

DIFFERENTIATING ASSESSMENT FROM EVALUATION AS CONTINUOUS IMPROVEMENT TOOLS

Peter E. Parker¹, Paul D. Fleming², Steve Beyerlein³, Dan Apple⁴ and Karl Krumsieg⁵

Abstract — Unleashing the potential of continuous improvement in teaching/learning requires an appreciation of the difference in spirit between assessment and evaluation. Assessment is frequently confused and confounded with evaluation. The purpose of an evaluation is to judge the quality of a performance or work product against a standard. The fundamental nature of assessment is that a mentor values helping a mentee and is willing to expend the effort to provide quality feedback that will enhance the mentee's future performance. While both processes involve collecting data about a performance or work product, what is done with these data in each process is substantially different and invokes a very different mindset. This paper reviews the principles and mindset underlying quality assessment and then illustrates how feedback can be enhanced within the framework of structured assessment reports.

Index Terms — assessment, evaluation, SII reports

INTRODUCTION

Higher education has been under scrutiny from many sectors of society as standards of performance in industry and government continually increase. Change within institutions of higher education is necessary to meet these higher standards [1,2]. Educators need tools and processes to help them implement cultural change so future college and university graduates can successfully meet these challenges [3,4].

To some extent, it is possible to increase performance by merely raising standards and punishing students and teachers for not meeting the standards. However, achieving significant gains in performance capability can only be grown in a student-centered environment that provides real-time, customized feedback [5,6]. This choice between the carrot and the stick in achieving our educational objectives is humorously illustrated in Figure 1 [7].

As educators, we use a variety of assessment practices to help our students build lifelong learning skills. We also can cajole movement in the same direction through our evaluation practices. Assessment is the process of measuring a performance, work product or a learning skill and giving

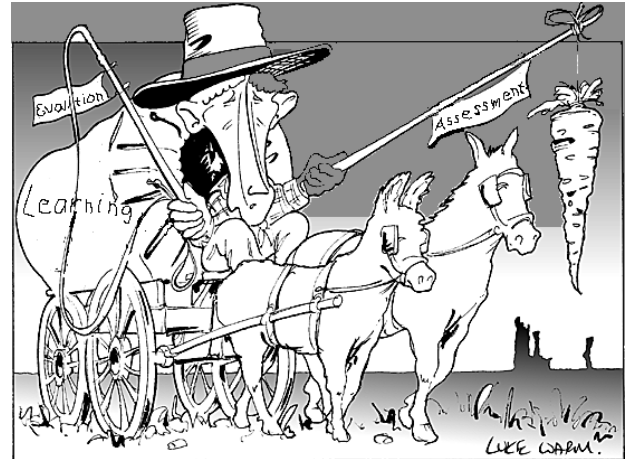


Figure 1
OPTIONS FOR IMPROVING LEARNING OUTCOMES

feedback, which documents growth and provides directives to improve future performance. Evaluation is a judgment or determination of the quality of a performance, product, or use of a process against a standard.

The distinction between assessment and evaluation can be seen in statements endorsed by the American Association for Higher Education (AAHE), the American Education Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME) [8,9]. Assessment is widely recognized as an ongoing process aimed at understanding and improving student learning. Assessment is concerned with converting expectations to results [10]. Evaluation is recognized as a more scientific process aimed at determining what can be known about performance capabilities and how these are best measured. Evaluation is concerned with issues of validity, accuracy, reliability, analysis, and reporting [11]. Quality assessment and evaluation are similar in that they both involve specifying criteria and collecting data/information. In most academic environments, they are different in purpose, setting criteria, control of the process, and response. Table 1 compares assessment and evaluation in each of these areas. Reflect on the contrasting roles for the assessor, assessee, evaluator, and evaluatee. Also note the implied relationship between the assessor and assessee.

¹ Peter E. Parker, Western Michigan University, College of Engineering and Applied Sciences, Kalamazoo, MI 49008, peter.parker@wmich.edu

² Paul D. Fleming, Western Michigan University, College of Engineering and Applied Sciences, Kalamazoo, MI 49008, dan.fleming@wmich.edu

³ Steve Beyerlein, Mechanical Engineering Dept, University of Idaho, Moscow, ID83844, sbeyer@uidaho.edu

⁴ Dan Apple, Pacific Crest, Corvallis, OR 97330, dan@pcrest.com

⁵ Karl Krumsieg, Pacific Crest, Corvallis, OR 97330, karl@pcrest.com

TABLE 1
COMPARISON BETWEEN ASSESSMENT AND EVALUATION

	Assessment	Evaluation
Purpose	To improve future performance	To judge the merit or worth of a performance against a pre-defined standard
Setting Criteria	Both the assessee and the assessor choose the criteria.	The evaluator determines the criteria
Control	The assessee --- who can choose to make use of assessment feedback	The evaluator --- who is able to make a judgement which impacts the evaluatee
Depth of Analysis	Thorough analysis by answering questions such as why and how to improve future performance	Calibration against a standard
Response	Positive outlook of implementing and action plan	Closure with failure or success

Assessment can be done anytime, by anyone, for anyone. The role of the mentor is to facilitate the mentee's success through quality feedback. The fundamental nature of assessment involves an assessor (mentor) expending the effort to provide quality feedback that will enhance an assessee's (mentee's) future performance based upon the needs expressed by the assessee. The assessor can be a single person, a group, or oneself. Many people can perform an assessment for a single person, or a single person can perform an assessment for many people. The focus of an assessment can be on a process, the resulting product or outcome from a process, or a specific skill or set of skills related to a process. In an educational context, students may be assessed with respect to the quality of learning they produce (their content knowledge), the quality of their use of a learning process, or with regard to the strength of specific skills which support a learning process. The timing of an assessment can vary with respect to the focus of the assessment. An assessment can be made in real-time (as the performance is occurring), at discrete intervals (formative assessment), or at the end of a process (summative assessment). The quality of an assessment process is a function of the following variables:

- the knowledge and skill level of the assessor with respect to the assessment process
- the expertise of the assessor with respect to what is being assessed (content)
- the preparation time available for designing an assessment, and

- the level of trust between the assessee and the assessor.
- The ability to enhance meaning through analyzing assessment data

This paper explores a method for enhancing the quality of assessment findings by the assessor and the usefulness of assessment reports given to the assessee. Key elements of the method are a standardized report format and a rubric for classifying the quality of assessment feedback. The method has been validated through classroom use with pre-engineering through graduate students.

ASSESSMENT METHODOLOGY

A quality assessment process has four steps: (1) develop guidelines for the assessor to follow when assessing a performance, (2) design methods used for the assessment, (3) collect information about the performance or work product, and (4) report findings to the assessee. Executing each of these steps insures that the assessor and the assessee agree upon the purpose of the assessment, what is to be assessed, and how the assessment is to be reported.

The first step in setting up an assessment is to define the purpose for the performance and the purpose for the assessment. With this information, the person being assessed (assessee) can better determine what is important to assess, and the person doing the assessment (assessor) is equipped to give correct and appropriate feedback.

In designing a method for assessment, both parties

should collaborate to generate a list of possible criteria that could be used by the assessor to give feedback to the assessee. From this list, both should agree and select the most important criteria that best meet the guidelines from the first step of the methodology. In most cases, this list should contain no more than four criteria. For each chosen criterion, determine appropriate factors to assess the performance and the appropriate scale to measure or determine the quality of each chosen factor. Note that in some cases, where the assessment is more spontaneous or narrowly focused, the criteria may be manageable enough without defining factors.

While the assessee is performing, the assessor must collect information consistent with the chosen criteria and factors. It is important for the assessor to note: (1) the strong points of the assessee's performance (things done well) and why they were considered strong, (2) the areas in which the assessee's performance can improve, along with how the improvement could be made, and (3) any insights that might help the assessee. By using this structure for data collection, the likelihood that positive changes in behavior of the assessee, as a result of the assessment, will increase.

The final step of the methodology is for the assessor to report findings to the assessee in as positive a manner as possible. We have found it valuable to structure these in an SII report that consists of three parts (Strengths, areas for Improvement, and Insights) [12].

- **Strengths** identify the ways in which a performance was of high quality and commendable. Each strength should address what was valuable in the performance, why this is important, and how to reproduce this aspect of the performance.
- **Areas for Improvement** identify the changes that can be made in the future (between now and the next assessment) that are likely to improve performance. Improvements should recognize the issues that caused any problems and mention how changes could be implemented to resolve these difficulties.
- **Insights** identify new and significant discoveries/understandings that were gained concerning the performance area; i.e., what did the assessor learn that others might benefit from hearing or knowing. Insights include why a discovery/new understanding is important or significant and how it can be applied to other situations.

The three elements of SII reports parallel the findings of Howard Gardner in his book [Extraordinary Minds](#) [13]. Based on extensive interviews, he noted that

extraordinary people stand out in the extent to which they reflect—often explicitly—on the events of their lives, the extent to which they consciously identify and then exploit their strengths, and the extent to which they their flexibility/creativity in learning from setbacks.

Elevating the quality of assessment feedback is critical to improving performance. The following rubric has been developed to enhance SII reports. Four levels are presented corresponding to the first four levels in Bloom's taxonomy[14]. Students and faculty can use this rubric to rate the quality of self- and peer-assessments and to suggest ways that assessment feedback can be made even more valuable.

Level I (Information)

Strengths and improvement areas are presented as simple statements or observations. Insights are rare.

Level II (Comprehension)

Strengths and improvements are clearly stated and the reasons for the strength and suggestions for improvement are given. Insights tend to be related to the specific context of the assessment.

Level III (Application)

The feedback now helps the learner to process future new contexts by illustrating application to additional contexts.

Level IV (Problem Solving)

The feedback now has evolved such that the learners comprehend the multiple uses in which they can leverage strong points, areas for improvement and insights across a variety of contexts through their newly gained expertise.

As assessments move up the scale from Level I, there is an observable shift from assessing effort to assessing performance. As a result, students tend to focus more on performance criteria and less on time spent as they receive assessment feedback at higher levels.

CLASSROOM APPLICATIONS

Angelo and Cross[5] present 50 classroom assessment techniques. Not all are useful for all classes and no instructor should attempt to use all of them in one course. All, however, are focused on helping the student learn. The SII assessment approach is readily adaptable to their techniques.

One of us (Parker) used the SII methodology with a modified form of Angelo and Cross' "Group Work Evaluation" to drive improvement in oral presentations. The class consisted of 24 first semester freshmen.

TABLE II
FRESHMAN SEMINAR ASSESSMENTS

Strengths	<ul style="list-style-type: none"> • Key points at the beginning of the report. This was helpful because it gave a good foundation for the rest of the report. (Level II) • All spoke clearly, so the audience was able to hear and be more involved. (Level II) • Good presentation setup. This made the presentation flow and made everything easier to understand. (Level I)
Improvements	<ul style="list-style-type: none"> • You may want to expand more than what is on the screen. Sometimes only the bullet points were read and not expanded upon. This would help others have a better understanding of the subject. (Level III) • Could have explained words better to allow audience to have a better understanding of the material. (Level I) • Some team members didn't have good eye contact. This made the audience feel less included. (Level II)
Insights	<ul style="list-style-type: none"> • I didn't know polymers were used in the medical industry. (Level I) • Group members shouldn't wander around while others are speaking. (Level I)

chemical engineers and 8 sophomore level transfer students. The class was divided into 9 teams of 2 - 4 students, which were permanent for the semester. (The target was 4 students per team for the original 36 enrolled student, but withdrawals dropped 1 team to 2 and 2 teams to 3 members.) Each team was assigned two projects during the course of the semester --- the first project was to research a particular chemical company (manufacturer or supplier) and the second was to research the production of a particular product (e.g. polyethylene). Each team prepared a written report, which was evaluated for both technical content and grammatical correctness. The team also presented their report to their classmates. Each presentation was then assessed by the instructor and the other teams. Both the teams and the instructor used the SII format to report their findings. All assessment material was collected and given to the presenting team. Typical responses from the second set of assessments are given in Table II. The level of assessment as estimated from the rubric is also given. Supplying this classification can help both the assessor and assessee identify specific actions and critical thinking questions they should consider in future performance.

The student assessments from the first presentations were primarily at Level I. Statements of strengths and improvement areas were simply listed without clarifying information. As is evident from Table II, students are starting to elevate their level of assessment by giving reasons why the particular item is a strength or area for improvement. In addition, the insights indicate that students are internalizing both information delivered and things about the process of giving a presentation. However, the level of critical thinking displayed in the insights is usually less than the strengths and improvements. From the instructor's vantage point, groups

that tended to deliver strong assessments were also strong performers. Groups that delivered weak assessments tended to give weak performances. End of course feedback indicates that the students felt that the assessment process helped them become better presenters of information.

Another of us (Beyerlein) has used the SII technique in a sophomore design course. The syllabus and course expectations are clearly stated and use the terms of EC 2000 (in particular, measurable outcomes) as well as various terms from Process Education, such as a Knowledge Map for the course and some indications of expected behavior changes (Way of Being). The design project is a vehicle for the students to apply various knowledge items (such as mathematical analysis; team dynamics) to a practical situation. One measurable outcome of the project was a score related to the performance of the design (in this case, a rocket to loft popcorn to a given height). One of the assessment areas was the project presentations. Table III presents the range of SII assessments—two from students and one from a senior who assisted as a class mentor.

The first two assessments in Table III illustrate the difference between Levels I and II. The first assessment is a simple restatement of knowledge and provides little help to the assessee for future growth. The second assessment indicates that the assessor has comprehended the assessment process and is providing feedback that will help the assessed team grow in future presentations. The third assessment (from the senior student) provides an example of a Level IV assessment and focuses not just on the current context, but offers pointers to future growth across many contexts.

Another of us (Fleming) used similar methods to assess individual projects in a graduate course. This

course on Nonimpact Digital Printing consisted of a diverse group of five students. Three were Paper and Imaging graduate students, two with paper backgrounds and one with a printing background. One of the students was a Law Faculty member, who specializes in cyber law and the other was an unconventional student with a Master's degree in accounting, who has an ongoing interest in printing. In this course, each student wrote a paper and made a presentation on a topic related to digital printing. The presentations were assessed by both the instructor and the students. The final evaluations of these work products were done by the instructor, taking into account the assessments. Typical responses of seminar participants are given in Table IV.

These assessments tend to be at Levels I and II with minimal exposition on the rationale behind the assessment. Higher level of content mastery therefore does not directly translate into ability to spontaneously give high quality feedback. Like many faculty in teaching workshops, these students had no prior experience with the SII model. It is not surprising that their results are initially at Level I. However, the ability of these individuals to elevate the level of their assessments likely exceeds that of the underclassmen involved in the other two classes. Nonetheless, quality assessment, like Aristotle's concept of virtue, is something that is achieved through exposure to quality assessment, practicing assessment (self and others), and being open to

TABLE III
DESIGN MIN-PROJECT ASSESSMENTS

Strength	The presentation was short. (Level I)
Improvement	The presentation could have been made more interesting. (Level I)
Insight	The score was not the only goal. (Level I)
Strength	The presentation precisely summarized the important design features. (Level II)
Improvement	The presenter's lack of enthusiasm detracted from the message. (Level II)
Insight	The team kept the spirit of the competition in mind, not just the score. (Level II)
Strength	A PowerPoint slide with bullets about key functional requirements, constraints, and manufacturing considerations is an efficient way to summarize a design solution. (Level III)
Improvement	By thoughtfully using one's voice, eye contact, and body movements, the delivery can actually enhance your message. (Level IV)
Insight	Communicating your interpretation of the underlying purpose of a competition helps all assess whether they could have learned more from the activity. (Level III)

TABLE IV
GRADUATE SEMINAR ASSESSMENTS

Strengths	<ul style="list-style-type: none"> • Good organization and clarity. (Level I) • Good transitions and visual aids. (Level I) • Comfortable presentation style. (Level I)
Improvements	<ul style="list-style-type: none"> • Assure that you're speaking to the level of the audience. (Level I) • Need to rely less on notes and trust one's knowledge more. (Level II) • Include links to Web sites in presentation (if network connection is available). (Level II)
Insights	<ul style="list-style-type: none"> • Relationships between patents and nonimpact printing. (Level I) • Few web sites provide detailed technical information about nonimpact printing. (Level I)

assessment by others. It is most thoughtful and most instructive when it is a day-to-day part of the classroom culture.

CONCLUSIONS

One of the driving forces for change in higher education is the need to develop students who are life-long learners so they can adapt to the ever, and rapidly, changing world around us. Quality self-assessment provides a solid foundation for such self-growth, as we need to know our strengths and areas for improvement (change). Regular use of assessment activities gives learners the practice and experience they need to become quality self-assessors and self-growers. Based on our experience, quality assessment can be implemented relatively easily at any level in the curriculum through SII reports. In these activities students have an opportunity to take more control over their own learning and instructors have an opportunity to coach them on their performance. We have found that students to learn more in an assessment culture, are more motivated to perform better, and seek to improve their own performance. This mindset provides the faculty with the opportunity to assume the role of mentor in the learning process.

REFERENCES

- [1] Gardner, D. P., et. al., *A Nation At Risk: The Imperative For Educational Reform*. Report to the Secretary of Education. 1983. Available at <http://www.ed.gov/pubs/NatAtRisk/letter.html>
- [2] Boyer, E., *Scholarship Reconsidered: Priorities of the Professoriate*. Carnegie Foundation for the Advancement of Teaching, 1990.
- [3] Cappelli, P., "Colleges, students, and the workplace: Assessing performance to improve the fit," *Change*, Vol. 14, p 55-61, 1982.
- [4] Ewell, P., "Assessment: Where are we? The implications of new state mandates," *Change*, Vol. 19, p 23-28, 1987.
- [5] Angelo, T., and Cross, K. *Classroom Assessment Techniques* Jossey-Bass. San Francisco, CA 1993.
- [6] Neff, G., S. Beyerlein, D. Apple, K. Krumsieg, "Transforming Engineering Education from a Product to a Process", Fourth World Conference on Engineering Education, October, 1995.
- [7] Adapted from Luke Warm, *Times Higher Education Supplement*, 17 October 1997.
- [8] Astin et al, "9 Principles of Good Practice for Assessing Student Learning," www.aahe.org/assessment/principi.htm
- [9] AERA, "Position Statement Concerning High-Stakes Testing," www.era.net/about/policy/stakes.htm.
- [10] Angelo, T., "Reassessing (and Defining) Assessment. AAHE Bulletin, Vol. 48, No. 2, pp 7-9, 1995.
- [11] *Standards for Educational and Psychological Testing*, American Psychological Association, 1999.
- [12] K. Krumsieg and M. Baehr, *Foundations of Learning*, Third Edition, Pacific Crest, Corvallis, OR, 2000
- [13] Gardner, H., *Extraordinary Minds*, Basic Books, New York, NY, 1998.
- [14] Bloom, B. S. (ed.), *Taxonomy of Educational Objectives: The Classification of Educational Goals, Handbook 1: Cognitive Domain*, McKay, New York, 1956