The Evolution of Responsive Education Program Evaluation at the National Science Foundation Engineering Research Center for Biorenewable Chemicals Research Experience for Teachers

Karri M. Haen[a] and Mari Kemis[b]

[a] *Department of Genetics, Development and Cell Biology, Iowa State University, Ames, IA 50011, United States*
[b] *School of Education, Research Institute for Studies in Education (RISE), Iowa State University, Ames, IA 50011, United States*

Email: Karri Haen, khaen@iastate.edu; Mari Kemis, mrkemis@iastate.edu

Abstract

The Iowa State University Research Institute for Studies in Education (RISE) has managed the assessment of the NSF Engineering Research Center for Biorenewable Chemicals (CBiRC) education programs for the center's five project years. A longitudinal study of the CBiRC Research Experience for Teachers (RET) highlights the evolution of the overall evaluation design, where the essential feature of the approach is responsiveness to key issues or methodological problems developed throughout the maturation of the teacher education program. The continual adaptation of the assessment's goals and data collection for the RET emphasizes the value of a highly integrated, yet flexible, evaluation framework, where the assessment rigorously generates and seeks to answer basic research hypotheses about the program's components. Research-based measures of programmatic success have stimulated continued innovation in the design features of both the evaluation and the education program. Here, we discuss an example of a responsive evaluation where an innovative methodology for addressing problems associated with evaluating outcomes of diverse, research-based professional development programs was utilized. The method discussed includes an authentic experimental design task that circumvents the difficulties associated with estimating increased science research skills with self-efficacy questions and which may allow for greater flexibility in determining longitudinal gains in teachers' science process knowledge. This research illustrates the value of responsive evaluation for NSF-funded education programs.

Evaluation Background

Despite the body of literature on the merits of responsive evaluation comprised since the 1970's, there has been a recent surge in interest in the standardization of evaluation instruments for NSF engineering education programs, including the highly diverse Research Experience for Teachers (RET) [1] Program. Unlike *preordinate* evaluation (formalized evaluation based only on measurements of pre-specified program objectives), responsive evaluation follows the primary components of the instructional program, where "the choice of tests and other data-gathering devices is made based on observation of the program in action and interaction with various interested groups" (Stake, 1972). Stake suggested a responsive evaluation would be particularly important during the formative phases of program implementation, where project staff might be unaware of how problems could arise (ibid). The current study addresses the extended value of a research-guided responsive evaluation for both the initiation and continuing innovation of the CBiRC RET program.

CBiRC Research Experience for Teachers Program

The mission of the Iowa State University Center for Biorenewable Chemicals Research Experience for Teachers Program is to improve the quality of secondary level education in science, technology, engineering and mathematics (STEM), particularly in Iowa public schools. The CBiRC RET program has three interconnected objectives: to alter high school teachers' 1) teaching philosophy, 2) pedagogical technique, and 3) STEM content/process knowledge. The CBiRC RET provides high school teachers with first-hand experiences in the design, methods, and analysis of research associated with biorenewable chemicals engineered for the purpose of clean bio-based energy resources. Relationships built in the CBiRC RET enable teachers to understand and communicate the latest developments in STEM fields, inspiring student enthusiasm for higher education and career tracks in science and engineering.

CBiRC's seven-week RET began in June 2009 and has since completed four years of the summer research program. Teachers conduct small independent research projects in one of the three research thrust areas under the mentorship of CBiRC associated faculty. Research lab specialty, and, thus, teacher research projects, range from the discovery of biocatalysts and microbial engineering (Thrust 1), or the chemical engineering of biochemical catalysts (Thrust 2), to life cycle assessments of carbon renewability and techno-economic feasibility (Thrust 3). The three research thrusts are highly interactive and research groups work in multiple disciplines simultaneously (Fig 1). Prior to work in the research lab, teachers attend a short training workshop that includes laboratory safety procedures, use of basic laboratory equipment, and data collection technique. Teachers also participate in an intensive workshop on a selected pedagogical topic, which is intended to encourage the translation of the research experience to the classroom.

[1] Based upon panel discussions at NSF Engineering Education and Centers conference, March 2012, Washington D.C.

Figure 1. Scientific and engineering disciplines involved in CBiRC research. Teachers participating in a CBiRC RET may work on a variety of interdisciplinary research problems, all with the goal of producing environmentally and economically sustainable biochemicals and clean bio-based energy resources. Hexagons in the figure represent the potential disciplinary areas in which teachers may gain experience.

Initial Evaluation of the CBiRC RET: Project Years 1-3

The standard CBiRC RET evaluation instruments have aimed to obtain information from two basic categories: 1) outcomes and impacts of the program (specified in accordance with the three CBiRC RET overarching objectives) and 2) RET program administrative issues and obstacles (Fig. 2). For the first three years of the CBiRC RET evaluation, teachers participated in online pre- and post-program surveys, which were largely based upon the Likert ranking of self-efficacy questions for a range of research topics. Additionally, teachers participated in a formative assessment in the form of six weekly reflections that report teachers' thoughts about their progression through the RET program. Each year the program ends in a summative focus group, which asks teachers to comment on their laboratory and pedagogical experiences, including plans for implementing their summer experience into their high school science or engineering curricula. Early in the second semester of the following academic year, a follow-up survey is administered to find out how the teachers have applied what they learned during the program in their classrooms. Finally, RET teachers' laboratory mentors respond to a short mentor survey to determine mentors' perceptions of the teachers' performance during the program.

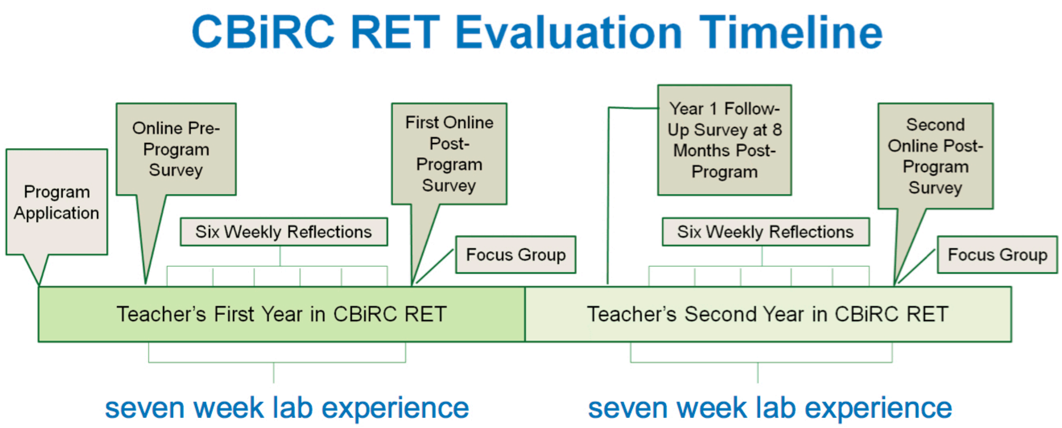

Figure 2. Standard evaluation timeline for the first two years of a CBiRC RET experience. During CBiRC project years 1-3 teachers participated in multiple formative and summative assessments throughout the RET program. These assessments included pre- and post-program surveys, weekly reflective journaling, a focus group and an eight-month follow-up survey. If a teacher elected to participate in a second year of the RET program, they would continue to participate in the same evaluation with exception of the baseline pre-survey. The timeline was modified to include an additional assessment in project year 4.

Revision to the RET Assessment & Evaluating Science Process Skills

The CBiRC RET assessment has evolved substantially since its first iteration in 2009, largely because the RET program is unusual in that it allows teachers to return for successive summers to work in CBiRC faculty research laboratories. Since responses to Likert-rated survey questions may only vary weakly with time, scaled-option questions for repeat participants had low resolving power for determining changes in teachers' perceptions in the longitudinal evaluation (see Fig. 3 for example). Self-efficacy questions were particularly problematic for the RET assessment, as first-year participants reported highly significant gains from the pre-program to end of the first year, but perceived gains in learning generally leveled off for second-year participants despite an abundance of qualitative data to the contrary (data not shown). Because we suspected the longitudinal data collected from the pre- and post-surveys did not adequately represent the reality of second year RET experiences, we sought additional means to quantify program-related changes in science skills.

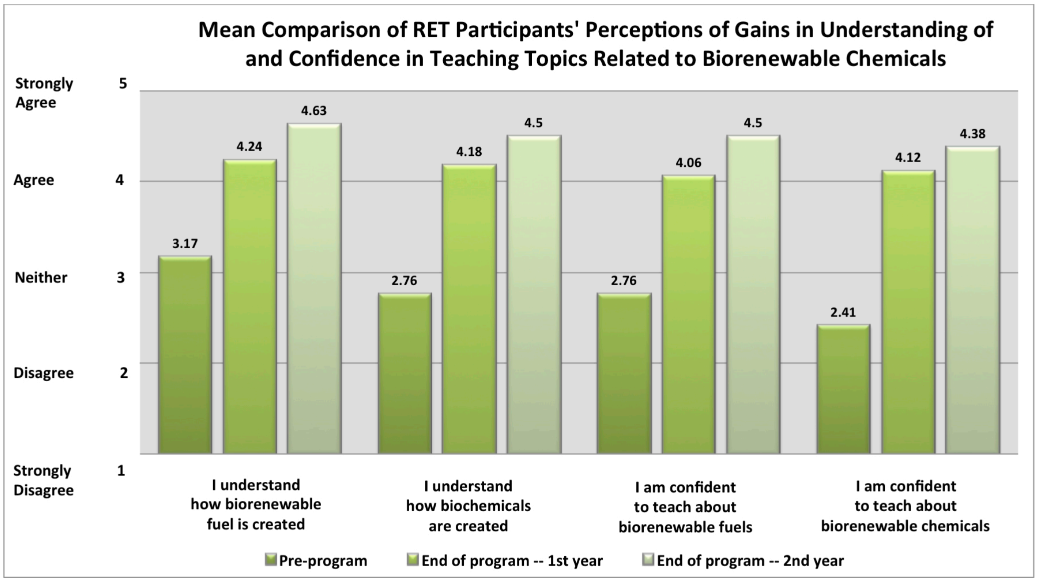

Figure 3. Mean comparison of RET participants' perceptions of gains in understanding and confidence in teaching topics related to biorenewable chemicals. This example is derived from data collected for RETs from 2009-2012 and shows gains in the understanding of, and confidence in teaching about, biorenewable chemcials are perceived to be significantly greater in the first year of the RET experience than in the second year.

Upon review of the breadth of potential teacher laboratory research experiences during the CBiRC RET, the evaluators realized there was no straight-forward way to implement a content knowledge test that would be valid for all teachers. Thus, a methodology for evaluating science process skills was considered. Science process skills are defined as "a set of broadly transferable abilities, appropriate to many science disciplines and reflective of the behavior of scientists" (Padilla, 1990). The immediate goals of research-based science professional development programs such as RETs emphasize the refinement of higher-order or "integrated" science process skills, including learning or re-learning how to formulate hypotheses, conduct experiments, control variables, and interpret scientific data (ibid). Immersion in real-world science and engineering research is thought to encourage the growth of this procedural knowledge, which may then be passed on to students in the classroom (Driscoll, 2005).

Several studies have looked at the effect of participation in science professional development programs on teachers' self-efficacy (Lumpe et al., 2000), motivation (Pop et al. 2012) and even the impact of RET programs on student achievement (Dubner et al. 2009); however, upon searching the current education and science literature, we were unable to find publications that provided quantitative data directly measuring science process skill gains from participating in RETs. Kishbaugh et al. (2012) found there are a "paucity of options [for measuring science process skills] in the published literature, and in particular [there are] few options for rubrics that are designed towards student learning of broader goals of scientific literacy in a variety of assignments."

Recently published rubrics for assessing science process skills are generally based upon Bloom's taxonomy of learning objectives (1956). These include, but are not limited to, one extensive rubric bank for assessing research assignments in the sciences (Kishbaugh et al. 2012) and various examples of research skill assessment tasks from the University of Adelaide's Research Skill Development for Curriculum Design and Assessment (RSD) (http://www.adelaide.edu.au/rsd/).

To compliment the standard assessment (Fig. 2), we have developed a novel rubric-guided scientific experimental research design assessment, for which the RET site administrator and the project evaluators successfully piloted a study during the summer of 2012. The new assessment required teachers to derive an analysis of an authentic research problem and determined gains in teachers' science process skills by rating teachers' understanding of three aspects of research problems: 1) defining a hypothesis, 2) experimental design, and 3) deriving a data analysis plan. Eighteen RET teachers participated in the study, and their combined scores across the assessment topics showed statistically significant improvement in performance on the experimental design task after participating in the RET program.

Participants

Eighteen 2012 RET teachers participated in this study. Of these, nine were participating in a RET for the first time, and nine had previously participated in the CBiRC RET at least once. Of the first-year teachers, two were female and seven were male. They had been teaching an average of 3.33 years, ranging from one to five years. Most taught high school science subjects including chemistry, biology, earth science, physics, and physical science, and two teachers taught pre-engineering. Seven reported that they had bachelor's degrees in science subjects. The second-year teachers included three females and six males, with an average of slightly over four years of teaching experience in 2012. Like the first-year teachers, they taught a variety of high school science subjects like biology, physical science, chemistry, advanced biology, and physics. Four had bachelor's degrees in biology, two in the environmental sciences, and one in zoology. Two teachers did not indicate their degree. One teacher also had a master's degree in secondary science education with a biology focus.

Methods

Pre- and Post-Program Surveys. Although we do not present additional data from these surveys beyond Figure 3, we describe the methodology here since this portion of the evaluation was critical to changes in the overall responsive evaluation. RET participants were asked to take a pre-survey at the beginning of the program. This survey asked for demographic data, educational and teaching histories, attitudes about teaching, methodologies used in the classroom, familiarity with laboratory techniques from multiple disciplines, professional development activities, and the general confidence level teachers have about teaching STEM topics. The survey consisted of 116 quantitative and short-answer questions. Quantitative questions were rated on a Likert-based scale, and included rated whether the respondent agreed/disagreed with statements about science education, was familiar/unfamiliar with laboratory concepts, felt methods were important/unimportant to teaching and learning, or felt comfortable/uncomfortable or prepared/unprepared to teach STEM concepts. Survey data collection was conducted online using Qualtrics survey software. E-mail notifications included a brief outline of the study objectives, confidentiality information, and a link to the Qualtrics website. Two reminders were sent to non-respondents. A post-survey with the same items was sent to RET participants during week seven of the program, using the same methodology. All methods described for the CBiRC RET evaluation were approved by the Iowa State University Internal Review Board (IRB) and were carried out with the written consent of the participants.

Experimental Design Task. The experimental design task was intended to measure teachers' science process knowledge while attempting to control for content knowledge-related bias due to teachers' differential research lab experiences. Thus, we selected research problems that were in no way related to the research RET teachers conducted in any of their laboratories, and that only required a basic understanding of the design of experiments. We searched recent scientific abstracts for qualified research problems and selected five possible publications upon which the pre- and post experimental design tasks would be based. A panel of two scientists (the first author and a doctoral candidate) utilized the experimental design rubric (described below) to rate the difficulty of each research problem. The two research problems that were rated the most similar by the panel were used as the basis for the pre- and post-program experimental design tasks. For the pre- and post-tests, teachers received a brief description of an authentic research problem and were asked to develop a hypothesis, design an experiment, and write about how they would analyze the data they proposed to collect. The 2012 pre- and post-tests were based upon the research problems from Mitro et al. (2012) and Wisemen et al. (2012), respectively. At the end of each test, teachers were asked whether they had seen or heard about the research publications used to derive the test problems.

Experimental Design Rubric. We used a similar strategy to Shadle et al. (2012) in the design of the rubric, but the criteria and descriptors for the 4-level analysis were derived particularly for the grading of a generalized experimental design problem. The rubric rated each teacher's experimental design task results on three experimental areas: 1) hypothesis development, 2) experimental design description and 3) data analysis plan, where each response to a topic was categorized as "emerging" (levels 1-2), "developing" (levels 2-3) or "mastering" (levels 3-4).

Teachers participated in the experimental design task on the first and last days of the RET program. The tests were passcode protected and given electronically in a proctored room. Teachers finished each task within 45 minutes. Eighteen experimental design task pre- and post-tests were coded with subject identifiers and graded by the first author (an expert in experimental design) according to the established rubric. After completion, subject identifiers were used to arrange scores into pre- and post-test responses for each teacher, and paired t-tests were used to determine statistical significance.

Results and Discussion

Due to difficulties with tracking repeat-year RET teachers' science skills growth with standard Likert-ranked self-efficacy survey questions (Fig. 3), and increased interest of NSF CBiRC site visit team members in higher fidelity longitudinal quantitative data tracking for RETs, the evaluation team designed a new method to determine changes in teachers' science process knowledge using authentic research problems from the current scientific literature. Research topics chosen for the assessment were determined to be outside the realm of expertise of RET-hosting laboratories and required only basic science process skills (not discipline specific knowledge) in order to formulate a potentially successful experimental design.

Data were collected from the pre- and post-program experimental design task for 18 RET teachers. Teachers made significant gains in their science process skills in all areas tested (Fig. 4), including the ability to formulate a correct scientific hypothesis that identified important aspects of the research problem (pre=2.83, post=3.72, $p<0.001$); ability to formulate an experimental design that appropriately described instruments and procedures with reasonable evidence of validity (pre=2.85, post=3.29, $p<0.05$); and the ability to describe data analysis procedures appropriate for the data to be collected (pre=2.17, post=2.61, $p<0.05$).

Overall, teachers made the largest gains in their ability to formulate a correct scientific hypothesis, enhancing their skills on the assessment 31.45% by the end of the program. For instance, in the pre-program test, five teachers were either unable to identify a research problem and produced hypotheses that addressed questions outside of the presented problem, or they produced hypotheses that were too general to guide the design of an adequate experiment. All of the teachers performed at the developing/mastering or mastering level by the end of the RET (a score of 3 or 4). Additionally, teachers made a 15.44% gain in their ability to derive a sound experimental design to address a research problem in the post-test. Although determination of a sound experimental design is a complex process that requires multiple scientific skills, it is perhaps not surprising that teachers performed better on the post-test since they all addressed the research problem in their hypotheses on the post-test. Identification of the correct research problem potentially led teachers to derive less complicated experimental design proposals. Finally, teachers made a 20.28% gain in their ability to derive a valid data analysis plan for the proposed experiment. On the post-test, four teachers provided the details of specific statistical analyses for interpreting data, but only two did this on the pre-test (achieving mastery level).

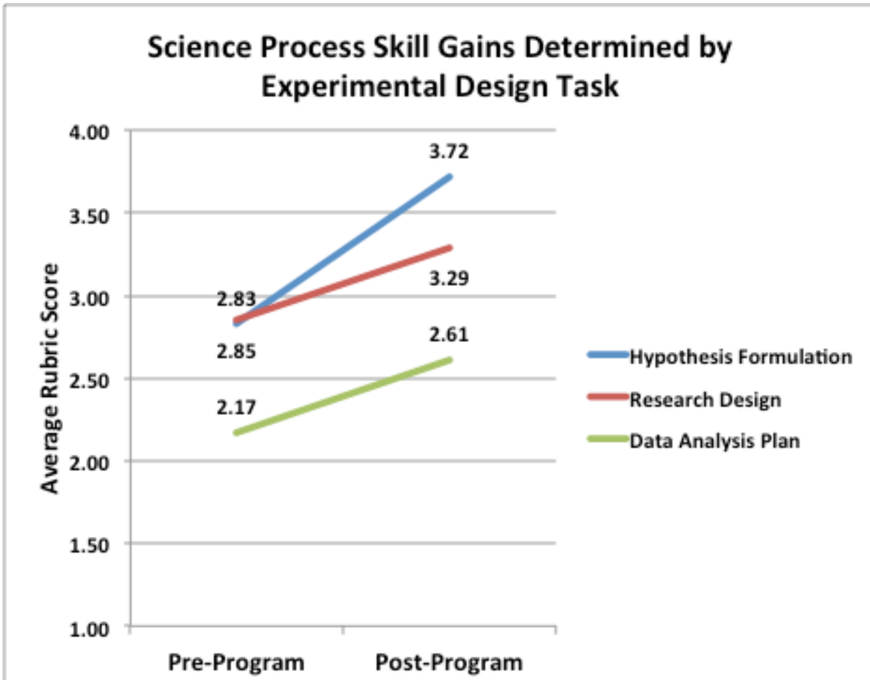

Figure 4. Science skill gains determined by the experimental design task and rubric for all 18 teachers. Gains in the ability to formulate a hypothesis (blue line, $p<0.001$); ability to formulate an experimental design (red line, $p<0.05$); and the ability to describe data analysis procedures (green line, $p<0.05$).

We additionally analyzed the experimental design assessment data by years of participation in the RET. In this analysis, nine teachers were in the RET for the first time and nine had participated in a prior research RET. Although we did not expect to find significant differences between the two groups due to the low population size (and other potential threats to validity), a significant difference was detected between the first-year RET teachers and the second-year RET teachers in the area of deriving a data analysis plan ($p<0.05$) (Fig. 5). This finding potentially indicates that an additional seven weeks of experience in a RET laboratory significantly alters teachers' ability to use statistics to interpret scientific data.

First-year teachers made significant pre- to post-test gains on the hypothesis formulation and experiment design portions of the assessment ($p<0.05$), while second-year RET teachers performed significantly better on the hypothesis formulation and data analysis plan portions of the post-test assessment ($p<0.05$).

Interestingly, the second-year RET teachers performed nearly equivalently on the experiment design portion of the assessment in the pre- and post-test (pre=3.06, post=3.17). Because the design of experiments is a complex process, it may be that teachers require much more time in the laboratory than an additional seven weeks to improve their performance on this component of the experimental design task.

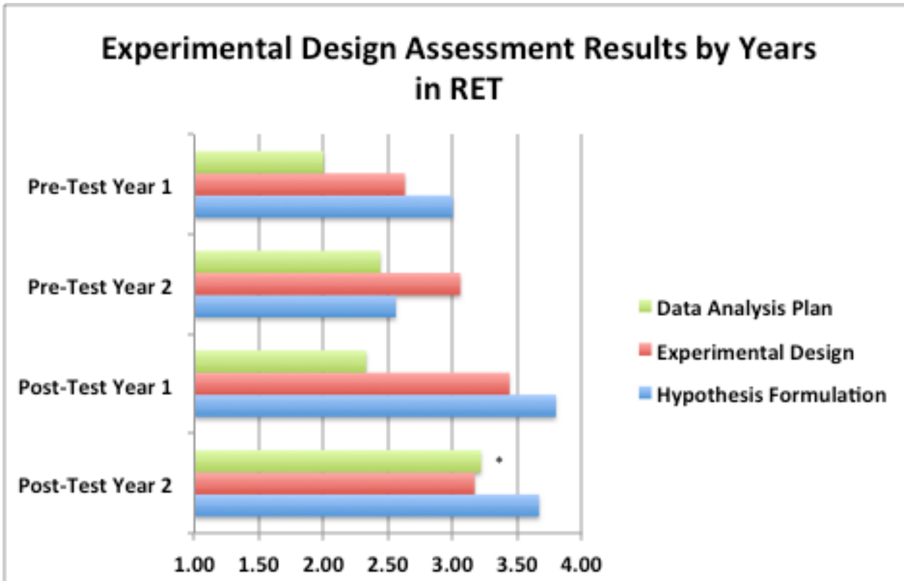

Figure 5. Experimental design assessment results broken down by number of years in RET. Nine teachers had previous RET experiences, while nine teachers participated in a research RET for the first time. The only significant difference between the first- and second-year teachers' post-test performance on the test was in the data analysis plan component (p=0.04). Overall, first- and second-year teachers increased their performance on the hypothesis development portion, while first-year teachers significantly improved in the area of experiment design and second-year teachers significantly improved in formulating a data analysis plan ($p<0.05$).

Conclusions and Future Directions

In this research we have created a science process skill assessment we have termed an experimental design task. The task uses authentic research from the scientific literature to derive research problems that are used as the basis for analyzing three test criteria that can be graded with a specialized rubric. This preliminary study shows the experimental design task can be used to measure change in science process skills, and may potentially detect differences in the performance of RET teachers who have participated in a RET program for multiple years versus only one year. Such results can be useful for not only determining program impacts, but can be used to guide the further development of RET programs by allowing the evaluation team to make better recommendations regarding the integrity of basic research skills that teachers gain during the program.

We caution that this is only a preliminary study with a relatively small number of teacher participants. In the next project year of this analysis, greater accuracy will be achieved by the use of multiple graders to score data for the experimental design task resulting in interrater reliability statistics. Additionally, it is currently unclear what the relationship is between teacher self-efficacy statements about science process skills and teachers' performance on the experimental design task. We will pursue this question by adding science process skills self-efficacy questions to the standard pre- and post-program surveys for the 2013 CBiRC RET. Finally, because creativity is an innate characteristic of the experimental design process, we plan to investigate teachers' understanding of the need for creativity in science and their performance in the design of experiments.

Acknowledgments: The evaluation team would like to thank the CBiRC Pre-College Education Director, Adah Leshem and CBiRC RET lead teacher Eric Hall for their help in administering the experimental research design assessment and the CBiRC University Education Programs director, Raj Raman, for helpful conversations regarding the overall CBiRC Center evaluation.

This material is based upon work supported by the National Science Foundation under Award No. EEC-0813570.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Citations

B.S. Bloom. (1956). Taxonomy of Educational Objectives, Handbook I: The Cognitive Domain. New York: David McKay Co Inc.

M.P. Driscoll (2005). Psychology of learning for instruction, 3rd ed. Boston MA: Pearson, Ed. Inc.

Tara L. S. Kishbaugh, Stephen Cessna, S. Jeanne Horst, Lori Leaman, Toni Flanagan, Doug Graber Neufeld and Matthew Siderhurst. (2012). Measuring beyond content: a rubric bank for assessing skills in authentic research assignments in the sciences. Chem. Educ. Res. Pract., 2012, 13, 268–276

Andrew T. Lumpe, Jodi J. Haney, and Charlene M. Czerniak. (2000). Assessing Teachers' Beliefs about Their Science Teaching Context. Journal of Research in Science Teaching. V. 37(3) pp. 275–292

Susanna Mitro, Amy R. Gordon, Mats J. Olsson, and Johan N. Lundström. The smell of age: Perception and discrimination of body odors of different ages. PLoS ONE, 7(5): e38110+, May 2012.

Michael J. Padilla. 1990. The Science Process Skills. National Association for Research in Science Teaching.

T.J. Posnanski. (2002). Professional development programs for elementary science teachers: An analysis of teacher self-efficacy beliefs and a professional development model. Journal of Science Teacher Education, 13(3), 189-220.

Susan E. Shadle, Eric C. Brown, Marcy H. Towns, and Don L. Warner. (2012). J. Chem. Educ., 89 (3), 319–325

S.C. Silverstein, Dubner, J., Miller, J. Glied, S., Loike, J.D. (2009). Teacher Research Program Participation Improves Their Students' Achievement in Science. Science 326: 440-442.

Robert Stake. 1972. Responsive Evaluation.
http://www.eric.ed.gov/contentdelivery/servet/ERICServlet?accno=ED075487
WWW resource last accessed on 10/22/12.

Richard Wiseman, Caroline Watt, Leanne Ten Brinke, Stephen Porter, Sara-Louise L. Couper, and Calum Rankin. The eyes don't have it: lie detection and neuro-linguistic programming. PloS one, 7(7), 2012.

Zembal-Saul, et. al. (2002) Web-Based Portfolios: A Vehicle for Examining Prospective Elementary Teachers' Developing Understandings of Teaching Science. Journal of Science Teacher Education, 13(4): 283-302.