A METHODOLOGY FOR EMBEDDED ASSESSMENT
Integrating Grading and Assessment

Name:	Craig A. Wendorf, Ph.D., Professor of Psychology
Location:	D240 Science Center, University of Wisconsin – Stevens Point
Contact:	cwendorf@uwsp.edu, 715-346-2304 (with voicemail)

Overview of Embedded Assessment

The General Purpose

This document describes an approach for conceptualizing embedded assessments of student performance. Specifically, it describes how both grading and assessment can be implemented in a course using a well-defined rubric for clearly articulated learning outcomes. Though it is described in conjunction with the common grading rubric used for all of my courses, the methodology applies to all rubric-based grading and assessment.

Assessment and Grading

In general, both assessment and grading have one very important thing in common: assessing levels of performance. However, the difference lies in the way the data are examined and what aspects of the performance are reported. In grading, the focus is on averaging the performance across outcomes for individuals; thus, the reported numbers conflate performance on learning outcomes. In assessment, the focus is on averaging the performance across individuals for the outcomes; thus, the reported numbers apply to outcomes and say little about students' individual performance relative to others.

Assessment can be and often is conducted separate from grading. For example, one form of assessment asks students to complete a task outside of class (cf., at the end of years of schooling in an exam of graduating seniors) and these results are then used as indicators of the program being reviewed. However, assessment can be embedded directly into the class and even done in concert with grading, but only if the learning experiences and graded work are representative of specific learning outcomes.

Assessing Levels of Performance

One way to integrate assessment and grading is the use of a well-defined, common rubric. In the rubric I use for my classes for example (Wendorf, 2007), the level of performance on any task can be rated as being one of the following:

- Unacceptable – Performance does not meet the quality expectations for the task
- Problematic – Performance meets low quality expectations for the task
- Satisfactory – Performance meets normal quality expectations for the task
- Good – Performance meets high quality expectations for the task

Of course, these qualitative judgments by themselves are broad and ill-defined. It is important for an instructor to be able to specify the criteria that constitute these levels of performance, especially if multiple raters of

performance would be used. Nonetheless, even a simple and broad rubric such as this can assist in creating a matrix that allows one to assess learning outcomes and grade student work in tandem.

Integrating Assessment and Grading

A Brief Example

To be more concrete, suppose that three students have completed a paper that is designed to assess three basic outcomes: 1) their knowledge of a particular concept, 2) their ability to apply the concept to a real-life example, and 3) their ability to provide empirical evidence for the concept. The table below reflects each student's performance on the three learning outcomes.

Student	Outcome 1	Outcome 2	Outcome 3	Grades
Student A	❑ Unacceptable ❑ Problematic ❑ Satisfactory ❑ **Good (3)**	❑ Unacceptable ❑ Problematic ❑ **Satisfactory (2)** ❑ Good	❑ Unacceptable ❑ Problematic ❑ Satisfactory ❑ **Good (3)**	Sum = 8 Percent = 88.9
Student B	❑ Unacceptable ❑ **Problematic (1)** ❑ Satisfactory ❑ Good	❑ Unacceptable ❑ Problematic ❑ **Satisfactory (2)** ❑ Good	❑ **Unacceptable (0)** ❑ Problematic ❑ Satisfactory ❑ Good	Sum = 3 Percent = 33.3
Student C	❑ Unacceptable ❑ Problematic ❑ **Satisfactory (2)** ❑ Good	❑ Unacceptable ❑ **Problematic (1)** ❑ Satisfactory ❑ Good	❑ Unacceptable ❑ Problematic ❑ **Satisfactory (2)** ❑ Good	Sum = 5 Percent = 55.6
	M = 2.00 SD = 1.00	M = 1.67 SD = .58	M = 1.67 SD = 1.52	

For the sake of grading, point values are attached to each judgment of performance (Unacceptable=0, Problematic=1, Satisfactory=2, and Good=3). Grades for each student are then determined as either the sum of the points across outcomes or converted into percentages of the available points. In this example, Student A did quite well (8 out of 9 points, for an 88.9%), whereas the other two students did not perform up to acceptable standards on average.

For assessment, however, the focus is not on the individuals but on the outcomes. Here the same point values are used but averaged across the individuals. In this example, the average point values for Outcome 1 were 2.00, which was higher than the averages for the other two outcomes. Nonetheless, performance was not particularly strong for any of the outcomes, and considerable variability existed in performance for each outcome.

Assessment Reports

Logically, the scoring procedure above is similar to a common statistical procedure called a Repeated Measures Design Analysis of Variance (RMD ANOVA). In an ANOVA, the focus is on typically comparing scores across multiple treatments or conditions – which here are the learning outcomes that were assessed – and not on comparing the average scores of the individuals (which are regarded in the ANOVA merely as individual differences). Thus, this common form of research methodology and statistical analysis is well-suited for directing attention toward assessment of learning outcomes and away from grading.