Int. J. Engng Ed. Vol. 17, No. 1, pp. 81–88, 2001
Printed in Great Britain.

0949-149X/91 $3.00+0.00

A Comprehensive System for Student and Program Assessment: Lessons Learned*

SUSAN DABNEY CREIGHTON, ROBERT L. JOHNSON, JAMES PENNY and EDWARD ERNST
University of South Carolina, Columbia, SC 29208, USA. E-mail: ische45@um.sc.edu

Recent initiatives by the National Science Foundation (NSF) and Accreditation Board for Engineering and Technology (ABET) have promoted changes in classroom practices to move instruction from a traditional lecture to structured activities that require students to use the knowledge, skills, and abilities required by engineers in the workplace. For instructors to make valid inferences about whether students have mastered these work-related skills, assessments must be aligned with these new curricular emphases. Performance-based assessments offer a more direct method of assessing student outcomes identified in the ABET criteria. Such a shift in instructional and assessment practices in the classroom requires a concomitant shift in the evaluation of the effectiveness of such engineering programs. This article summarizes lessons learned by the authors in two disciplines, engineering and educational research, in the implementation of an assessment system and the use of performance-based assessments to evaluate student and program outcomes.

LITERATURE REVIEW

THE ABET criteria stipulate that engineering programs should evaluate the success of students in meeting program objectives. Colleges must have '. . .published educational objectives that are consistent with the mission of the institution . . .' and the Engineering Criteria 2000 (EC 2000). The criteria also require an ongoing evaluation to gauge the effectiveness of the program. The objectives of ABET accreditation include [1]:

1. '. . . that graduates of an accredited program are prepared adequately to enter and continue in the practice of engineering.'
2. '. . . stimulates the improvement of engineering education.'
3. '. . . encourages new and innovative approaches to engineering education.'

The ABET criteria also require engineering programs to include a continuous quality improvement process. In this model, program evaluation documents progress towards achievement of goals established by the engineering program and uses this information to improve the program. The criteria require that programs demonstrate student outcomes of such complex skills as:

- an ability to design and conduct experiments, as well as to analyze and interpret data, and
- an ability to design a system, component, or process to meet desired needs.

Types of evidence advocated by ABET to document these student outcomes include:

- portfolios;
- design projects;
- nationally-normed, content examinations;
- alumni/alumnae and employer surveys.

To advance the criteria, ABET and NSF have promoted diversity in classroom practices that move instruction from a traditional lecture to structured activities reflecting what engineers do in the workplace. These initiatives promote changes in classroom practices to reflect the knowledge, skills, and abilities required by engineers to conceptualize, articulate, and implement a solution for engineering problems. The reform movement advocates that engineering curricula incorporate a variety of teaching methods to involve students in active learning, design projects, technology use, and multidisciplinary teams [2, 3]. Performance-based assessments, in the form of design projects, portfolios, and lab reports, are more direct measures of student learning than multiple-choice exams. Such assessments enable faculty to more closely link student competencies with the expectations of the workplace.

As classroom practices change, engineering programs have begun to document student competencies through the use of performance assessments [4] and portfolios [5, 6]. A comprehensive assessment system, then, must address issues in the implementation of performance-based assessments in the classroom and in program evaluation. The following sections review lessons learned in the development of performance-based practices from two perspectives: engineering education and educational research.

LESSONS LEARNED IN ENGINEERING EDUCATION

During the last three years the College of Engineering and Information Technology at the

* Accepted 20 March 2000.

University of South Carolina (USC) implemented an infrastructure for supporting a continuous quality improvement system for the evaluation of student progress and program improvement. Following are lessons learned at USC from this experience. These lessons derive from qualitative and quantitative data based on participant observations, student and faculty interviews, stakeholder surveys, a computer database, and collaborations with other engineering institutions.

New appointment

The comprehensive assessment system became institutionalized within the College of Engineering by the employment of a professional to guide its development and implementation. The College created a Director of Assessment position to maintain a continuous quality improvement process by:

- providing technical support for faculty implementing assessment processes in each program and within their classrooms and
- implementing a college-wide assessment infrastructure.

The assessment director develops instruments, such as stakeholder surveys, and organizes employer focus groups. In addition, the director collaborated with a faculty assessment committee and the Office of Institutional Planning and Assessment to develop a course survey and a student longitudinal tracking system. Data are analyzed and reported to faculty and administration with feedback playing a key role in program improvements.

One example of a change made as a result of student feedback within the assessment system is a modification in the advisement process within the Electrical/Computer Department. Survey responses of graduating seniors indicated that students were not satisfied with the advisement process and the lack of faculty participation in that process. In response to student dissatisfaction, the department modified its advisement system. The director developed an instrument to monitor the advisement modification. Evaluation of the new advisory process indicated that student perceptions improved as a result of the change, and these positive results were reported to the department and the Assessment Committee. Having the appropriate personnel available to undertake the assessment brings closure within ABET's two-loop approach that links college-wide and program evaluation.

New challenges

Getting faculty and programs to implement the recommended changes in teaching and assessment has offered challenges. Although aware of the new accreditation criteria, many instructors have not perceived the benefits that accrue when alternative teaching and assessment methodologies are implemented in the classroom. When asked in what ways they have changed the way they teach during the past three years, approximately 30% of the responding faculty members noted no change in the implementation of these ABET-recommended approaches. In addition, 64% or more of the College faculty indicated they had not modified their teaching styles in the following areas:

- using a variety of assessment methods;
- accommodating differences in student learning styles;
- using team teaching;
- providing evaluation feedback to students.

Data from a survey of seniors indicated that less emphasis has been given to the implementation of student teams; integration of non-technical issues such as ethics and current events; and interaction with students. However, about 60 percent or more of the faculty reported being more inclined to:

- using technology to deliver instruction;
- incorporating computer exercises into coursework;
- using a variety of teaching strategies in the classroom.

Faculty targeted four areas for improvement to support implementation of teaching and assessment approaches advocated by ABET:

1. Place more emphasis on education in the tenure and promotion system.
2. Increase the number of faculty.
3. Provide funding or release time to develop innovative instructional approaches.
4. Provide state of the art classrooms and labs.

To promote faculty implementation of the desired changes in teaching and assessment, engineering programs may want to consider these faculty suggestions.

New opportunities

Implementation of the comprehensive assessment system offered faculty a variety of opportunities for professional development. Faculty in the College of Engineering and Information Technology at USC completed a short needs assessment to identify the topics and training that would enhance their teaching-assessment capabilities. A majority (95%) of the respondents indicated that their professional degrees did not include educational or teaching instruction; only 10% of the responding faculty had participated in workshops, seminars or conferences in which assessment topics were the focus. Many engineering instructors, therefore, may have been hesitant to implement changes in the classroom because they lacked practice in alternative assessments such as question and answer dialogues, administering classroom teamwork, or evaluating oral presentations.

To address this need, the College organized in-house conferences, workshops, and seminars for faculty and staff members. Targeted faculty members also received support to increase their competencies in a number of ways:

- attending conferences or workshops to learn more about assessment;
- communicating with faculty in Colleges of Education and with other Colleges of Engineering to discuss reforms and how changes have been implemented;
- seeking information on the web regarding particular topics such as ethics;
- piloting one new task or assignment each semester.

Faculty indicated that these experiences increased interaction among colleagues within the College, exposed them to new teaching strategies, provided them with opportunities to collaborate with faculty at other institutions, and presented them with potential areas for research and publication. However, thus far, the limited participation of faculty has minimized program impact and plans are underway to encourage more faculty involvement.

New attitudes

Some students were resistant when alternative teaching and assessment methodologies were introduced into their courses. Recent research involving the implementation of engineering portfolios indicates that students are resistant to change, wanting to retain the traditional lecture teaching model [5]. In their evaluations of one engineering course at USC, students indicated a preference for the lecture format because it requires less active involvement and is predictable. Teamwork, design projects, oral presentations, and written reports, however, are the types of skills that employers expect engineering graduates to possess, and students cannot achieve competency in these areas without practice. Students can gain experience through participation in professional organizations, internships and co-op jobs, but engineering colleges must also provide ample practice opportunities within the curriculum.

Student comments on some surveys demonstrated resistance to development of these multifaceted skills. Providing his evaluation of a course, one student suggested that the professor 'needs to teach more, not just assign projects and expect us to do them on our own.' Regarding the implementation of a writing component within some of his courses, another student explained his attitude: 'I feel like writing is not what I went to school for. I felt that my major concern in college was to learn mechanical engineering concepts.' USC experience has shown that students perceived a task more favorably if they understood how it could benefit them in seeking a job during and after the completion of their engineering degree. Such is the case for writing of proposals or memos. For example, a senior provided the following synopsis:

> Based on my limited work experience (co-ops, internships, etc.) I have found that engineers in general place a high value on the written document. I agree with my field that conveying thoughts on paper is equally as important as analyzing data. After all, the greatest discoveries are useless if no one else can comprehend your description of them.

Faculty experience in implementing change suggests that providing a comprehensive explanation of the assignment at the beginning of the semester can reduce student anxiety and provide structure for the new process [5]. Experience at USC affirms these previous findings. For example, the Director of the Professional Communications Center attempts to reduce student resistance to oral and written forms of assessment by presenting information early in the semester in a non-threatening manner. One method used to introduce writing into the curriculum is to make presentations in the engineering classrooms and to engage students in informal discussions. Such methods have helped to reduce student anxiety and increased their understanding of the relationship between the new student competencies and workplace expectations.

New relevance

Engineering alumnae/alumni endorsed experiential learning and performance-based assessments as relevant to their work in business and industry. Former students stated that their lab experiences, design projects, oral presentations, and writing skill development were the most useful learning activities in helping them prepare for the engineering profession. Alumni/alumnae expressed that courses with 'hands-on' activities and assessments more accurately replicate the 'real world' situations of the engineering that they experience. One mechanical engineering graduate stated:

> I feel that all of the classes and activities that generated a hands on experience are the major aspects of my undergraduate program that have contributed to my satisfaction in the engineering field. Though the lab-oriented classes are time consuming, there is no lesson like the one you can relate to by doing it yourself. Courses that emphasize design projects, CAD, and computer analysis are some of the aspects that I have found valuable as a young engineer. There have been many occasions where I could say to my supervisor that I could perform a particular job because I have had a chance to work in an area before.

Another alumnae discussed the value of the lab and her evaluation of the opportunity:

> Unit operations lab. This was a tremendous learning experience. [The] lab has prepared me for engineering tasks which I encounter on a daily basis. Textbook education, coupled with hands-on, real life experiences helped to create an awesome learning environment. Engineering is not all textbook and predetermined calculations and equations and we caught a glimpse of the real-world engineering as

students in this class. This way, I had a true understanding of what I was going to be getting into.

The pattern of responses suggests that USC alumnae/alumni believe that a beneficial learning environment blends the workplace and teaching. Examples of how this process can be accomplished, according to former students, includes conducting seminars with industry professionals and bringing equipment into the classroom or organizing field trips to view it. As one graduate emphasized, relating theory to practical applications through the use of examples 'would be beneficial to both the student's understanding and ability to apply this knowledge.' Another former student emphasized, 'It is easier to learn a subject when you know where it is used in the real world.'

New skills

Implemented to improve student writing, the Professional Communications Center (PCC) benefited students through increasing their awareness of the importance of oral and written communication in the marketplace. The USC College of Engineering PCC offers an oral and written communications program that emphasizes the practice-to-learn instructional approach. The PCC provides several advantages:

- a mechanism for integrating student instruction and consultation into the regular engineering course offerings;
- a feedback opportunity for individual students who seek assistance;
- in-class presentations by PCC experts;
- appropriate strategies for assessing writing and oral presentations;
- an easily accessible location for consultation.

For faculty, the PCC offers:

- the opportunity for collaboration in implementation of written and oral communication projects within the classroom;
- a consulting resource when help is needed for grant proposals and journal articles.

For the College, it also offers an effective liaison between engineering programs and the English department and its writing programs.

When seniors were asked their opinion regarding the amount of experience they received in oral communications, 79% or more of the students indicated they received an adequate amount of practice in their engineering coursework. Approximately 15% stated they could have benefited from more oral communications instruction and/or practice. When rating the amount of training provided in the area of written communications, approximately 72% of the seniors believed they received an adequate amount of experience in writing technical reports, memos, proposals, etc. About 11% of the students, however, reported that there was not sufficient experience provided in this area.

Seniors also indicated their satisfaction with the level of competency they achieved in oral and written communications as a result of their college education. Regarding oral communications, 90% of the students were 'satisfied' or 'completely satisfied' with their level of competency. Likewise, approximately 91% gave a positive rating for written communication skills.

Seniors cited oral and written communication experiences as one of the program components that they found most useful in their preparations for becoming an engineer. When asked for recommendations to improve the educational experience for future engineering students at USC, one student provided this comment: 'Writing and presentation skills should continue to be stressed heavily if not more heavily.' The increased presence of PCC personnel within engineering classrooms makes this recommendation a realization.

LESSONS LEARNED IN EDUCATIONAL RESEARCH

A continuous quality improvement process, as stipulated in the ABET criteria, requires design of an evaluation that documents progress towards achievement of program goals and involves staff in the use of evaluation information for program improvement. To achieve this, an evaluation must be aligned with the curricular emphasis that is placed on active learning, a mode of instruction that reflects the real-world experiences of engineers. More specifically, the assessments used in the evaluation must address the complex skills, such as designing an experiment, that students gain through active learning; and it is unlikely that a traditional pencil-and-paper examination will adequately measure such skills.

Two research studies related to the use of performance assessments in program evaluation form the basis of the following discussion. In one study, the staff of a family literacy program designed and implemented a portfolio assessment to document changes in the parenting practices of participants [7, 8]. In the other study, a series of investigations examined the effect of scoring practices on the reliability of essay scores [9–12]. A synthesis of the findings from the two studies offers implications for designing a program evaluation that is aligned with the curricular emphases described in Lessons Learned in Engineering Education, above, as well as implications for improving the quality of performance-based assessments used in the evaluation.

The technical quality of performance-based assessments for use in program evaluation benefited from the collaborative effort of a multidisciplinary team

ABET standards for student outcomes are performance-based, and a key component in the design of performance-based assessments for program evaluation is a collaborative, multi-disciplinary team. Faculty in a discipline, such as engineering, contribute an understanding of the important knowledge and skills that capable students should be able to demonstrate, and team members with expertise in assessment and evaluation contribute specialized knowledge necessary for [13]:

- aligning the evaluation with the agency's mission and goals;
- providing the technical support in the implementation and review of the measurement system;
- interpreting results for making program decisions.

For example, skills such as construct definition and instrument development are necessary for devising assessments to be used in an evaluation; however, there is little evidence that program staff or faculty bring these skills to an evaluation. Indeed, as mentioned earlier, a needs assessment conducted by USC's College of Engineering found that few professors indicated any prior training in assessment.

An example of the benefit of a multidisciplinary team is provided in the evaluation of the family literacy program. In this study, staff and evaluators collaborated to design a portfolio assessment to document the program's impact on the development of literacy and parenting skills of participants [8]. In each step, family literacy staff contributed substantive knowledge about parenting skills that participants should develop as a result of their involvement in the program. The evaluators contributed to the understanding of the measurement issues necessary in designing an assessment that could provide the staff with information for making decisions about program impact. As such, the staff and evaluator collaborated to [8]:

- identify prerequisite assessment skills and appropriate professional development activities required for completing the task;
- establish goals for assessing development of family literacy;
- identify types of artifacts appropriate to include in the portfolio;
- monitor the development of the portfolios;
- create record-keeping documents;
- develop analytic and holistic scoring rubrics;
- score the portfolios.

The collaboration of stakeholders in the development of the goals and the implementation of the assessment provided several benefits. In response to a survey, staff members indicated that the portfolio assessment served to focus program activities on the goals that the staff had identified as being important for parents to develop. They also indicated that the collaboration increased their understanding of assessment methods, such as developing rubrics (i.e., analytic and holistic), and that they were able to share assessment ideas with their colleagues. Staff involvement in the evaluation appeared to also increase the staff's sense of ownership in the evaluation as theorized by writers Barrick and Cogliano [14] in the field of evaluation. Indicative of their sense of ownership, program staff began to assume greater responsibility for assuring the quality of the assessment by, for example, designing a portfolio log to guide the selection of items that were appropriate for inclusion in the portfolio. The comprehensive assessment systems implemented by engineering faculty may also profit from similar collaborative efforts.

The collaborative development of the evaluation and its concomitant assessments precluded many of the common pitfalls of evaluation

Based on their experience in evaluating a nursing education program, Barrick and Cogliano [14] described problems that may occur when stakeholders are not involved in the creation of an evaluation framework. Such problems included a lack of consensus about the goals of the evaluation and, subsequently, no implementation of the evaluation plan. When issues about the evaluation of the parent literacy program were raised, staff concerns were addressed by reviewing the fact that staff had collaborated with the evaluator to establish the goals, determine the methods of data collection, and design the scoring procedures. Resistance to the evaluation was avoided, and staff focused their efforts on implementing necessary changes to improve the evaluation. The collaboration thus avoided many of the pitfalls that can be encountered in evaluations where staff are less involved. Collaboration between evaluators and faculty in the development of a continuous quality improvement process for engineering may prove similarly beneficial.

When rater consistency for assigning scores to open-ended assessments was examined for both holistic and analytic rubrics, the highest level of inter-rater reliability was achieved when scores were based on summing ratings across the evaluative dimensions of the analytic rubric

The Program Evaluation Standards [15] are prescribed by the professional organizations associated with education. These standards stipulate that the reliability of new assessments must be addressed as the measures are incorporated into an evaluation. In the case of open-ended tasks, such as lab reports, raters must score student responses and the raters' subjective decisions potentially introduce measurement error into the assessment, reducing reliability (i.e. consistency) of those scores. If inter-rater reliability is low, then it is possible that were another rater to score the student response, the decision would differ. In the use of unreliable assessment results to make programmatic decisions, separating differences in student performance due to rater judgement from those differences attributable to program impact can be problematic.

To improve reliability of rater decisions in the scoring of essays, student responses are generally scored by two or more raters, a practice followed by 78% of the state departments of education that included open-ended items in their testing programs.9 Raters typically scored open-ended items by the application of a scoring rubric consisting of 4 to 6 points. These rubrics are categorized into two types: analytic or holistic. An analytic rubric separately lists each skill, or evaluative criterion, to be scored in a task. For example, in the study of essays, an analytic rubric listed four evaluative criteria [16]:

- content/organization,
- style,
- conventions of written language,
- sentence formation,

and raters awarded a score for each criterion. In contrast, the evaluative criteria in a holistic rubric are presented collectively in a descriptive narrative. Thus, in scoring with a holistic rubric the evaluative criteria, such as topic development, sentence formation, and word choice, are evaluated simultaneously and one score awarded to reflect the skill level demonstrated in the student response.

In one investigation of reliability in scoring essays, an inter-rater reliability of 0.83 (as measured by a Spearman correlation) was achieved when two trained raters scored essays with a 6-point holistic rubric [11]. It was estimated that three raters were required to achieve a reliability of 0.88, and that little could be gained by adding a fourth rater because inter-rater reliability only increased to 0.91. In another investigation, two raters scored essays using the previously-described analytic rubric. Each of the four evaluative criteria, or analytic domains, received a score ranging from 1 (Inadequate), 2 (Minimal), 3 (Good), to 4 (Very Good) [12]. Reliability for the domain-level scores were similar to the holistic rubric: 0.78 (Style), 0.80 (Sentence Formation), 0.83 (Content), and 0.85 (Conventions). Inter-rater reliability, however, rose to 0.90 when scores were added across domains and those total scores correlated between raters.

In the family literacy study, raters first scored the portfolios using an analytic rubric comprised of six program goals. Raters then applied a holistic rubric to obtain an overall rating that reflected the participant's level of family literacy. Inter-rater reliability of the goal-level scores (also referred to as domain scores) ranged from 0.65 to 0.83 [7]. When scores were summed across the goals, the reliability of the total score was 0.86. The reliability of the summed score was slightly higher than that of the holistic score ($r = 0.82$). Other studies have also found that combining analytic scores across goals or domains improves the reliability of portfolio scores [17]. Thus, evaluations of engineering programs should consider the use of several raters in combination with analytic rubrics to produce reliable scores when using performance assessments, such as design projects, to measure student outcomes.

In the scoring of performance assessments, inter-rater reliability also improved by allowing raters to augment integer-level scores

Augmenting scores is analogous to the use of *plus* and *minus* in the awarding of letter grades; augmentation allows the rater to increase the score using a '+' if the paper appears a bit superior to the exemplar papers and to lower the score using a '−' if the paper appears a bit inferior to the exemplars. For instance, a student interpretation of lab data may be sufficient for receiving a 4 on a 4-point scale, but may omit some minor points of analysis, suggesting that additional interpretation skills are needed. Such a performance could possibly warrant a score of 4− instead of 4.

Penny, Johnson, and Gordon [11] demonstrated that the use of augmented scores can increase indices of inter-rater reliability. Thus, augmentation may improve the reliability in the scoring of the performance tasks that are used to monitor student outcomes in the evaluation of engineering programs.

When raters initially assigned different scores to a student response, inter-rater reliability was most improved by obtaining the judgment of a third rater and averaging the three scores

If two or more raters independently score an open-ended task and assign different scores, one of several things can happen to resolve the score discrepancy. Usually, this resolution process involves an additional review by a third, more expert, rater. Often the decision of this expert determines the final score of the paper. In other cases, the original raters discuss the components of the paper and mutually decide what score is appropriate.

Johnson et al. [9] in a review of state testing agencies identified several methods by which the disagreements of raters were resolved. These included:

- averaging the two raters' scores;
- substituting the rating of a third expert rater for the two original scores;
- averaging the scores of a third expert rater and the original two raters;
- averaging the score of a third expert rater with the original rating that was closest to that of the expert.

In an empirical study of these score resolution methods, Johnson, Penny, and Gordon [12] offered evidence suggesting that the highest reliability could be achieved by averaging the three scores (third method above). Hence, when raters assign different scores to performance assessments, it appears that program evaluations should solicit the independent judgement of a third expert rater and average the three scores.

IMPLICATIONS FOR RESEARCH

As engineering programs move to incorporate the types of activities that students will experience in the workplace, methods for assessing students are likely to become more open-ended, reflecting the problem-solving aspects of the ABET criteria. This investigation found support for experiential learning among former students who are now practicing engineers. If assessment systems are to document changes in students and programs, such systems also need to include open-ended assessments that parallel classroom activities. For these assessments to provide useful information for program planning in engineering, the data should be reliable. Lessons from educational research and measurement offer implications for best practices in scoring. These lessons provide possible directions for research-based investigations that improve the assessment of student outcomes in engineering.

REFERENCES

1. ABET, *Engineering Criteria 2000: Criteria for Accrediting Programs in Engineering in the United Sates*, 3rd ed., Engineering Accreditation Commission, Accreditation Board for Engineering and Technology, Inc. Baltimore, MD, December 1997.
2. E. Ernst, Review of reports, studies, and conferences on engineering education 1981–1997, *Proceedings from Realizing the New Paradigm for Engineering Education, Engineering Foundation Conference*, (June 3–6, 1998).
3. John W. Prados, Action agenda for systemic engineering education reform: next steps, *Proceedings from Realizing the New Paradigm for Engineering Education, Engineering Foundation Conference*, (June 3–6, 1998).
4. J. McGourty, C. Sebastian and W. Swart, Developing a comprehensive assessment program for engineering education, *J. Eng. Educ.*, **87**, 4, (1998) pp. 355–361.
5. A. Christy and M. Lima, M. The use of student portfolios in engineering instruction, *Journal of Engineering Education*, **87**, 2 (1998) pp. 143–148.
6. B. Olds and R. Miller, Portfolio Assessment: Measuring Moving Targets at an Engineering School, *NCA Quarterly*, **71**, 4 (1997).
7. R. Johnson, F. McDaniel and M. Willeke (at press) Using portfolios in program evaluation: an investigation of interrater reliability, *Amer. J. Evaluation*.
8. R. Johnson, M. Willeke and D. Steiner, Stakeholder collaboration in the design and implementation of a family literacy portfolio assessment, *Amer. J. Evaluation*, **19**, 3 (1998) pp. 339–353.
9. R. Johnson, J. Penny and C. Johnson, *Score resolution in the rating of performance assessments: practices and issues*, presented at the Council of Chief State School Officers National Conference on Large Scale Assessment (June, 1998).
10. R. Johnson, J. Penny and B. Gordon, *Score resolution methods and the reliability of holistic scores*, presented at the Annual Meeting of the American Educational Research Association, Montreal (April, 1999).
11. J. Penny, R. Johnson and Gordon, *Score augmentation and the reliability of operational scores: an empirical study of a holistic rubric*, presented at the Annual Meeting of the American Educational Research Association, Montreal (April, 1999).
12. R. Johnson, J. Penny and B. Gordon, (in press) The relationship between score resolution methods and score reliability: an empirical study of an analytic scoring rubric, *Applied Measurement in Education*.
13. R. O'Sullivan, From judges to collaborators: Evaluators' roles in science curriculum reform. In R.G. O'Sullivan (Ed.), Emerging roles of evaluation in science education reform, *New Directions for Program Evaluation*, **65** (1995) Jossey-Bass: San Francisco, pp. 19–29.
14. C. Barrick and J. Cogliano, Stakeholder involvement: mythology or methodology? *Evaluation Practice*, **14**, 1 (1993) pp. 33–37.
15. The program evaluation standards (2nd ed.) Joint Committee on Standards for Educational Evaluation, Thousand Oaks, CA: Sage (1994).
16. Georgia Department of Education, *Georgia High School Writing Test: Assessment and Instructional Guide* (1993).
17. D. Koretz, B. Stecher, S. Klein and D. McCaffrey, The Vermont portfolio assessment program: Findings and implications, *Educational Measurement: Issues and Practice*, **13**, 3 (1994) pp. 5–16.

Susan Dabney Creighton is the Director of Assessment for the College of Engineering and Information Technology. She is completing her Ph.D. in the College of Education at the University of South Carolina. She received a M.Ed. in Educational Research in 1977. Prior to her employment with the College of Engineering, Susan worked with various state agencies and school systems as an evaluator of federal and state grants as well as coordinating the implementation of continuous quality improvement systems.

Robert Johnson is Assistant Professor in the Educational Psychology Department at the University of South Carolina. He is also the Co-director of the Center for Excellence in the Assessment of Student Learning. His current research interests include the development and scoring of performance assessments.

Jim Penny is a Research Scientist with the Center for Creative Leadership in Greensboro, North Carolina. His research interests include the statistical demonstration of test fairness and educational measurement.

Edward W. Ernst received the BS, MS, and Ph.D. degrees in electrical engineering from the University of lllinois. He was a member of the faculty in Electrical and Computer Engineering at the University of lllinois from 1958 until his retirement in 1989. He was Program Director for Undergraduate Engineering Education at the National Science Foundation from 1987–90. He has been Allied Signal Professor of Engineering at the University of South Carolina since October, 1990. He was IEEE Vice President for Educational Activities for 1981–82. His service with ABET covers the period from 1977 to 1992 and includes: Chair, Engineering Accreditation Commission (1985–86); President (1989–90). He was the founding Editor of the ASEE Journal ofEngineering Education (1991–1996). He is a member and Vice-Chair of the International Engineering Consortium (IEC) Board of Directors. He is a Fellow of IEEE, AAAS, ASEE, ABET and IEC, and a member of Sigma Xi, Tau Beta Pi, Eta Kappa Nu and Phi Kappa Phi. Other awards he has received include the Halliburton Award for Leadership in Engineering Education (1983), the IEEE EAB Meritorious Achievement Award in Accreditation Activities (1985), the IEEE Education Society Achievement Award (1989), the Linton E. Grinter Distinguished Service Award from ABET (1992), the Distinguished Alumni Award from the Electrical and Computer Engineering Alumni Association (University of lllinois) in 1994, and the ASEE Recognition of Contributions as Editor of the Journal of Engineering Education (1996).