or life experiences, information in given diagrams, etc.) to support reasoning regardless of correctness of answer

NONE, 0 pts.	PARTIAL, 1 pt	COMPLETE, 2 pts.
No evidence is used to support reasoning; student simply restates information given in posed question	Use of scientific vocabulary and/or concepts with no explanation of knowledge of said terms (e.g., "gravity", "energy", "waves", force", etc.)	Use of scientific vocabulary and/or concepts with thorough explanation of knowledge of said terms; analysis of diagrams, charts, etc. supports reasoning; all aspects of reasoning are justified through academic content and/or life experience

CORRECTNESS

Student's answer to the proposed question is accurate and appropriate, given the student's grade-level and expected acquisition of knowledge

NOT CORRECT, 0 pts.	PARTIALLY CORRECT, 1 pt	CORRECT, 2 pts.
Student response does not contain accurate information, scientific concepts, or vocabulary	Student response may contain accurate information, scientific concepts, and/or vocabulary; answer is incomplete or partially inaccurate based on student's expected acquisition of knowledge	Student response contains accurate information, scientific concepts, and vocabulary; question is answered completely for the student's grade-level and expected acquisition of knowledge

TABLE 2. Summary Statistics for Pre-CFA Scores

	# Observations	Mean Score	Std Dev	Min	Max
Treatment Group	100	4.63	1.89	1	10
Control Group	100	3.80	1.76	0	9

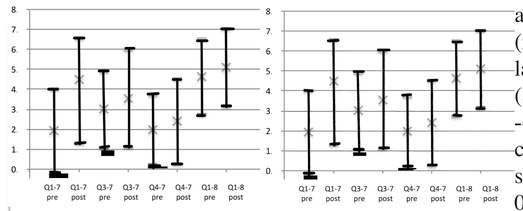
TABLE 3. Summary Statistics for Gain Scores

	# Observations	Mean Gain Score	Std Dev	Min Gain	Max Gain
Treatment Group	100	0.48	2.09	-4	6
Control Group	100	0.19	1.91	-6	5

TABLE 4. Paired t-test for gain scores 8th grade ($t_{df=99} = 1.01, p = 0.31$)

	# Observations	Mean Gain Score	Std Error	95% Confidence Interval	
Treatment Group	100	0.48	0.21	0.07	0.89
Control Group	100	0.19	0.19	-0.18	0.57

The graphs to the left show CFA scores of treatment teachers' students (left)



and control teachers' students (right). The effect sizes were calculated. We use Cohen's assessment (see Cohen's Table 1) of very large --> small, adding a "very small" category: [5] very small, < 0.10; small; 0.10 to 0.35; medium, 0.35 to 0.65; large, 0.65 to 1.00; very large, > 1.00. The results for the four characteristics are characterized in Table 5.

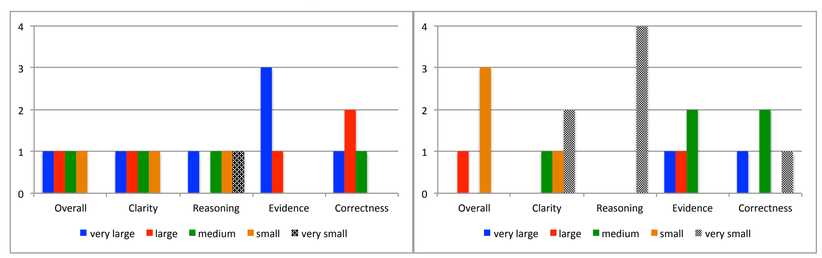
TABLE 1: Thresholds for interpreting effect size

Test	Relevant effect size	Effect size threshold Small	Medium	Large	Very large
Standardized mean difference	d, Δ, Hedges' g	.20	.50	.80	1.30
Correlation	r	.10	.30	.50	.70

Notes: The rationale for these benchmarks can be found in Cohen (1988) at the following pages: d (p.40) and r (pp.79-80). Supplementing Cohen's (1988) original small, medium and large effect sizes, Rosenthal (1996) added a classification of very large, defined as being equivalent to, or greater than d = 1.30 or r = .70.

Table 5

	Overall	Clarity	Reasoning	Evidence	Correctness
pre --> post treatment					
very large	1 (1.19)	1 (1.71)	1 (1.03)	3 (1.21, 1.17, 1.05)	1 (1.82)
large	1 (0.95)	1 (0.78)	0	1 (0.89)	2 (0.95, 0.90)
medium	1 (0.35)	1 (0.45)	1 (0.55)	0	1 (0.47)
small	1 (0.11)	1 (0.13)	1 (0.24)	0	0
very small	0	0	1 (-0.14)	0	0
pre--> post control					
very large	0	0	0	1 (1.41)	1 (1.68)
large	1 (0.73)	0	0	1 (0.76)	0
medium	0	1 (0.55)	0	2 (0.65, 0.49)	2 (0.54, 0.50)
small	3 (0.26, 0.23, 0.11)	1 (0.15)	0	0	0
very small	0	2 (0.03, -0.07)	4 (0.03, 0, -0.02, -0.11)	0	1 (-0.03)



Treatment Control

REFERENCES

1. G. J. Aubrecht and B. Schmitt, "IMPACT 2: Inquiry Model for Professional Action and Content-rich Teaching -- A partnership to help reform middle school teaching and learning," *AURCO J*. **16,** 1-17 (2010). G. J. Aubrecht, "Changing the way teaching occurs in an American middle school," *Lat. Am. J. Phys. Educ*. **6,** Supp. 1, 23-26 (2012). G. J. Aubrecht, "Three steps to successful change," in Dvořák, L. and Koudelková, V., eds., *Active learning – in a changing world of new tech-nologies*, p. 268-278 (Prague, ICPE, 2014). G. J. Aubrecht, J. Esswein, J. Creamer, and B. Schmitt, "Revealing effects of changes in middle school science teachers' practices," in A. D. Churukian, D. L. Jones, and L. Ding, eds., *PERC Proceedings 2015*, 39-42.
2. G. J. Aubrecht, B. Schmitt, and J. E. Esswein, "Using common formative assessments with middle and high school students to inform pedagogical approaches for teaching scientific content," in M. Fatih Taşar, ed., *Proceedings of the World Conference on Physics Education 2012*, p. 593-600 (Ankara, Turkey: Pegem Akademi, 2014). G. J. Aubrecht, J. Esswein, B. Schmitt, and J. Creamer, "Using common formative assessments (CFAs) as a means to quantify perceived student changes in IMPACTed teachers," *AURCO J*. **21**, 1-18 (2015).
3. G. J. Aubrecht, , J. Esswein, J. Creamer, and B. Schmitt, "Quantifying middle school students' reasoning abilities," submitted to *Lat. Am. J. Phys. Educ*.
4. G. J. Aubrecht, J. Creamer, J. Esswein, and B. Schmitt, "Recognizing changes in middle school science teachers' practices," *AURCO J*. **22**, 1-26 (2016). G. J. Aubrecht, , J. Esswein, J. Creamer, and B. Schmitt, "Seeing effects of changes in middle school science teachers' practices," submitted to *Lat. Am. J. Phys. Educ*.
5. J. Cohen, *Statistical Power Analysis for the Behavioral Sciences* (London: Routledge,1988)