ERC BEST PRACTICES MANUAL
CHAPTER 4: EDUCATION PROGRAMS

4.6 ASSESSMENT AND EVALUATION

4.6.1 Rationale and Definitions for Assessment and Evaluation

NSF recognizes the importance of assessing the impact of all ERC University and Precollege Education programs and the General Outreach to involve precollege students in the ERC activities that it supports. Accordingly, in the 2015 ERC solicitation it was required that ERCs assess, evaluate, and track the impacts of educational and outreach programs on program participants; requirements in this area will be specified in each class' solicitation. Having an assessment and evaluation plan in place not only ensures that the center meets NSF requirements, but it is also key in determining which education programs are helping the center meets its mission and which should be modified or terminated. It is the best way to gather data that can direct the development of effective programs, and goes beyond the use of anecdotal information about program satisfaction towards a more data-driven approach for assessment and evaluation of impact.

As a starting point, it is important to define assessment and evaluation, because although often used interchangeably, the two terms define different processes. Contemporary definitions vary as a function of the context in which the assessment or evaluation occurs and in terms of the assessed content. In general, *evaluation* refers to a summative judgment of worth, merit, and value at the end of a project, while *assessment* is more formative (occurring as the project progresses) and guides improvement over time. Specific to the ERCs, assessment should guide continuous improvement of the ERC's University and Precollege Education programs and the General Outreach activities involving precollege students, while measuring the programs' impacts over time. NSF ERC Program-level evaluations of educational impact are carried out by the ERC Program.

Program evaluation and assessment are not just the evaluator's or assessment officer's responsibility. Education Directors also must understand the process, design, and content of program assessment and evaluation and how to present the results effectively to inform all involved. Additionally, the center's full Leadership Team should be included in assessment and evaluation efforts. Regular communication of efforts and results is recommended.

ERC impacts are most often related to college and career trajectories in engineering and related fields. ERCs contribute to both industry and academia via their precollege and university student alumni. Assessment is an important way to demonstrate this impact. In the sections that follow and in the appendix to this section, general guidelines, key features, processes, procedures, and examples are presented to guide ERC personnel in developing and implementing assessment and evaluation plans.

The structure of Gen-3 ERC education programs should inform this process—i.e., University Education (undergraduate and graduate), Precollege Education (RET and Young Scholar programs), and General Outreach designed to engage precollege students in the ERC's research area to stimulate interest in engineering careers. In addition, the University Education programs are designed so that the ERC graduates acquire skill sets needed to be effective in industry and creative and innovative in both academe and industry. This structure is often difficult for faculty to understand and the assessment/evaluation officer can be especially helpful in working with the ERC's Education Director in designing and assessing the impacts of this new approach.

4.6.2 General Guidelines

In order to develop an effective Assessment and Evaluation Plan, all stakeholders should be involved at the earliest stages of program development, including representatives from each partner institution. This will include Education Directors, assessment officers, program coordinators, and program evaluators. To ensure that all possess a clear understanding of the assessment purpose and planning process, the following steps are suggested:

- Assessment personnel and the related assessment plan should demonstrate an understanding of NSF requirements, including quantifiable outputs and the educational impact of study components.
- Assessment planning should co-occur with overall center programmatic planning and should be in place on the first day of center operation.
- Desired outcomes, perceptions, and expectations of the University and Precollege Education programs and the General Outreach to involve precollege students in the ERC activities, must be determined.
- Appropriate methodology and assessment tools must be selected for each activity.
- Timelines for each assessment component are a must.
- The center program Evaluation/Assessment Officer should be someone who is trained in qualitative (e.g., interview) and quantitative methodologies (survey design, psychometrics), and corresponding statistical and narrative analysis. Evaluators can be external or internal and each center must decide whether to have an external evaluator based upon discrete ERC needs. This person must understand the goals of the ERC program and the center in order to develop an appropriate design and instrumentation. Be mindful that many professionals trained in this field are accustomed to measuring learning outcomes, which is not a goal of ERC education programs, per se. Thus, precollege and university level programs will have different outcome goals that should be carefully determined using the ERC Program's performance assessment criteria and the center's own programs' goals.
- All projects must meet the Institutional Review Board (IRB) approval for not only the lead university, but also for the partnering universities and industry. Furthermore, it is a requirement that IRB approval be obtained before conducting any publishable research with human subjects.

One of the most challenging parts of the assessment process is determining appropriate expected outcomes. It is often the case that faculty tend to set unrealistic expectations and over-promise results. A common example is to list changes in state-wide standardized test scores as a result of a center program. Given the large number of variables that impact test scores, it is not reasonable to assume that a small-budget (in comparison to the total education budget that impacts test scores) intervention will have an impact on state-wide standardized test scores. It is therefore important that the assessment director work closely with research faculty to ensure that expected outcomes match the time, duration, and budget of the intervention.

Appendix section 4.6.1 provides several examples of program-wide education assessment and evaluation at different ERCs.

4.6.3 Assessment Design

There are multiple levels of information that can help guide the ERC education programs' development. Front-end evaluation is a useful tool, similar to market research. An example case where this would be useful is in the development of course materials that the ERC plans for adoption by a wide audience. Front-end evaluation would involve surveying the potential users (faculty) of the new materials about what topics they would like to see covered. Also, surveying potential students about what their existing level of knowledge about a topic is would uncover misconceptions that the developers could address. Incorporating end users into the design process results in better materials and facilitates adoption.

Many times, valuable information can be gleaned from informal quick studies with small numbers of participants. For example, prior to making a website or on-line unit public, it is always helpful to have small numbers of the intended audience beta test the site or materials. Problems with navigation and function can be easily corrected before "going public." Also, quick, short surveys can help guide programming. For example, finding out how current students learned about the center can help recruiters identify useful recruiting avenues that should be continued, as well as identify less productive methods that should be abandoned.

Formal assessment will also be appropriate in many cases. Pre- and post-assessment of knowledge and skills utilizing objective instrumentation is an accepted way to measure student learning outcomes. Instrumentation typically includes items testing for specific content knowledge, and over time and with due diligence, instrumentation can be revised and modified to enhance validity and reliability (Drummond & Jones, 2010).

Both quantitative (e.g., scales, rankings, etc.) and qualitative (e.g., focus groups, interviews) methods are useful. Quantitative designs can fail to capture the richness of phenomenological experiences best offered up through personal narrative, so supplementing quantitative measures with qualitative methods can produce a more complete description of outcomes. Guided discussion can bring about descriptive data useful to the assessment process (Vacc & Juhnke, 1997). These mixed-method designs, when properly done, result in rich quantitative and qualitative data which are mutually supportive, thus enhancing design internal consistency and validity and increasing results generalizability (Hanson, Creswell, Plano Clark, Petska, & Creswell, 2005).

At a minimum, mixed-method assessment designs should include clearly articulated goals and student's gaining skill sets. The essential goals are to determine whether the (i) mission statement is being properly addressed and (ii) students are gaining the desired skill sets. Content-specific instrumentation measuring teaching (i.e., educational activities) and learning (i.e., skill sets) is useful. Complementary case-by-case interviews or focus groups are also helpful.

There are useful frameworks to help organize the assessment and evaluation plan. One example is the Kellogg Logic model, which provides stakeholders with a visual template that connects activities to expected outcomes.[1]

4.6.4 Suggested Instrumentation

Instrumentation construction can often feel like a daunting task; however, the primary necessity for proper construction is time. Gen-3 ERCs are funded under cooperative agreements with an initial time line of five years and renewals can extend that to 10 years. Support is provided in annual increments. The first renewal review is during the third year, where NSF expects that the assessment program has been set up and is functioning effectively to guide practice. Three years is more than adequate to initialize and "study" instrumentation developed specifically for use within ERC education programs. Other requirements for instrumentation development include a good understanding of student learners, their backgrounds, and prior knowledge base, as well as the desired learning outcomes.

Table 1 provides suggested measures for assessing major education programs. Besides quantitative methods (e.g., survey), qualitative methods such as in-depth interviews are also useful to identify students' learning processes, outcomes and concerns. Initially, it is often a good idea to conduct a qualitative assessment due to the small sample size of most education programs.

Table 1. University and Precollege Assessment

Program	Example	Selected Measures
Undergraduate Program (NSF requirement)	• Academic-year Undergraduates Survey • Summer Research Experiences for Undergraduates (REU) Survey	• Career Path • Concept Inventory (e.g. Hestenes, Wells, & Swackhamer, 1992) • Research ability • Attitudes (Hilpert, J., Stump, G., Husman, J., & Kim, W., 2008). • Self-efficacy (Bandura, 2006) • Professional development (Rubric for evaluation of

[1] http://www.wkkf.org/resource-directory/resource/2006/02/wk-kellogg-foundation-logic-model-development-guide

Program	Example	Selected Measures
		presentation, self-assessment of key professional skills) • Creativity • Descriptive metrics: publications, presentations, attending graduate school/ industry
Graduate student skill-sets defined by each center (NSF Requirement)	• Entry survey • Exit survey • Employee assessment	• Longitudinal tracking on the progress of graduate students' skill-sets defined by your center (e.g., creativity, innovation, analytical skills, problem solving, leadership, motivation, communication skills) before starting and after graduating from graduate school. • Employee assessment on students' skill-sets at workplace. • Quantitative metrics: participations of professional training, publications, internship, and awards. • Engineering Global Preparedness (EGPI: Ragusa 2010, 2011) • Engineering Creativity and Propensity for Innovation (ECPII; Ragusa, 2011) • Course/program specific concept inventories
Precollege programs (NSF requirement)	• Young scholar program (YSP) survey • Research experience for teachers (RET) survey • Precollege partnerships • Portfolio assessment	• YSP: pre-and post-measurements on engineering knowledge, interest, research ability, attitudes, and future plan. Quantifiable metrics: publications, presentations, and persistence of interesting in studying in STEM. • RET: pre-and post-measurements on teaching efficacy , professional development, and engineering knowledge. • Quantitative metrics: impact of classroom curriculum development or research publications. • For students of RET Teachers: o Science literacy- specifically, science vocabulary, reading comprehension, science writing; (Ragusa, 2012) o Motivation for Science Questionnaire (Ragusa, 2012)
General outreach (NSF Requirement)	• Summer camps • Lab tours • Field trip • Community outreach	• Summer camps: pre and post measurement on engineering knowledge, interest in learning specific activity, self-efficacy in STEM, attitudes, & career/major preference. • Other outreach: post-measurement on interest in learning more, basic knowledge, and participant's feedback on overall programs.

Instrument sharing among ERCs is strongly encouraged. Granted, measurements will discretely vary according to ERC scientific and research orientation; nonetheless, assessment officers can talk among themselves to determine instrument-sharing advisability. The American Psychological Association recommends the following protocol for instrument sharing:[2]

1) Contact the instrument author to discuss instrument sharing.
2) Be mindful of copyright issues and obtain written permission from the instrument author prior to using the instrument.

As mentioned, instrumentation is generally discrete to each ERC research/scientific agenda. Thus, issues of fair use of copyrighted material must be considered. In short and when engaged in instrument sharing, the borrowing ERC, in collaboration with the instrument author, should discuss the likelihood of—or need for—instrument adaptation and discern the necessity for and ensuing transformative nature of those adaptations. A full explanation of fair use practices with copyrighted materials may be found at www.copyright.gov.

Overall, survey instruments should be carefully designed by the following steps:

[2] http://www.apa.org/science/programs/testing/find-tests.aspx?item=4

1) Determine the evaluation goals or purpose of assessment;
2) Gather and study existing assessment reports from NSF (e.g., REU program, RET program, and YS program);
3) Use published (validated and reliable) scales from the fields of education, engineering education, sociology and psychology for specific measures you are interested in; and
4) Finalize the survey by pre-testing on a small pilot set of representative students.

Structured interviews are one methodology for discovery in the assessment process, particularly when the interview questions are predicated on a specific taxonomy for learning or criteria for assessment (Vacc and Juhnke, 1997). Case studies, phenomenological interviewing, or focus groups can be used for structured interviews. Once again, guiding questions derive from a good understanding of (i) student learning, (ii) learning outcomes or skill sets, and (iii) and mission statement concepts. To be effective, the person guiding interviews must be professionally qualified for individual interviewing and managing group dynamics.

Appendix section 4.6.2.1 gives an example of the development of an education program assessment instrument by an ERC.

4.6.5 Data Collection and Management

Creating a systematic and organized method of tracking all the education information and data through websites or other web tools (for example, Google Docs, or Survey Monkey[3]) across university partners is crucial. Data collection and management plans should be developed as part of the ERC proposal process.

Documenting photos, videos, and other form of evidence for each program is beneficial for writing the annual reports and renewal proposals. Cloud computing can be used to share photos across partner universities, if permitted by the institutions, although photo release forms and signed forms should always be stored with photos.

With quantitative designs, SAS, Mplus, SPSS, or Microsoft Excel can be utilized to analyze pre-/post-data. In raw form, these data should be housed in a locked office, on a password-protected desktop of the Assessment Officer. Once analyzed and ideally, aggregated, data should be (i) transferred to the ERC reporting database, (ii) reported at the annual conference, and (iii) reported in the engineering education literature.

Qualitative data such as interviews or focus groups' narratives can be audio-taped and, when possible, should be video-taped for data collection. Software also exists for analyzing and presenting qualitative data, (for example: http://provalisresearch.com/products/qualitative-data-analysis-software/). All data must be stored according to IRB requirements and retained for the time period required by each university.

4.6.6 Using Assessment Data

Assessment is intended not only to measure impacts on students and teaching efficacy but also to gauge programmatic effectiveness. Modifying and improving programs is best done through systematic data collection, management, and analysis.

NSF requires center reporting on an annual basis, and this includes Assessment and Evaluation activities and results. Assessment results may be used by the Site Visit Team to evaluate the effectiveness of the education programs. It is recommended that the center strive to exceed NSF expectations, highlighting signature programs by reporting data through graphics, tables, and longitudinal assessment; the ERC should focus upon the broader impact of educational and outreach activities specific to these signature programs.

[3] http://www.surveymonkey.com

Besides assessing participants' gains in learning, interest, attitudes, teaching efficacy, or future career goals, it is important to evaluate each education program as a whole in order to identify the weakness and strength of program design. Program evaluation will serve the purpose of improving logistics and program design.

4.6.7 Notes

Rationale and Definitions for Assessment and Evaluation

The National Academy of Engineering has recommended both best practices and attributes for engineering education in their *Engineers for 2020* report (NAE, 2005). Additionally, the Academy defined Discipline-Based Education Research (DBER) (National Research Council, 2012). The NAE together, with the National Research Council, identified "assessment best practices," (NRC, 2011) as an important component of DBERs. In 2006, the Educational Testing Service (ETS) published three issue papers describing a Culture of Evidence—or evidence-centered design—as a methodology for systematically assessing post-secondary education effectiveness across institutions of higher education. Evidence-centered designs link institutional or programmatic vision and mission with student learning outcomes, which in turn are aligned with discipline-specific professional standards, and measured by, or exemplified through, concrete evidence (Millett, Payne, Dwyer, Stickler, & Alexiou, 2008). See section 4.6.8 below for reference citations.

Framers of the ETS papers emphasized that "at the heart" of an evidence-centered design is the issue of validity, whereby evidences measure or exemplify that which they purport to measure or exemplify. Evidence could include (a) annual data collection with valid/reliable instrumentation; (b) pre-/post-test designs using instruments with multiple forms; (c) a variety of assessment formats, including asking questions; and (d) "peer group comparisons." The goal of evidence-centered assessment is to produce valid and reliable data for decision-makers to determine higher education and programmatic effectiveness (Dwyer, Millett, & Payne, 2006; Millett et al., 2008; Millett, Stickler, Payne, & Dwyer, 2007).

Suggested Instrumentation

Resources for survey design and scale development from sociology and psychology disciplines:

Internet, Mail, and Mixed-Mode Surveys: The Tailored Design Method by Dillman, Smyth, and Christian.

Scale Development: Theory and Applications by DeVellis.

Reliability and Validity Assessment by Garmines and Zeller.

Psychometrics Theory by Nunnally and Bernstein

4.6.8 Selected References

Bandura, A. (2006). Guide for constructing self-efficacy scales. Self-efficacy beliefs of adolescents, 5, 307-337.

Drummond, R.J., & Jones, K.D. (2010). *Assessment Procedures for Counselors and Helping Professionals (7th ed).* Upper Saddle River, NJ: Pearson Merrill Prentice Hall.

Dwyer, C.A., Millett, C.M., Payne, D.G. (2006). A culture of evidence: Postsecondary assessment and learning outcomes. Princeton, NJ: ETS.

Hanson, W.E., Creswell, J.W., Plano Clark, V., Petska, K.S., & Creswell, J.W. (2005). Mixed methods research designs in counseling psychology. *Journal of Counseling Psychology, 52,* 224-235.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory.The physics teacher, 30, 141.

Hilpert, J., Stump, G., Husman, J., & Kim, W. (2008, October). An exploratory factor analysis of the Pittsburgh freshman engineering attitudes survey. In Frontiers in Education Conference, 2008. FIE 2008. 38th Annual (pp. F2B-9). IEEE.

Millett, C. M., Payne, D. G., Dwyer, C. A., Stickler, L. M., & Alexiou, J. J. (2008). A culture of evidence: an evidence-centered approach to accountability for student learning outcomes. Princeton, NJ: ETS.

Millett, C. M., Stickler, L.M., Payne, D. G., & Dwyer, C. A. (2007). A culture of evidence: critical features of assessment for post-secondary learning. Princeton, NJ: ETS.

NAE, (2005), Educating the Engineer of 2020, National Academy of Engineering, The National Academy Press, Washington DC.

National Research Council (2011). Promising Practices in Undergraduate STEM: Summary of 2 Workshops. Washington, DC: The National Academies Press.

National Research Council (2012). Discipline-Based Education Research: Understanding and Improving Learning in Undergraduate Science and Engineering. Washington, DC: The National Academies Press.

Ragusa, G. (2010) Preparing University Students For Global Workforces: Comparisons Between Engineering and Business School Students. 40^{th} ASEE/IEEE Frontiers in Education Conference Proceeding, 2011, Session F4J. Louisville: KY.

Ragusa, G. (2011) Engineering Creativity and Propensity for Innovative Thinking In Undergraduate and Graduate Students. 2011 American Society of Engineering Education Conference Proceedings. Session AC 2011-2749. British Columbia, Vancouver, Canada.

Ragusa, G. (2011) Engineering Preparedness for Global Workforces: Curricular Connections and Experiential Impacts. 2011 American Society of Engineering Education Conference Proceedings. Session AC 2011-2750. British Columbia, Vancouver, Canada.

Ragusa, G. & Lee, C.T. (2012) The Impact of Focused Degree Projects in Chemical Engineering Education on Students' Achievement, and Efficacy. Education for Chemical Engineers.7(3) 69-77.

Ragusa, G. (2012) Teacher Training and Science Literacy: Linking Teacher Intervention to Students' Outcomes in STEM Courses in Middle and High School Classes. 2012 American Society of Engineering Education Conference Proceedings. Session AC 2012-5242. San Antonio, TX.

Segreto-Bures, J., & Kotses, H. (1982). Experimenter expectancy effects in frontal EMG conditioning. *Psychophysiology, 19(4),* 467-471.

Thompson, B. (2002). "Statistical," "practical," and "clinical": how many kinds of significance do counselors need to consider? *Journal of Counseling and Development, 80,* 64-71.

Vacc, N. A. & Juhnke, G. A. (1997). The use of structured clinical interview for assessment in counseling. *Journal of Counseling and Educational Development, 75,* 470-480.

APPENDIX 4.6
ASSESSMENT AND EVALUATION EXAMPLES

4.6.1 Program-wide Education Assessment and Evaluation

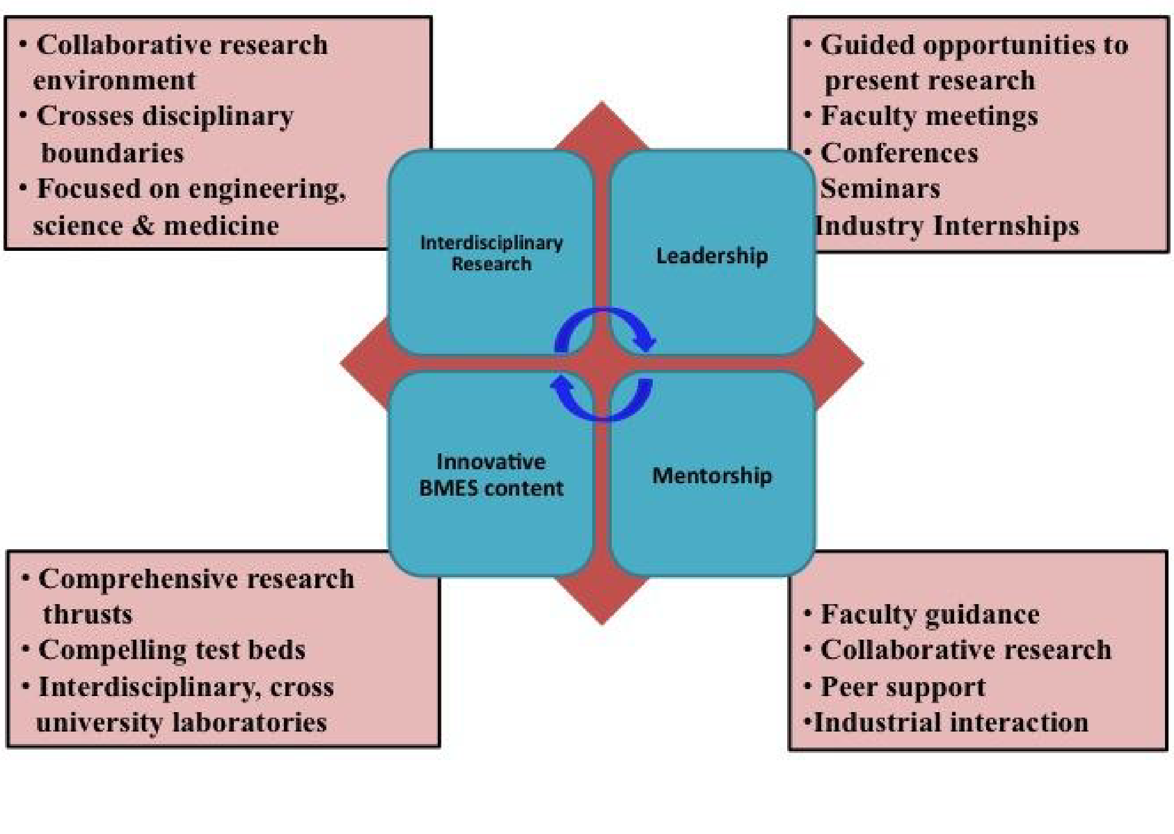
4.6.1.1 ***Center:*** Biomimetic MicroElectronic Systems (BMES) ERC, a graduated ERC (2003–2013)

Lead Institution: University of Southern California
Center Director: Dr. Mark Humayun, Dept. of Ophthalmology and Biomedical Engineering
Name of Program: BMES ERC Assessment Examples–Education and Outreach (E&O)

Type of Program: Assessment examples (across all E&O programs)

Program Synopsis: For the past 10 years, the BMES ERC cumulative educational assessment has focused on the following important areas: a) student knowledge gained resulting from BMES ERC-initiated courses; b) preparedness metrics aligned with recommendations of the National Academy of Engineering (*Educating the Engineer of 2020*, 2005) and the Accreditation Board of Engineering and Technology (*ABET Manual*, 2011); c) measures of students' creativity, propensity for innovation, and entrepreneurship; d) measures of preparedness of engineers for a global economy/workforce; and e) impact assessments that focus on the careers of Center alumni. Our education assessment results are presented in these areas for this report. These efforts have addressed both the Generation Two (Gen-2) and Generation Three (Gen-3) ERC requirements outlined in the NSF guidelines. Assessment results are presented for undergraduate, graduate, and alumni activities. Measurement and instrumentation is described for each program focus. The figure below illustrates the cumulative interconnectedness of our ERC education programs and associated assessment metrics.

Assessment of Interrelated Focus of Education Programs



Contact person/website: Gigi Ragusa (ragusa@usc.edu)

Dates of Operation/Timeframe: The assessment program began in year one of the ERC. It is ongoing and fully developed.

Background: Comprehensive assessment measures and procedures with foci on BMES ERC research areas enable us to assess the effectiveness of and continuously improve our undergraduate, graduate, and K-12 education and outreach efforts. It should be noted that, while Dr. Ragusa was funded under BMES core funds as the staff Assessment specialist, BMES also had a number of other grants to develop these assessments. These included the US Department of Education Institute of Education Sciences (IES) and the Department of Energy's Energy Frontier Research Centers (EFRC), as well as NSF's REE, RET, REESE, TuES, and the older CCLI and IEECI programs. While it is important to do assessment and evaluation, it is not expected that ERCs will always develop new tools out of ERC funds, but rather will use relevant, existing tools and other funding where available.

Methodology: These efforts are described in two major categories below: university-level assessments and outreach/precollege assessments. All assessments are psychometrically sound (Wilson, 2012) per Item response theory, valid, and statistically reliable. They are aligned with National Academy of Engineering, National Research Council (NRC), and American Engineering Association (AEA) best practices. All partner institutions in BMES ERC have Institutional Review Board (IRB) clearance for these assessments.

UNIVERSITY-LEVEL ASSESSMENTS

Biomedical Engineering (BME) Concept Inventories: Concept inventories were designed and administered to measure the knowledge gain associated with BMES ERC biomedical engineering courses. A concept inventory is a conceptual measure of students' course-specific knowledge (Hestenes et al 3]. Concept inventories are administered at the beginning and at the end of a course to measure gains in conceptual understanding as a result of completing the course. These inventories are best used for undergraduate courses that have primary foci on developing conceptual understanding. The concept inventories were aligned with the primary research areas of the BMES ERC (Hestenes, et al, 1992), and were the first developed in biomedical engineering.

ERC Student Research Experiences: We also measured the effects of undergraduate and graduate students' research experiences on their skills and perceptions of the Center's research program by administering an interdisciplinary research questionnaire. This was aligned with NAE's Engineers of 2020 recommendations. These results were compared to a rubric rating of the ERC student research teams' annual presentations.

The Engineering Global Preparedness Index: In concert with NAE recommendations and guidelines for NSF Gen-3 ERCs, we measured our BMES ERC students' preparedness to work in global workforces. An 18-item questionnaire with a six-point Likert-type scale was developed. This assessment has also been used at thirteen universities nationally as a measure of achievement of the Accreditation Board of Engineering and Technology (ABET) outcomes F (ethics), G (communication), and H (global preparedness) using workforce preparedness as an assessment frame.

Four subscales are used as constructs in the engineering global preparedness index (EGPI). These scales are aligned with important skills needed by both engineers and other professionals who work in global marketplaces (and aligns with Gen-3 ERC requirements). A description of each EPGI construct/subscale follows.

- **Engineering Ethics:** Depth of concern for people in all parts of the world; moral responsibility to improve conditions and take action in diverse engineering settings (ABET OUTCOME F).
- **Global Engineering Efficacy:** Belief that one can make a difference; support for personal involvement in local, national, international engineering issues and activities towards achieving greater good using engineering technologies (ABET OUTCOMES G AND H).
- **Engineering Global-centrism:** Valuing what is good for the global community in engineering related efforts, not just one's own country or group; making judgments based on global needs for engineering and associated technologies, not ethnocentric standards (ABET OUTCOME H).
- **Engineering Community Connectedness:** Awareness of humanity and appreciation of interrelatedness of all peoples and nations and the role that engineering can play in improving humanity and meeting human needs. (ABET OUTCOME G)

Engineering Creativity and Propensity for Innovation: Instruments were developed to measure students' creativity and propensity for innovation. This instrument was designed and tested because of ongoing conversations with engineering educators nationally and the desire to assess the role that comprehensive educational and engineering experiences have in important industrial and academic skill sets: creativity and innovation. This instrument is currently being used at two other ERCs and three additional universities with other programs. This instrument is aligned with several theoretical perspectives on creativity research (Torrance, 1974; Abedi, 2007; and Khatena, 1999). Constructs included in the revised Engineering Creativity and Propensity for Innovation Index (ECPII) are engineering fluency, flexibility, disciplined imagination, originality, and design thinking. These constructs are closely aligned to the combined research on creativity and innovation and have domain specificity to engineering.

ECPII's five theoretically grounded constructs (measured in subscales) are described below. Problem solving is the focus of this metric.

- **Engineering Fluency:** Students' breadth of understanding of diverse aspects of the engineering disciplines.
- **Engineering Disciplined Imagination:** Students' ability to imagine diverse/unique possibilities within engineering disciplines.
- **Engineering Flexibility:** Students' ability to imagine diverse problem-solving approaches within the engineering discipline coupled with ability to use a diverse engineering problem-solving skill set in the face of distractors.
- **Engineering Originality:** Students' ability to develop and design problem-solving approaches that are unique.
- **Engineering Design Thinking:** Students' ability to design with both breadth and depth using an advanced problem-solving approach.

The instrument has 37 items on the ECPII with three to five items per subscale as described above. This item distribution and scale total is supported by item response theory for designing difficult to observe (i.e., soft skill) constructs. A minimum of two items per subscale in the index are: a) reverse-scored items in support of best practices in survey development; and b) true measurement of students' ability (rather than student perception) beyond what is self-reported. A six point Likert-type scale was employed for the majority of the ECPII items. A final set of items situated at the end of the index are open-ended and include the requirement that respondents read a context and discipline embedded-scenario and solve an engineering problem via a listing or illustration of steps to problem-solving. This subset of items is rated using a six-point checklist aligned with the subscales in the ECPII.

Research Experience for Undergraduates (REU): As assessments for the BMES REU program, both a questionnaire (also used with our graduate and undergraduate students as described above) and an annual focus group with the participants while they were participating in the REU were used.

Alumni Assessment: Assessing and tracking BMES ERC student alumni since the ERC's conception has been conducted. A survey is done of graduates to assess their career trajectories and includes alumni who took ERC courses, engaged in an ERC lab experience, or both. They were also followed via social working sites.

Education Assessment Summary: Our cumulative assessments across ERC education programs continue to reveal impactful results. Our ERC students are prepared for industry employment and graduate school and have benefited from the diverse, interdisciplinary education and research activities provided by the ERC. It is important to note that the development of assessment instruments is an iterative process and modifications are often required before the most useful length and format is developed.

OUTREACH/PRECOLLEGE ASSESSMENTS

Science for Life: Content specific concept inventories for our Science for Life program in grades 3-5 were developed. A science and engineering interest survey for students to measure changes in students' interest in science and engineering across the three grade levels is used. Students' California Standards Test (CST, an achievement test) scores are tracked across years in school to monitor changes for those participating in the program.

Engineering for Health: In our Engineering for Health program—a high school "Young Scholars" type of program—we also track students, CST scores across years in school to monitor changes for those participating in the program. We also have a high school science motivation scale that we use to monitor changes while participating in the program.

Research Experience for Teachers (RET): A science literacy qualitative inventory to measure participants' science teaching and performance and students' (of RET teachers) science motivation and engagement and science literacy was developed and tested for reliability and validity.

Five assessment metrics were used to judge the success of the teacher and associated student intervention programs (RET and California Postsecondary Education Commission Improving Teacher Quality [CPEC ITQ]). These include: a teacher instructional performance measure: science teaching efficacy measure: student science concept inventory: student science literacy measure: and a student science motivation, engagement, and interest measure.

Teacher Metrics

- **Teacher Instructional Performance Metric:** We used a rubric-scored observational assessment of science teacher instructional performance aligned to California's teacher performance assessment entitled Performance Assessment of California's Teachers (PACT).
- **Science Teaching Efficacy Beliefs Instrument Revised (STEBI-R):** This instrument is a teacher metric and is a measure that assesses the teacher's efficacy in teaching science to middle and high students. It includes personal science teaching efficacy and science teaching outcome expectation, and is administered as a pre- and post-test to all teacher participants, then compared to non-participant science teachers that match the participant teachers demographically (based on national averages).

Student Metrics

- **Science Qualitative Reading Inventory:** This metric measure students' science literacy by grade level. It includes a measure of science vocabulary, reading comprehension, and science writing and is matched in terms of grade level science content and vocabulary.
- **Grade and Content Specific Concept Inventories:** These concept inventories measures of grade level concepts critical to scientific understanding.
- **Motivation for Science Questionnaire:** This questionnaire measures students' interest, motivation, and engagement in science.

Impact/benefits: This comprehensive suite of assessment provides both formative and summative impact-focused assessments of all education and outreach programs at BMES ERC.

Sustainability: This assessment suite for our programs has been institutionalized at USC and with our partner institutions (university, community college, and K-12).

Tips: Start designing assessments based on research efforts early, obtain IRB clearance at all sites in our ERC partnerships, and work at sustainability/institutionalization early on (at least by year three).

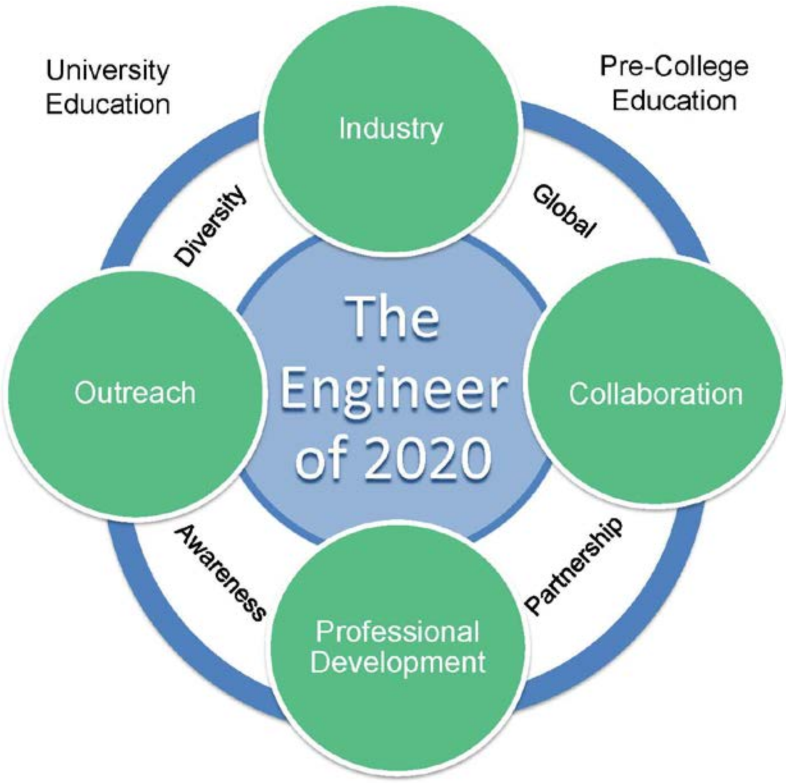
4.6.1.2 ***Center:*** Center for Integrated Access Networks (CIAN) ERC

Lead Institution: University of Arizona
Center Director: Dr. Nasser Peyghambarian, Department of Materials Science and Engineering
Name of Program: CIAN Assessment Examples–Education and Outreach (E&O)

Type of Program: Assessment examples (across all E&O programs)

Program Synopsis: CIAN is a Gen-3 ERC and, as such, its university education programs assessment must address the requirements delineated by NSF. Essential to CIAN's systematic assessment plan was to develop early-on clear objectives of CIAN's education program. CIAN's assessment efforts measure the extent to which these objectives are achieved by measuring outcomes of CIAN's education programs, such as increases in knowledge and awareness and changes in attitudes or beliefs. CIAN's education objective is to engender in its students the National Academy of Engineering's *Engineer of 2020* attributes. This objective drives the activities CIAN develops and nurtures. The outcomes of producing engineers with these skill sets are assessed using an array of tools, such as self-report surveys, interviews and focus groups, graduate student portfolios, and alumni tracking efforts. The figure below shows a quadrilateral paradigm that visually demonstrates the rationale behind CIAN's program development. It shows how CIAN's education activities are mapped to the CIAN objective it is designed to engender.

Quadrilateral Paradigm Tied To CIAN ERC Education Objectives



Contact person/website: Allison Huff MacPherson (allison@optics.arizona.edu)

Dates of Operation/Timeframe: The CIAN ERC was founded in 2008.

Background: CIAN's education programs strive to engender the skill sets consistent with the *Engineer of 2020* attributes in its students, as outlined in the figure below. Programs and activities are developed and assessed with this in mind.

Engineer of 2020 **Attributes**

		Strong Analytical Skills	Practical Ingenuity, Innovation, & Creativity	Communicati on	Business, management and leadership	Standards and Professionali	Dynamism, Agility, Resilience, & Flexibility	Life Long Learners
INDUSTRY ACTIVITIES								
Workshops on Industrial Practices	• Post-Activity Survey	✓	✓		✓		✓	
Industry Webinars	• Data Collection • Graduate Student Portfolio		✓	✓	✓			
Internships	• Participant Reports • Data Collection • Alumni Surveys	✓	✓	✓		✓		

Industry Speed Networking	• Data collection		✓	✓			✓	✓
Graduate Student Portfolio	• Interview • Data Collection		✓	✓	✓			✓
COLLABORATION ACTIVITIES								
SLC Organizing Committee	• Graduate Student Portfolio			✓	✓	✓		
Student Web Presentations	• Data Collection • Graduate Student Portfolio	✓	✓	✓				✓
Student Travel Grants	• Data Collection • Participant Reports*		✓				✓	✓
SLC Student Retreat	• Post-Activity Survey • Data Collection		✓	✓	✓		✓	✓
IAB Annual Meeting	• Post-Activity Survey • Data Collection		✓	✓	✓		✓	
Cross-CIAN and Cross-Discipline Collaboration Efforts	• Data Collection • Deliverables *	✓	✓	✓	✓	✓	✓	
Undergraduate Research Fellowships	• Data Collection • Participant Reports*	✓		✓		✓		✓
REU	• Pre-Post Surveys • Deliverables • Alumni Surveys	✓		✓		✓		✓
Student Presentations	• Data Collection • Graduate Student Portfolio • Faculty	✓	✓	✓				✓
PROFESSIONAL DEVELOPMENT ACTIVITIES								
Mentoring REU/RET/Young Scholars	• Participant Mentor Surveys* • Graduate Student Portfolio			✓	✓	✓		✓
Professional Development Workshops	• Post-Activity Survey • Graduate Student Portfolio	✓	✓	✓	✓	✓	✓	✓
Complete Online Training Module for Responsible Conduct of Research	• Data Collection • Graduate Student Portfolio					✓		✓
Education Presentations	• Graduate Student Portfolio • Data Collection		✓					✓
Perfect Pitch Competition	• Graduate Student Portfolio • Presentation critique rubric	✓		✓			✓	✓
Super-course Modules	• Course Evaluation • Faculty Evaluation • Graduate Student Portfolio • Data Collection	✓	✓	✓		✓	✓	✓
OUTREACH ACTIVITIES								
Presentation and Demonstrations	• Short oral pre-post survey* • Data collection		✓	✓	✓	✓		

RET	• Pre-Post Surveys • Deliverables • Data Collection • Follow-up interviews • Participant Mentor Evaluation*	✓	✓	✓		✓	✓	✓
Young Scholars	• Pre-Post Survey* • Deliverables • Data Collection • Monthly report of experiences • Follow-up Interviews	✓	✓	✓		✓	✓	✓
Summer Camps	• Pre-Post Surveys • Data Collection • Deliverables	✓	✓	✓		✓	✓	✓
Community Outreach Activities	• Data Collection		✓	✓	✓	✓		✓
Tours	• Data Collection			✓	✓			

Note: For additional relevant information, see Appendix 4.5, part A: Programs to Produce Desired Skill Sets, in the Graduate Education section of the chapter.

Methodology: The types of assessments CIAN uses are described below. Most assessment tools are self-report/survey items. Institutional Review Board (IRB) clearance at each partner university should be obtained.

Formative (or process) evaluation provides ongoing measurement while developing CIAN education activities for the purpose of monitoring and improving the achievement of CIAN's educational objectives of developing an education program that graduates students who are more effective in industrial and academic practices. Formative tools are used to assess CIAN's education program and assure that its activities are aligned with the desired skill sets CIAN's education program is engendering in its students. Results from the process evaluation are used mainly to improve the quality of the activities and reinforce our educational strategy.

Summative (or outcome) evaluation provides a comprehensive evaluation at the conclusion of an activity that measures the achievement of CIAN's education outcomes of producing engineers who are creative, innovative and adaptive with skill sets in line with the *Engineer of 2020* attributes. Insight from participants and examples of their efforts are typical tools CIAN uses to measure the achievement of learning outcomes.

Some amount of change is expected from engagement with each of the main CIAN activities; however, students who participate the most in CIAN's education activities will likely see the biggest achievement of learning outcomes. These students tend to be our Student Leadership Council (SLC) officers and representatives. Tracking these students post-graduation is important. The SLC represents the nexus of the ideal research, industry, and education overlap of goals and opportunities. The long-term intent for assessment of CIAN graduates will include all students (undergraduates and graduates) and post docs active in CIAN research labs and enrolled in CIAN-related courses. However, the focal point of the longitudinal assessment will be the members of the SLC.

The assessment tools used to measure the outcomes of each activity are included in the method to assure proper program management.

Impact/benefits: Program evaluations provide valuable feedback from participants, enabling CIAN to modify its education programs each year and continually improve outcomes (e.g., immediate jobs in industry for CIAN graduates, REUs prepared for graduate school, or K-12 students interested in pursuing education in optics).

Sustainability: With the help of a graduate assistant working specifically on assessment and evaluation, CIAN's education director has clarified the goals of CIAN education assessment and begun a targeted effort to improve assessment strategies, develop new assessment tools, and obtain more quantifiable data going forward. These assessment methods require minimal modification each year, thus lending to the ease of this assessment plan's sustainability.

Tips:

- Collaboration with more established ERCs is recommended in order to exchange ideas regarding measurement tools and assessment plans; likelihood of success is increased when building on others' time-tested methods.
- It is best practice to stagger the distribution of surveys and other self-report tools several months apart to avoid "burn-out" and low response rates from the same pool of respondents. To maximize response rates, make sure to send out one or two reminder emails, along with a deadline, following the initial invitation to participate in surveys.
- Preparation of the assessment tools should begin long before the first day of programs so there is sufficient time to address "hiccups," unforeseen obstacles, and make edits to surveys and other tools.

4.6.1.3 ***Center:*** ERC for Revolutionizing Metallic Biomaterials (ERC-RMB)

Lead Institution: NC A&T State University
Center Director: Dr. Jagannathan Sankar, Department of Mechanical Engineering
Name of Program: Educational Activities and Assessment

Type of Program: Assessment examples of Research Experience for Undergraduates (REU) / Research Experience for Teachers (RET) / Young Scholars (YS) Programs

Program Synopsis: The ERC-RMB has supported REU/RET/Young Scholars summer programs at North Carolina A&T State University (NCAT). These programs link ERC-RMB researchers/scientists, NCAT faculty, and graduate students with undergraduate, teacher, and young scholar participants in teaching and learning activities for six weeks. REUs have come from the ERC-RMB lead institution (NCAT), partnering institutions (University of Cincinnati, University of Pittsburgh), and from universities from across the country. RETs come from K-14 environments, and Young Scholars were middle and high school students, from the local school district. During the six-week program, REU/RET/YS participants engaged in laboratory work, classroom instruction, seminar discussion, and field trips to local industry in bioengineering. Participants had research responsibilities carried out under the guidance of the ERC-RMB mentors (i.e., research/scientists). REU/RET/YS participants completed pre-/post-assessment surveys to determine changes in their understanding related to bioengineering, creativity and innovation, diversity of thinking, entrepreneurship, and ethics. RETs were assessed for their perceptions of programmatic activities upon K-14 teaching and learning. REU/RET/YS participants engaged in weekly focus groups, and assessment personnel gathered observational data from laboratory activities.

Contact person/website: Robin Guill Liles, Associate Director for Educational Assessment (rgliles@ncat.edu), http://erc.ncat.edu/

Dates of Operation/Timeframe: The REU/RET/YS program associated with the ERC-RMB has been operational each summer since 2009. The program is designed to last at least six weeks. REU/RET/YS program is a residential camp whereby participants live on the NCAT campus. Activities were planned from 8:00am–9:00pm.

Background: Initially, the REU/RET/YS program served as the gateway educational activity for the ERC-RMB Education and Outreach (E&O) program. This program was included in the original ERC-RMB E&O conceptual plan. Over the years, the REU/RET/YS program has continued to grow and diversify. Faculty and research/scientist mentors have refined their teaching modules, both within and out of the laboratory. Field trips and other extracurricular activities have become more sophisticated. Most notably, framers of the ERC-RMB E&O programs continue to recruit and retain record numbers of participants from underrepresented populations.

Assessment Methodology: In the first year of the Center's REU/RET/YS program, the assessment team implemented pre-/post assessment interviews. Interview questions flowed from the ERC-RMB E&O mission statement to "to train future engineers for industry, research, and development in a multidisciplinary environment that values diversity of thinking, creativity and innovation, and entrepreneurship." Interview transcripts revealed frequently occurring themes and factors. In year two, a written pre-/post- assessment instrument was developed focusing upon these themes and factors. The 80-item survey was administered electronically through Survey Monkey. In addition, the assessment team initiated a series of weekly focus groups. In an effort to refine assessment data, satisfaction questionnaires were implemented in the third year of the REU/RET/YS program. Their purpose was to determine which extracurricular activities participants deemed most beneficial to their REU/RET/YS experiences. From the first three years of assessment data, it became evident that REU/RET/YS participants found laboratory experiences important to their positive change in understanding and knowledge related to bioengineering. For this reason, in the fourth year, the Home Observation Scale (HOME) was adapted to use in a series of laboratory observations. Two members of the assessment team systematically observed REU/RET/YS participants as participants engaged in their laboratory activities. Once observations were completed, team members compared observations. Only in those observations with 100% observer rate agreement were reported. In year five of the ERC-RMB REU/RET/YS program, a concert of assessment instruments were used, including pre-/post survey assessment, focus groups, satisfaction questionnaires, and laboratory observations.

Impact/benefits: Assessment data from the REU/RET/YS program provide impressive evidence that a portion of participants go on to self-select into academic and career bioengineering strands. From the ERC-RMB ethical perspective, this selection process has particular value in bringing members of under-represented populations into the field of bioengineering.

Tips: Developing assessment strategies for the first iteration of educational and outreach programs and activities can be a time-consuming and even overwhelming process. Many young ERC E&O personnel report confusion about the timeline and impending site visit. Having a professional with particular expertise in assessment is helpful. In addition, good note- and record-keeping is a must for writing cogent end-of-year reports.

4.6.2 Education Assessment Research

4.6.2.1 ***Center:*** ERC for Quantum Energy and Sustainable Solar Technologies (QESST)

Name of Program: Conceptions of Engineering Scale

Type of Program: Precollege engineering education research project

Program Synopsis: This project addressed an engineering education knowledge gap through the development of a valid and reliable instrument to measure middle school students' conception of engineering based on analysis of three conceptual frameworks currently used in K-12 engineering education.

Contact person/website: Michelle Jordan (michelle.e.jordan@asu.edu), Chrissy Foster (christina.foster@asu.edu), Jenefer Husman (jenefer.husman@asu.edu)

Dates of Operation/Timeframe: 2013 & continuing

Background: Educational research results suggest exposure to engineering needs to begin in middle school in order to provide equitable opportunities for students to consider majoring in STEM fields. However, students' conceptions of engineering may be inaccurate, resulting in dropping out of engineering degree programs. It is important to understand students' early conceptions of engineering and how educational experiences influence those conceptions. Although inroads have been made in identifying developmentally appropriate knowledge and skills related to engineering, there is a need for a conceptually sophisticated understanding of engineering knowledge to guide K-12 standards and assessment. To date, few researchers have attempted to measure middle school students' conceptions of engineering.

The student sample targeted for the study is of particular interest as they are representative of populations currently under-represented in STEM careers (Gibbons, 2009) and often under-prepared for STEM careers by K-12 schools (Museus et al, 2011).

Note: This project was funded with a separate grant focused on engineering education research, not from ERC core funds. The ERCs are not expected to conduct engineering education research with core funds but rather, hopefully, to use the outcomes and results of prior or parallel research to help improve the execution, monitoring, and improvement of their programs.

Methodology: Assessment Scale Development

Six broad aspects important to middle school students' conceptions of engineering were identified:

- The purpose of engineering to create a solution to a problem that exists at multiple levels of society
- The nature of engineering as a systematic, iterative, and creative process with requirements and constraints
- Phases/elements of the design process
- Engineers use certain kinds of tools in particular ways for specific purposes
- Engineering is a highly social activity requiring collaboration and communication with diverse people for multiple purposes using a variety of mediums
- Engineers develop habits of mind (e.g., optimism, creativity, systemic thinking)

Having identified these six broad aspects of engineering education, they were considered in relation to common misconceptions of engineering previously identified. Additional items were constructed that asked respondents to reflect on structures, behaviors, and functions of engineering. Initial items were field-tested using interviews with practicing engineers. Items were added, deleted, and revised based on their feedback. These activities resulted in the development of an initial 25-item multiple-choice instrument comprised of six subscales to measure students' conceptions of engineering.

Notes: Scale development was based in part on findings from 2011-2012 data collected in nine Math Engineering Science Achievement (MESA) Clubs. Study participants were recruited from middle school students who participated in afterschool engineering clubs during 2013-2014. Likert-items, open-ended survey questions, and interviews yielded insights into middle school students' conceptions/ misconceptions of engineering, self-perception of themselves as engineers, and perceptions of their MESA experiences. Analysis suggested that students valued their learning from hands-on collaborative challenges. However, they varied considerably in the sophistication of their understanding of what engineering is and what engineers do (Jordan, Foster & Husman, 2013a, 2013b).

Impact/benefits: The Conceptions of Engineering Scale was field tested by being administered to 100 participants diverse in age (sixth through eighth graders), ethnicity, gender, and socio-economic status (SES). Fifty-nine respondents had participated in an engineering after-school club (MESA); 41 respondents had not participated. MESA is a co-curricular afterschool program, currently offered in eight states. MESA clubs are typically advised by volunteer teachers. MESA traditionally supports educationally disadvantaged students in Title I schools by providing pathways for minority, low-income, and first generation college-bound students to succeed in STEM disciplines. Students collaborate in design challenges and compete in statewide completions. Results are currently being analyzed.

Interviews were conducted to probe students' reasoning related to their responses. Using think-aloud protocols, we asked students to explain their answers to three to six survey items (e.g., "What were you thinking when you responded to this question?" and "Why did you select that response?").

The development of a valid and reliable instrument to measure middle school students' conceptions of engineering can be used to a) design and evaluate the impact of in-school and after-school programs that emphasize collaborative project-based, engineering design learning experience, b) improve outreach programs which have as their goal to improve knowledge and understanding of engineering and promoting engagement in STEM, c) influence research about STEM activities and coursework as it relates to the recruitment and retention of middle school students' from diverse backgrounds, and d) identify new information about students' existing knowledge and understanding of engineering that middle school science and math teachers will find useful for assessing conceptions of engineering as they begin to implement the Next Generation Science Standards.

Tip: Further scale development is planned. Pre- and post-administration of the scale to a group of 600+ middle school members of MESA clubs is planned for 2013-2014.