Section 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 4.1 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 4.1: 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 (Wells, & Swackham • Research ability • Attitudes (Hilpert, J. Husman, J., & Kim, • Self-efficacy (Bandu

Program	Example	Selected Measures
		• Professional develo evaluation of preser assessment of key skills)
		• Creativity
		• Descriptive metrics: presentations, atten school/ industry
Graduate student skill-sets defined by each center **(NSF Requirement)**	• Entry survey • Exit survey • Employee assessment	• Longitudinal tracking of graduate student defined by your cen creativity, innovation problem solving, lea motivation, commun before starting and from graduate scho • Employee assessm skill-sets at workpla • Quantitative metrics professional training internship, and awa • Engineering Global (EGPI: Ragusa 201 • Engineering Creativ Propensity for Innov Ragusa, 2011) • Course/program spe inventories
Precollege programs **(NSF requirement)**	• Young scholar program (YSP) survey • Research experience for teachers (RET) survey	• YSP: pre-and post-r engineering knowle research ability, attit plan. Quantifiable m publications, preser persistence of intere in STEM.

Program	Example	Selected Measures

	• Precollege partnerships
	
	• Portfolio assessment
		• RET: pre-and post-m... teaching efficacy , p... development, and e... knowledge.
		
		• Quantitative metrics... classroom curriculu... or research publicat...
		
		• For students of RET...
		
		   ○ Science liter... science voc... comprehens... writing; (Rag...
		
		   ○ Motivation fo... Questionnai... 2012)

General outreach **(NSF Requirement)**	• Summer camps • Lab tours • Field trip • Community outreach	• Summer camps: pre... measurement on en... knowledge, interest... specific activity, self... STEM, attitudes, & ... preference. • Other outreach: pos... on interest in learnir... knowledge, and par... feedback on overall...

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:

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.

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. 40th 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.

1http://www.wkkf.org/resource-directory/resource/2006/02/wk-kellogg-found...

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

3https://www.surveymonkey.com

Source URL: https://erc-assoc.org/best_practices/section-46-assessment-and-evaluation