EVALUATING PROFESSIONAL DEVELOPMENT WORKSHOPS QUICKLY AND EFFECTIVELY

Charles Hayward and Sandra Laursen
Ethnography & Evaluation Research, University of Colorado Boulder

Abstract

Many funding agencies require evaluation of the impact of professional development projects they support. However, improved student outcomes, the ultimate goal, may take longer to be realized than the project time frame allows. Instructors need time to implement and refine new skills before positive student outcomes are realized, a delay that may be exacerbated in classes that are not taught frequently. We report on one example of an efficient and cost-effective self-report measure designed to detect the initial changes in teaching practices that lead to improved student outcomes over time. We discuss the ability for timely and accurate measures through this instrument. Results support the interpretation that instructors' reported teaching practices show changes consistent with methods taught at professional development workshops on Inquiry-Based Learning in mathematics. Additionally, correlations with self-reported level of implementation suggest that instructors are reporting honestly, and not just socially desirable changes consistent with their concept of "real Inquiry-Based Learning."

Key words: Evaluation, Measurement, Professional Development, Inquiry-Based Learning

Background

After decades of innovation and research it is clear that certain reforms of classroom practice improve undergraduate education in science, technology, engineering and mathematics (STEM). Research in cognitive science and education offers persuasive evidence that students can and do learn better through active, student-centered forms of instruction (Hake, 1998; Springer, Stanne, & Donovan, 1999; Bransford, Brown, & Cocking, 2000; Prince, 2004; Ruiz-Primo, Briggs, Iverson, Talbot, & Shepard, 2011; Deslauriers, Schelew, & Wieman, 2011). Yet relatively few students experience these proven, "high-impact" educational practices during college (Kuh, 2008). The President's Council of Advisors on Science and Technology (PCAST, 2012) advocated active learning strategies in order to meet its goal of an additional 1 million STEM graduates over the next decade.

Uptake of these methods by large numbers of faculty at diverse institutions is now the bottleneck in improving STEM higher education (Fairweather, 2008; Henderson & Dancy, 2007; 2008; 2011). To broaden uptake of student-centered teaching and learning approaches, professional development of college instructors (CIPD) is crucial. But efforts to broaden the reach of CIPD must be based on good evaluation evidence about whether it is having the desired effect on teaching. While improved student learning is the ultimate goal of CIPD, measuring student outcomes directly is not always feasible due to the cost and complexity. Additionally, the impact on students lies far downstream from the intervention itself (Guskey, 2002) as instructors must apply and refine the methods before positive student results can be detected. Instructors may not teach a course every year, resulting in a large lag between the CIPD and any detection of positive student outcomes.

Given this time lag, positive student outcome results may help to demonstrate the merit of a particular CIPD intervention, but they do not provide formative feedback to help instructor developers to diagnose or improve a particular intervention. Measuring the impact of CIPD through its effect on student outcomes can contribute to the research base, but does not

Hayward, C. & Laursen, S. (2014, February - March). Evaluating professional development workshops quickly and effectively. Paper presented at the Conference on Research in Undergraduate Mathematics Education. Denver, CO.

provide a nuanced and flexible evaluation tool that is responsive on the time scale on which CI developers—those who plan and lead the professional development—must adapt, refine, and report results of their intervention. Rather, we need reliable and valid evaluation tools to measure whether and to what extent CIs have made changes to their instruction as a result of CIPD interventions. Measuring changes in instructional practices is challenging. Multiple observation protocols have been developed to measure changes in teaching practices, however they are often time-consuming and disruptive to classrooms (Hora & Ferrare, 2012). Surveys are easier to conduct, but some have argued that they are inaccurate due to respondent biases (Desimone, 2009).

This study reports on evaluation of professional development workshops for college instructors on inquiry-based learning (IBL) in mathematics. IBL is a form of active learning that helps students develop critical thinking through ill-defined problems and by constructing and evaluating mathematical arguments. IBL is based on the teaching practices of R.L. Moore (1882-1974), a mathematician at the University of Texas, Austin. His teaching method involved students using definitions, logic, and precise language to prove mathematical theorems (Mahavier, 1999). Students worked independently and were not allowed to consult other students or textbooks. They then presented proofs in class and were critiqued. Today, this method is typically modified to allow more student collaboration and is referred to as the Modified Moore Method. IBL has emerged as a broader umbrella term encompassing Moore's method as well as others that share the spirit of student inquiry through deep engagement with mathematics and collaboration with peers (Yoshinobu & Jones, 2013). [For an example, see (Schumacher, 2010).] In all forms of IBL, students learn through analyzing ill-defined problems and constructing and evaluating arguments (Prince & Felder, 2007; Savin-Baden & Major, 2004). This supports deep learning of mathematical concepts (Moon, 2004; McCann, Johannessen, Kahn, & Smagorinsky, 2004). To teach in this manner, many instructors must transition from traditional lecture methods to more student-centered teaching approaches. [For an example, see (Retsek, 2013).] The professional development workshops we have studied aim to help instructors make this transition.

Conceptual Framework

Guskey (2000) classifies evaluation of professional development (PD) into five levels of increasing complexity. At higher levels, evaluation requires increased time and resources. Each level builds upon those before it and varies as to the questions evaluators address, how the data is gathered, what is measured, and how the information is used. The first level comprises participants' immediate reactions to the PD, while Level 2 goes further to address what participants have learned from the PD. In Level 3, evaluation measures organizational support and change. Participants' use of new knowledge and skills is measured in Level 4, and Level 5, the most complex, addresses student learning outcomes. Our larger project evaluates professional development workshops for college instructors on Inquiry-Based Learning in mathematics at Levels 1 through 5. Since student learning (Level 5) is very difficult to assess within the timeframe of grant-funded projects, this report focuses on the next highest level (Level 4), evaluation of participants' use of new knowledge and skills.

While many evaluation efforts at this level use classroom observation protocols to assess participants' implementation of methods presented during professional development workshops, these are time- and resource-intensive and may interfere with normal classroom dynamics (Guskey, 2000). On the other hand, surveys are cost-effective but rely on self-report, which may be prone to bias. Participants are not always good at judging their own learning since they do not yet have accurate benchmarks (Kruger & Dunning, 1999). Self-report may also be affected by social desirability if participants feel pressure to answer a certain way (Desimone, 2009). However, when instructors report concrete behaviors without evaluative components, self-reports correspond well with observations (Desimone, 2009). Therefore, our instrument minimizes these self-report pitfalls by having teachers report on their use of concrete teaching practices, rather than subjectively judging their own knowledge or evaluating the quality with which they implement new techniques.

While evaluators should be concerned with the quality of implementation, this will likely improve over time and may require repeated measurements. Many professional development programs are funded through short-term grants. Time is spent developing and conducting the professional development, and therefore, not much time is left for participants to use and develop the new knowledge and skills in their own classrooms before evaluation must be conducted. As a result, professional development workshops may be deemed ineffective when in reality, skills may continue to develop and benefits may not be fully realized until well after the project has ended.

In this paper, we report on a self-report measure for IBL workshops that is designed to quickly and accurately detect the initial changes in teaching practices following professional development workshops. Our main research questions are:

1) Are cost-effective and efficient self-report measures of changes in teaching practices an accurate way to evaluate outcomes of professional development?
2) In what capacity can evaluation efforts assess the implementation of new knowledge and skills from professional development workshops within the short-term cycle of grant funding?

Research Methods

The study sites were four workshops for mathematics faculty, led by universities with IBL Mathematics Centers where an extensive menu of IBL courses had been developed and taught over several years. Thus, faculty with expertise on IBL were available to lead each workshop. Through funding from the National Science Foundation, the universities developed and implemented annual IBL workshops from 2010 to 2013 for four cohorts of math faculty new to IBL. Workshops spanned four or five days and included a mix of invited talks, open discussions, video observations, expert panels, hands-on exercises, and work time. Each workshop had a slightly different style; the 2010 and 2012 workshops were highly interactive, while the 2011 and 2013 workshops were more conference-like.

As evaluators for the workshop project, our team conducted pre- and post- workshop surveys at each workshop. We also conducted one-year follow-up surveys for the first three cohorts (2010 through 2012). All three surveys included both quantitative items and open-ended questions. Evaluation instruments addressed Levels 1-5 in Guskey's framework. Level 1 was assessed on post-workshop surveys where participants rated and commented on the quality of the workshop and logistics and the aspects they found most and least helpful. Level 2 was measured with Likert-scale items to reflect participants' knowledge, skills, and beliefs about inquiry teaching, as well as their motivation to use inquiry methods. By assessing these items before participants attended the workshop, immediately afterwards and again one year later, we could identify significant changes in their knowledge and perceptions. Participants also wrote definitions of IBL on each survey to reveal their current perception and level of understanding. To assess Level 3, participants rated the levels of support for IBL from their departments, their chairs, their colleagues, and their institutions. Participants also completed open-ended responses about ways they have and have not been supported in implementing IBL. On follow-up surveys, thirteen Likert-scale items and two open-ended items addressed student gains (Level 5) from IBL.

A large portion of the follow-up survey was aimed at Level 4 evaluation. In one item, participants reported if they had not implemented IBL techniques, implemented some IBL techniques, or implemented one or more fully-IBL courses. Teachers also rated the frequency with which they used eleven different teaching practices. Both pre-workshop and follow-up surveys included these eleven items, so comparisons allowed us to detect changes for each individual instructor. Some items described teaching behaviors consistent with IBL methods while other items were not. Open-ended questions collected data on the challenges and supports experienced in implementing IBL techniques in the first year after the workshop.

Results

Participants

In total, 167 participants attended the workshops. They came from a variety of institutions. Most taught at four-year colleges (37%), Ph.D.-granting research universities (37%), or master's-granting comprehensive universities (23%), and a small number taught at two-year colleges (4%). About 13% of participants taught at minority-serving institutions. Many were tenure-track faculty that were not yet tenured (35%); some were tenured (34%), and some were not tenure-track (27%). A small number were high school teachers (<1%) and graduate students (3%). The largest group had between 2 and 5 years of teaching experience (27%), while some had less than 2 years experience (20%), and many had more: 19% had 6-10 years experience, 18% had 11-20 years experience, and 16% had more than 20 years experience. A small number had experienced IBL classes as a student (25%) and almost half had some prior experience using IBL methods as an instructor (46%).

Most participants were male (58%), but the percentage of women (42%) was higher than among math faculty as a whole (National Science Foundation, 2008). Most participants identified as of European descent (74%), and a small percentage were of Asian descent (10%). These proportions are about the same as in U.S. math faculty as a whole (National Science Foundation, 2008).

While pre-workshop, post-workshop, and follow-up surveys were all collected anonymously, they were matched using two pieces of non-identifying individual information. Details about the numbers of surveys collected from each workshop are presented in Table 1.

Table 1. Survey response rates as a percentage of attendees.

Cohort	Attendees	Pre-Surveys	Post-Surveys	Matched Pre/Post	Follow-Up	Matched Pre/FU
2010	42	37 (88%)	41 (98%)	33 (79%)	31 (74%)	23 (55%)
2011	55	47 (85%)	43 (78%)	29 (52%)	36 (65%)	21 (38%)
2012	42	40 (93%)	41 (95%)	38 (88%)	29 (69%)	25 (60%)
2013	28	22 (79%)	26 (93%)	21 (75%)	*TBD, Fall 2014*	
Total	167	146 (87%)	151 (90%)	121 (72%)	96 (69%)	69 (50%)

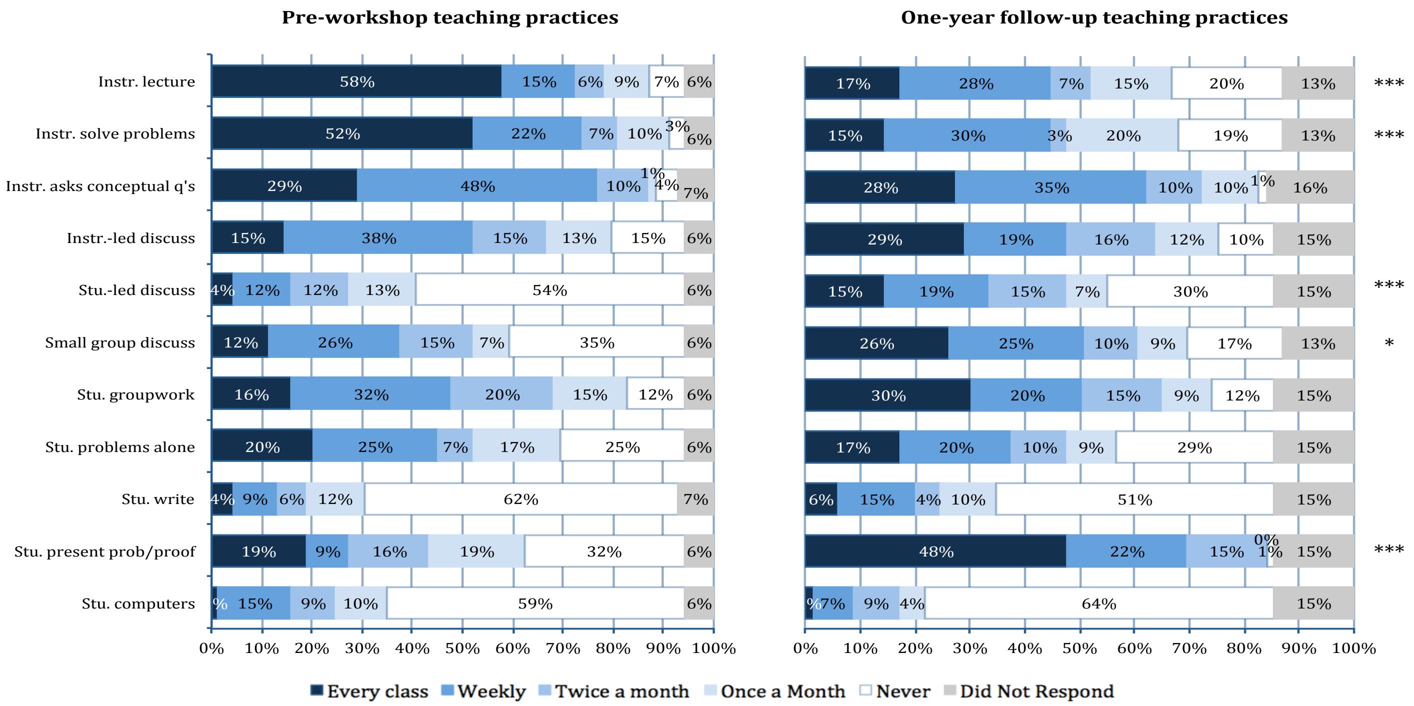
Teaching Practices: Changes in Practices

Participants reported teaching practices on both pre-workshop surveys and follow-up surveys. For each specific practice, respondents indicated on a 5-point scale whether they did it in 'every class' (5), 'weekly' (4), 'Twice a month' (3), 'Once a month' (2), or 'Never' (1). On follow-up surveys for the first three cohorts, 69% of participants responded, of which 69 (50% of attendees) supplied matching pre-workshop and follow-up surveys. We used Wilcoxon Signed Rank tests to measure changes in the ordinal data.

Teachers reported changes in practices consistent with IBL teaching. Overall, participants reported significant decreases in the frequencies of *Instructors lecturing,* and *Instructors solving problems at the board.* They also reported significant increases in *Student-led whole class discussions, Small group discussions,* and *Students presenting problems or proofs.*

There were no significant differences in practices that are not specific to IBL methods, including *Instructors asking conceptual questions, Instructor-led whole class discussions, Students solving problems individually, Students writing individually, or Computer assisted learning.* These non-IBL items were added to detect response biases; non-significant changes suggest that instructors are using the full scales and are reporting honestly. Indeed, in our other evaluation of professional developments, we have found that college instructors tend to be more self-critical on surveys and use the full range of scales, whereas K-12 teachers tend to use just the extreme answers (Hayward, Laursen, & Thiry, 2013). Additionally, *Student collaborative work in small groups* is characteristic of some types of IBL teaching. While participants did report increased use of this strategy, the difference was outside the range of statistical significance (p=0.086).Teaching practices on pre-workshop surveys and one-year follow-up surveys are compared below in Figure 1.

Figure 1. Self-reported teaching practices.

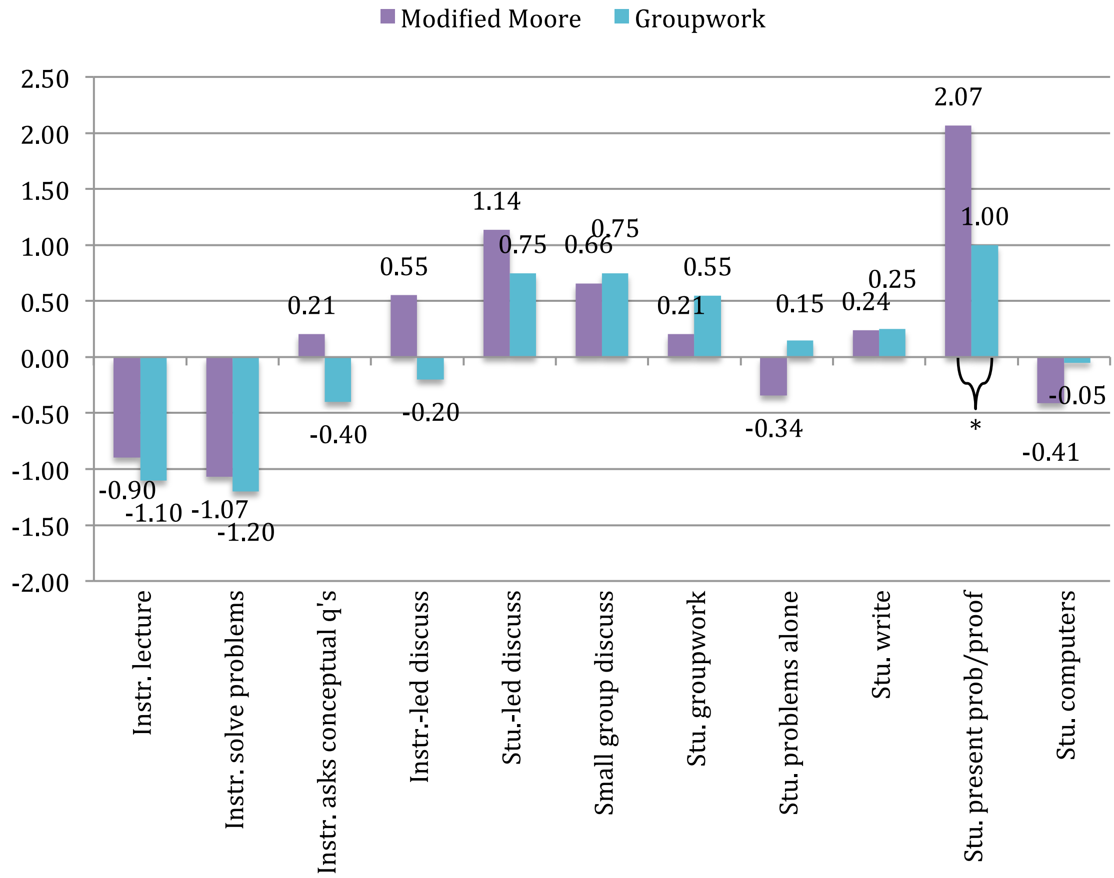


$^{*}p$<0.05, $^{**}p$<0.01, $^{***}p$<0.001

Teaching practices data were also analyzed for differences between two different types of IBL presentations. All participants in the 2011 workshop and about half of the participants in the 2012 workshop were presented with a groupwork-centered version of IBL (n=20 respondents), while all participants at the 2010 workshop and the other half of participants in the 2012 workshop were presented with a Modified Moore-method approach to IBL that uses individual student presentation of proofs or solutions followed by class discussions (n= 29 respondents).

When slicing up the data this much, sample sizes were too small to detect significant differences for most items, but the data suggest that participants presented with the Modified Moore method reported greater increases in the frequencies of having students present problems or proofs compared to participants presented with a groupwork-centered version of IBL. These changes are detailed in Figure 2.

Figure 2. Changes in teaching practices by cohort.



*p<0.05, **p<0.01, ***p<0.001

Taken together, these findings suggest that our self-report instrument may be sensitive to differences in workshops. Practices that are common to both versions of IBL, including decreases in instructors lecturing and solving problems, showed significant changes regardless of the type of IBL presented. However, participants who were presented with a Modified Moore Method reported increases that were statistically greater than those participants presented with groupwork-centered IBL.

Teaching Practices: Accuracy of Self-Report

In order to assess accuracy of self-reported implementation of IBL techniques, two approaches to probing participants' use of IBL were compared. Participants reported in a single item on the follow-up survey whether they had not implemented IBL, implemented some elements of IBL, or fully implemented IBL in one or more courses. These were converted to numerical answers of 1, 2, or 3, respectively. This single item was compared to the eleven separate items on specific teaching practices.

The results showed significant negative correlations between implementation level and frequencies of *Instructors lecturing* (r = -0.62, p < 0.001), *Instructors solving problems at the board* (r = -0.49, p < .001), *Instructors asking conceptual questions* (r = -0.46, p < 0.001), and *Students solving problems individually* (r = -0.53, p < 0.001). The only significant positive correlation was with *Students presenting problems or proofs* (r = 0.39, p < 0.01).

These features are consistent with a definition of IBL aligned with that of its founder, R.L. Moore, in which the instructor takes a less active role and class time is largely composed of students presenting and critiquing proofs that they have worked on before class (Jones, 1977). So, while participants seem to have self-identified their one-year follow-up teaching practices by using the Moore Method as a benchmark for "real IBL," the changes in their teaching practices in comparison to pre-workshop surveys encompass a broader definition of IBL inclusive of both whole-class and small group discussions. Since instructors have reported teaching practices different than those they consider to be "real IBL," this suggests that participants are providing honest responses rather than the socially desirable responses.

Implications for Future Research

Initial results suggest that in addition to being cost-effective and efficient, this self-report measure of changes in teaching practice shows promising indicators of accuracy. Changes from pre-workshop surveys to one-year follow-up surveys were consistent with the teaching practices presented at the IBL workshops. Trends in data suggest that self-report measures may also be sensitive to the type of IBL presented. Correlations on one-year follow-up surveys between self-reported implementation levels of IBL and teaching practices indicate that participants largely consider the Moore Method to be the "real" IBL. So, participants are reporting changes in their teaching practices in line with those presented at the workshop, but they do not necessarily identify themselves as doing "real IBL." One of the main critiques of the accuracy of self-report measures is that participants often report only socially desirable answers (Desimone, 2009). However, these nuanced differences indicate that for this measure, instructors may be accurately reporting teaching practices that are not consistent with the socially desirable definition of "real IBL." These surveys will continue to be used to evaluate future workshops, and increased sample sizes should provide greater statistical power. Additionally, we have received NSF funding to formally validate the survey instrument through comparisons with classroom observations.

This measure was simple and efficient to administer. While we did not address the quality of implementation, we were able to measure the extent of change in practice following the professional development workshops in a timely and cost-effective manner. This self-report tool is especially useful given the short time frame and tight budgets of many grant-funded PD projects. While this particular survey is best suited for IBL workshops, it could easily be adapted to professional development workshops on other topics or in other disciplines by changing the target instructional practices.

References

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). *How people learn: Brain, mind, experience, and school.* Washington D.C.: National Academy Press.

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Education Researcher , 38* (3), 181-199.

Deslauriers, L., Schelew, E., & Wieman, C. (2011). Improved learning in a large-enrollment physics class. *Science , 332* (6031), 862-864.

Fairweather, J. (2008). *Linking evidence and promising practices in science, technology, engineering, and mathematics (STEM) undergraduate education: A status report for the National Academies Research Council Board of Science Education.* Retrieved April 23, 2011, from http://www7.nationalacademies.org/bose/Fairweather_CommissionedPaper.pdf

Guskey, T. R. (2000). *Evaluating professional development.* Thousand Oaks, CA: Crowin Press.

Guskey, T. R. (2002). Professional development and teacher change. *Teachers and Teaching: Theory and Practice , 8* (3/4), 381-391.

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics , 66* (1), 64-74.

Hayward, C., Laursen, S., & Thiry, H. (2013). *Survey of teacher participants in professional development workshops conducted by the Biological Sciences Initiative, 2010 – 2012: Final report.* Report to the Biological Sciences Initiative. Boulder, CO: Ethnography & Evaluation Research.

Henderson, C., & Dancy, M. H. (2007). Barriers to the use of research-based instructional strategies: The influence of both individual and situational characteristics. *Physical Review Special Topics - Physics Education Reform , 3* (2), 0201102.

Henderson, C., & Dancy, M. H. (2011). *Increasing the impact and diffusion of STEM education innovations.* Retrieved July 1, 2013, from Comissioned paper for Forum on Characterizing the Impace and Diffusion of Engineering Education Innovations: http://www.nae.edu/File.aspx?id=36304

Henderson, C., & Dancy, M. H. (2008). Physics faculty and educational researchers: Divergent expectations as barriers to the diffusion of innovations. *American Journal of Physics , 76*, 79-91.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force Concept Inventory. *The Physics Teacher , 30* (March), 141-158.

Hora, M. T., & Ferrare, J. J. (2012). *A review of classroom observation techniques used in postsecondary settings.* White paper. Prepared for the Measurement of Teaching Practices in Undergraduate STEM workshop hosted by AAAS/NSF.

Jones, F. B. (1977). The Moore method. *The American Mathematical Monthly , 84* (4), 273-278.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology , 77* (6), 1121-1134.

Kuh, G. (2008). *High-impace educational practices: What they are, who has access to them, and why they matter.* Washington D.C.: AAC&U.

Libarkin, J., & Anserson, S. W. (2005). Assessment of learning in entry-level geoscience courses: Results from the Geoscience Concept Inventory. *Journal of Geoscience Education , 53*, 394-401.

Mahavier, W. S. (1999). What is the Moore Method? *Problems, Resources, and Issues in Mathematics Undergraduate Studies , 9* (4), 339-354.

McCann, T. M., Johannessen, L. R., Kahn, R., & Smagorinsky, P. (Eds.). (2004). *Reflective teaching, reflective learning: How to develop critically engaged readers, writers, and spearkers.* Portsmouth, NH: Heinemann.

Moon, J. A. (2004). *A handbook of reflective and experiential learning: Theory and practice.* New York: RoutledgeFalmer.

National Science Foundation. (2008). *TABLE 3. Employed doctoral scientists and engineers in 4-year educational institutions, by broad field of doctorate, sex, faculty rank, and years since doctorate: 2008*. Retrieved July 29, 2013, from Characteristics of Doctoral Scientists and Engineers in the United States: 2008: http://www.nsf.gov/statistics/nsf13302/pdf/tab3.pdf

National Science Foundation. (2008). *TABLE 4. Employed doctoral scientists and engineers, by selected demographic characteristics and broad field of doctorate: 2008*. Retrieved July 29, 2013, from Characteristics of Doctoral Scientists and Engineers in the United States: 2008: http://www.nsf.gov/statistics/nsf13302/pdf/tab4.pdf

PCAST. (2012). *Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics.* Executive Office of the President.

Prince, M. (2004). Does active learning work? A review of the research. *Journal of Engineering Education , 93* (3), 223-231.

Prince, M., & Felder, R. (2007). The many facets of inductive teaching and learning. *Journal of College Science Teaching , 36* (5), 14-20.

Retsek, D. Q. (2013). Chop wood, carry water, use definitions: Survival lessons of an IBL rookie. *Problems, Resources, and Issues in Mathematics Undergraduate Studies , 23* (2), 173-192.

Ruiz-Primo, M. A., Briggs, D., Iverson, H., Talbot, R., & Shepard, L. A. (2011). Impace of undergraduate science course innovations on learning. *Science , 331* (6022), 1269-1270.

Savin-Baden, M., & Major, C. H. (2004). *Foundation of problem-based learning.* Maidenhead, UK: Open University Press.

Schumacher, C. E. (2010, June 29). *Instructor's resource manual for use with Chapter Zero - Fundamental Notions of Abstract Mathemeatics, 2e.* Retrieved August 7, 2013, from http://www2.kenyon.edu/Depts/Math/schumacherc/public_html/Professional/Research/Zero/guide.pdf

Springer, L., Stanne, M. E., & Donovan, S. S. (1999). Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis. *Review of Educational Research , 69* (1), 21-51.

Yoshinobu, S., & Jones, M. (2013). An overview of inquiry-based learning in mathematics. *Wiley Encyclopedia of Operations Research and Management Science* , 1-11.