Embedded Assessment: A Measure of Student Learning and Teaching Effectiveness

Maureen A. McCarthy,1 Daniel M. Niederjohn,1 and Theodore N. Bosack2

Abstract
Evidence of student learning has increasingly become the focus of external constituents, yet methods of documenting student learning outcomes are considered expensive, onerous, or both. This article provides a brief review of the limitations associated with traditional student opinions or course evaluations as evidence of teaching effectiveness, and it suggests a second strategy for evaluating student learning. In addition, we propose a model for using existing assessments to provide evidence of student learning and teaching effectiveness.

Keywords
assessment, evaluation, student learning outcomes

External constituencies such as the Council on Higher Education Accreditation, federal legislators, and state legislators are calling for evidence that students graduating with a bachelor’s degree possess basic skills (e.g., communication, quantitative reasoning, and information literacy). Increasingly, this call for accountability affects departments, and ultimately increases pressures on individual faculty to document effective teaching (Middaugh, 2001). Documentation of student learning is important not only for responding to external constituents but also for helping to improve teaching and learning through an iterative process.

How do we respond to external constituents and improve teaching effectiveness? Faculty and administrators often turn to the time-honored student evaluations as a key measure of teaching effectiveness. Although student evaluations provide information that reflects student opinion about the course, student feedback does not provide evidence of learning. McKeachie (1997) cautioned against ignoring student ratings entirely, and he cited major validity problems associated with use of mean scores as a method of assessing teaching effectiveness. Despite overwhelming evidence that student evaluations are not an effective measure of student learning or teaching excellence (Ballard, Rearden, & Nelson, 1976; Beyers, 2008; Buskist, Keeley, & Irons, 2006; Germaine & Scandura, 2005; McKeachie, 1997), student ratings persist as a primary measure of teaching effectiveness. In other words, evaluation of teaching efficacy using current practice does not typically include an assessment of student learning outcomes. If, instead, we expand the definition of teaching effectiveness to include a measure of student learning, then student evaluations or opinions about the course are woefully inadequate as a valid measure of student learning (Ballard et al., 1976; Buskist et al., 2006).

Despite the inadequacies inherent in using student evaluations for purposes of faculty review (Eckert & Dabrowski, 2010), student opinions remain among the most widely used measures of determining teaching effectiveness and merit. Nevertheless, federal legislation continues to increase the emphasis on outcomes as a measure of teaching effectiveness. Resnick (2009) and LoSchiavo, Shatz, and Devereaux (2008) contend that the emphasis on student learning outcomes is related to funding in the Race to the Top initiative and in the future reauthorization of the Elementary and Secondary Education Act. In fact, states must: (a) create assessments that are matched with college readiness, (b) ensure that students have effective teachers, and (c) implement longitudinal measures to assess student progress and teacher effectiveness (Resnick, 2009). If federal regulations are already tightly linked to student performance in the K-12 system, can higher education be far behind?

Eckert and Dabrowski (2010) review the practice of evaluating teachers based on student performance, and they conclude that value-added measures are essential. If higher education is to respond to increased pressure to ensure student 1 Kennesaw State University, Kennesaw, GA, USA
2 Providence College, Providence, RI, USA

Corresponding Author:
Maureen McCarthy, Department of Psychology, Kennesaw State University Box 2202, Kennesaw, GA 30144-5591
Email: Mmmccar10@kennesaw.edu
learning, then it seems necessary to improve systematic measures of student learning as a measure of teaching effectiveness. Given the inadequacy of student evaluations as a measure of teaching effectiveness, we are proposing that embedded assessment is a second and more valid measure of student learning and faculty effectiveness. Embedded assessment is simply an analysis of the formative measures of student learning outcomes already present in the classroom (e.g., exams, assignments, papers). Allen (2004) suggested that embedded assessment of student learning offers a rich set of data that can more effectively measure faculty effectiveness and student performance. Embedded assessments also offer advantages over traditional methods. These advantages include directly tying student outcomes to course goals and testing progress toward the goals (student outcomes) at the same time as assigning student course grades, thereby saving time by reducing the need to generate separate instruments for multiple purposes.

**Specifying Student Learning Outcomes**

The first step in accurately evaluating student learning is to identify clearly what we want students to learn, in other words—specify student learning outcomes. Several resources can aid in developing course-based student learning outcomes including the *Guidelines for the Undergraduate Psychology Major* (American Psychological Association [APA], 2007); *Teaching, Learning, and Assessing in a Developmentally Coherent Curriculum* (APA, 2008); *A Rubric for Learning, Teaching, and Assessing Scientific Inquiry in Psychology* (Halonen, Bosack, Clay, & McCarthy, 2003); and locally developed goals and outcomes. The *Undergraduate Guidelines* offer broad goals for courses. The APA (2008) report on the *Developmentally Coherent Curriculum* offers specific guidance for developing course-specific student learning outcomes, and if a course includes significant emphasis on scientific reasoning, Halonen et al. (2003) provide specific developmentally appropriate student learning outcomes for the scientific reasoning domain.

In our embedded assessment model, we suggest that student learning outcomes can be used as the basis for measuring student learning and as the foundation for assignment of students’ grades. We also illustrate how evidence of student learning may be useful as a measure of teaching effectiveness. Our proposed embedded assessment model illustrates how to link student learning outcomes to existing assessment measures (e.g., exams, assignments, papers) as a way to evaluate student learning and provide faculty with self-evaluation data that will ultimately improve teaching effectiveness.

**Developing Student Learning Outcomes**

Using the aforementioned resources for student learning outcomes, we can derive specific criteria (i.e., student learning outcomes) that students should achieve in our courses. As the first step in planning for systematic assessment, selection of appropriate outcomes is paramount. Ideally, we not only identify student learning outcomes as a guide for our planning, but we communicate our expectations for student learning to students at the beginning of the semester by detailing specific student learning outcomes on our course syllabi. In our example, we specify developmentally appropriate learning outcomes for an experimental psychology course (see Table 1) as indicated in the rubric for scientific literacy (Halonen et al., 2003). We also derive student learning outcomes from campus-based expectations. We then link these outcomes to existing course assessments. Linking outcomes

### Table 1. Linkage of Student Learning Outcome to Assessment

<table>
<thead>
<tr>
<th>Student Learning Outcome</th>
<th>Exams</th>
<th>Paper</th>
<th>Lab Reports</th>
<th>Poster</th>
<th>Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Objective measures</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Explain the logic of hypothesis testing</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Operationalize independent and dependent variables</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use statistical programs to perform analyses</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Describe the benefits and limitations of statistical techniques</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Identify potential confounds in experimental studies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Apply statistical probability to statistical problems</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recognize and apply ethical research practices</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Describe relationships between theories, experimental, and correlational techniques</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td><strong>Qualitative measures</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Literature search</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Consistent use of APA style</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Use of formal and professional expression</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Develop plausible arguments to support research proposal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relate content from several sources</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Develop logical arguments for conclusions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Demonstrate effective poster presentation skills</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Articulate sound hypotheses</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Select and apply appropriate experimental design and statistical analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>
to assessments when planning a course helps faculty be clear about what students should learn and how student performance will be measured.

Some student learning outcomes are more easily assessed through traditional objective measures or exams, as illustrated in Table 1, whereas learning that is more complex often requires qualitative assessments. For example, we linked the learning outcome of explaining hypothesis testing to a particular outcome measure—to the exam rather than to a specific question on an exam. Linking outcomes to specific items becomes too cumbersome and potentially unreliable. Therefore, linking the learning outcome (i.e., explaining the logic of hypothesis testing) to several questions is a more manageable approach. In fact, using several homogeneous items to measure a single outcome (Nunnally, 1978; Pedhazur & Schmelkin, 1991) increases reliability. Table 1 includes each of the student learning outcomes and linkages to existing measures (i.e., exams) of student performance. As such, the table is a planning document for the course.

The embedded assessment model is useful because we plan the course with attention to student learning outcomes, while at the same time we plan for assessing progress toward these goals. For example, average performance on an exam might serve as an indication of student performance while simultaneously indicating progress toward a student learning outcome. In other words, we might specify teaching effectiveness as a score (i.e., criterion) on an assessment measure. Although some (Tomcho & Foels, 2008) suggest that a pre- and posttest measure is the only valid measure of student learning, applied research is subject to a host of confounding factors (e.g., students are enrolled in multiple courses). LoSchiavo et al. (2008) suggest that using scientifically valid experimental designs will become increasingly important. They suggest that evaluating pedagogical innovations by using two groups (i.e., treatment and control) may be the best approach. A criterion-based measure of student learning ensures that students leave the course with the requisite skills. Although not every student will meet the minimum specified criterion (i.e., 70% correct) for every aspect of a course, we can use these data to examine our teaching practice and to make adjustments to improve our teaching when students are not meeting the minimum criteria.

One potential criticism of this approach is a lack of standardization. Classroom evaluations are not typically standardized, so using exams and assignments as a measure of student learning may be challenged as a method for assessing student learning. Nevertheless, faculty already rely on classroom assessments as a means for assigning grades that are meant to reflect the level of student learning. If grades reflect acquired knowledge, then it follows that current assessment techniques possess some level of reliability and validity. It is possible to derive evidence of reliability using available test-scoring software (i.e., coefficient alpha). Coefficient alpha is an estimate of internal consistency reliability that reflects homogeneity of items (Pedhazur & Schmelkin, 1991). By using existing testing software, it is relatively easy to examine internal consistency reliability for the set of items linked to the learning outcome we are measuring.

Content validity is present to the extent that there is agreement among faculty that the exam questions reflect the content of the course. When designing embedded measures, student learning outcomes should be carefully linked to items in the assessment (Eckert & Dabrowski, 2010). Additional efforts to ensure breadth and depth of items should also be undertaken. External evaluation of the items by colleagues may provide a reliability check for content validity. Ideally, all faculty teaching multiple sections of a given course might come to consensus on assessment measures. Realistically, only a subset of common items is likely to be used in the assessment of student learning. This extra step, seeking agreement on items that measure student learning, may help alleviate the self-serving problem inherent in faculty reporting their own students’ progress toward outcomes.

Using existing exams as a measure of content knowledge is a useful way to assess student progress toward the intended student learning outcomes; however, learning outcomes that specify skills (i.e., written products) can only be assessed using qualitative measures. For example, if we want students to demonstrate knowledge of APA format, then an objective measure may suffice. If, however, we are interested in how well a student can apply APA writing style (e.g., economy of expression, precision, clarity) in their writing, then we need to evaluate student learning outcomes using a more qualitative approach, or a written product. Establishing reliability and validity for qualitative evaluations is more challenging. Faculty are already overwhelmed with grading student papers, so finding a second instructor to assist in establishing interrater reliability is daunting. It is difficult to achieve consistency for our own grading (i.e., intrarater reliability) and even more challenging to establish reliability between graders (i.e., interrater reliability). However, although Stellmack, Konheim-Kalkstein, Manor, Massey, & Schmitz (2009) caution that even with a rubric, it is difficult to obtain high levels of reliability, they also concede that clear criteria help to improve the overall level of reliability for writing assignments. Thaler, Kazemi, and Huscher (2009) found that when they used carefully crafted rubrics, they were able to obtain high levels of interrater reliability. We suggest that faculty specify a minimum criterion for demonstrating acceptable levels of writing when using the embedded assessment model. At the conclusion of the semester, the percentage of students meeting the minimum performance criteria, or achieving the learning outcome, can be derived as evidence of student learning and, ultimately, evidence of teaching effectiveness.

**Discussion**

We propose that this embedded model of assessment is effective for two purposes. First, we can evaluate student progress relative to specified student learning outcomes. Second, although grades, as a measure of student learning, reflect
student and instructor factors, (Tomcho & Foels, 2008), we can use student performance data to demonstrate teaching efficacy and to engage in self-evaluation of teaching effectiveness. These two purposes are linked if, through this method of analysis, an instructor discovers a low level of student performance on a measure of student learning and, therefore, adjusts his or her teaching to produce a better product. This intentional and iterative process of identifying student learning outcomes, linking the outcomes to course assessments, and examining the overall levels of student learning can inform teaching.

Why use this model? It is clear that pressures to conduct sound assessments of student learning will continue (Middaugh, 2001). First, a more deliberate use of existing measures of student success can provide incremental evidence of student learning and move us toward meeting the call for accountability. Second, not only can we use embedded assessment to provide evidence of student learning, but used appropriately, evidence of student learning is also a reflection of teaching effectiveness. Third, using embedded assessment allows faculty to take an active and intentional role in specifying student learning and determining whether students meet the specified criteria.

Despite the promise of this embedded assessment approach to inform teaching and learning, we want to acknowledge the inherent limitations of this model. Classroom tests and grades on papers are not as reliable as nationally standardized measures. Therefore, measurement of student learning is necessarily constrained by the reliability of the instruments (e.g., exams, papers) derived at the campus level. Nevertheless, we can easily use scoring software to obtain an estimate of reliability for classroom exams and we can enlist our colleagues to assist in providing minimal evidence of interrater reliability for qualitative measures. Content validity is present to the extent that the assessment matches the student learning outcomes.

Embedded assessment offers one tool to measure student learning, and faculty can use embedded assessments to provide evidence of student learning. A by-product of this embedded assessment technique is that it provides clear directions to students about the objectives that are important for the course and it offers faculty a mechanism for evaluating progress toward those objectives. For example, careful attention to specifying student learning outcomes for the course helps clarify student proficiencies that should be evident at the end of the course. If an instructor continually links student learning outcomes to assessments throughout the semester, then outcomes of the course are clearly communicated. Thus, a student can use the specific outcomes and performance on assignments to articulate competencies. In other words, students might feel more invested if they have a clear understanding of how they can articulate their competencies at the conclusion of the course. In this way, embedded assessment can also build a sense of engagement in the classroom that allows all students to feel a greater sense of accomplishment. As a result, faculty can use embedded assessments to provide evidence of student learning and innovative teaching methods, thereby enhancing their own efficacy as teachers.

Declaration of Conflicting Interests
The authors declared no potential conflicts of interests with respect to the authorship and/or publication of this article.

Financial Disclosure/Funding
The authors received no financial support for the research and/or authorship of this article.

References
Additional Resources for Using Embedded Assessments


Bios

Maureen A. McCarthy is the former Associate Executive Director of Precollege and Undergraduate programs for the American Psychological Association. She currently holds the position of Professor of Psychology at Kennesaw State University.

Daniel M. Niederjohn holds a doctorate in Clinical Psychology from the University of Tennessee. He is currently an Associate Professor at Kennesaw State University.

Theodore N. Bosack holds the doctorate from Brown University in child psychology and is Emeritus Professor of Psychology at Providence College. Currently, he is Executive Director of the Society for the Teaching of Psychology (Division 2 of APA).