The Innovative Instructor

Center for Educational Resources

# Best Practice Forum

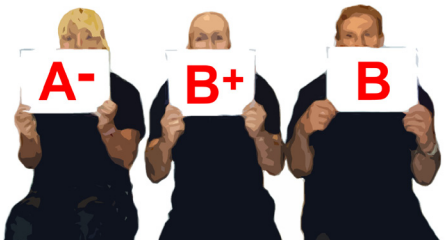
March 2011

## Calibrating Multiple Graders

*Pamela R. Bennett, Assistant Professor of Sociology*
*Andrew J. Cherlin, Professor & Chair of Sociology*
*Michael J. Reese, 5th Year Graduate Student in Sociology*

## What it is

Assessing student work in large classes can be complicated when several faculty or multiple teaching assistants share the responsibility. In a calibration exercise, multiple individuals work together to score a sample of student submissions before dividing and individually grading the remaining student work.




## Why it matters

Calibrating multiple graders helps to standardize assessment of student submissions. The practice minimizes variation in interpretations of students' work across graders. It also allows instructors and teaching assistants to field questions regarding scores from any student, not simply those whose work they personally graded.

## How to do it

The process of calibrating graders involves several steps that can be adjusted based on the course content, structure of the assignment, and individuals involved in grading. Below is the process used by the faculty and teaching assistants in an Introduction to Sociology course.

### I. Create a Rubric

"A *rubric* is an explicit set of criteria for assessing a particular type of work or performance"[1] A rubric provides a common definition of what is right or wrong and standards by which to rate the quality of students' work. It clearly describes what students must include to receive points and/or why points should be deducted. Some rubrics even include example answers.

#### Sociology Exam Rubric

The following exam question was worth ten points. The answer key defines a basic answer that is to be awarded a maximum of 7 points, and provides examples of elaboration that are worth a maximum of 3 additional points.

1. *Give two examples of important changes in the American family over the past half-century.*

| | Basic Answer | Elaboration |
|---|---|---|
| Answer Key | There are serveral "important" changes, any two of which would constitute a basic answer of **7 points**: rise in cohabitation, rise in divorce, the increase in single-parent families, postponement of marriage (and rise of early adulthood as a new life stage); movement of married women into the work force. Other responses could be acceptable – when in doubt, check with the instructor. | What will make the difference for the last **3 points** is whether students write something about each of their two examples that suggests that they understand the significance of the change, such as effects on children, changing roles of husbands and wives, increasingly long period of time before one becomes an "adult," etc. Many answers are possible, but they should go beyond generalities such as "caused great change." |

[1] http://www.tltgroup.org/resources/flashlight/rubrics.htm

## What this is

*The Innovative Instructor* is a forum that publishes articles related to teaching excellence at Johns Hopkins

## About the CER

The Center for Educational Resources partners with faculty and graduate students to extend instructional impact by connecting innovative teaching strategies and instructional technologies

For information on how to contribute to *The Innovative Instructor* or to access archived articles,

please visit our website
- www.cer.jhu.edu/ii

or call Cheryl Wagner
- (410) 516-7181

## Forum categories

***Pedagogy Forum***
Hopkins professors share successful strategies for teaching excellence

***Technology Forum***
Information about emerging technologies, who is using them, and why you should know

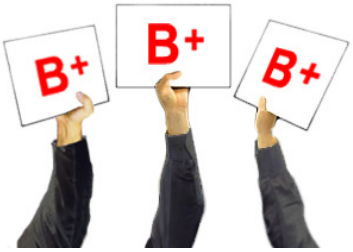
***Best Practice Forum***
"How To" workshops on using technologies and applying innovative instructional methods

THE SHERIDAN LIBRARIES

JOHNS HOPKINS UNIVERSITY

## II. Preparing for the Calibration Session

1. Identify several student submissions to be scored by the group. To quickly identify a sample of submissions that will likely generate marks across the range of scores as defined by the rubric, target submissions based on previous grades for homework or in-class participation.
2. Make copies of the submissions for each grader without including student identifiers to minimize grader bias.




## III. Calibration Session

During calibration, the group scores the sample student work either together or apart. If apart, the group reassembles and each person communicates how s/he interpreted the rubric and assigned a score. The group discusses differences in assigned scores and comes to consensus on a final score, using a standard interpretation of the rubric or a modification of it. (It is not unusual to modify the rubric; identification of weaknesses in the rubric is another benefit of the exercise.) Below is an example of how a session could be conducted.

1. The instructor who created the rubric reviews it with the group and explains the scoring criteria.
2. Graders individually read submissions and score the work based on the rubric.
3. Each person shares the score s/he assigned to the first component of the work (e.g., first question on an exam). If there is variation in assigned scores, then each grader explains how s/he arrived at the score assigned. The group then reaches consensus on a score based on a common interpretation of the rubric (or its modification). The group moves to the next component of the work. Graders should make notes on the rubric during this discussion to consult when scoring student work on their own.
4. Once calibration is accomplished, the remaining student submissions are divided among the group. Option: Assign student submissions to graders unfamiliar with the students to reduce bias based on previous interactions. This may be useful for large classes with small group sections assigned to TAs.

## IV. Staying Calibrated

As graders review students' work, they should note submissions for which they found assigning a score difficult. This may result from uniqueness of a student's response or difficulties in interpreting the rubric. Graders should communicate with the group about how to address the issue before final grades are submitted, as the resolution may affect how other graders score.

To check grader calibration, the lead grader can spot check scores assigned by different graders. Another option is to compare the mean and standard deviations of the scores assigned by each grader. Some variation will naturally occur, but if extreme outliers are identified, those graders' scores can be reviewed.

# Additional Resources

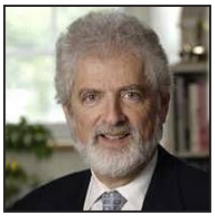
- Example rubric from California State University at Long Beach, Computer Engineering department: http://www.csulb.edu/colleges/coe/cecs/views/programs/undergrad/grade_prog.shtml

# Authors' Backgrounds


### Dr. Pamela R. Bennett
Assistant Professor of Sociology

Pamela Bennett is an assistant professor of Sociology. She teaches several courses related to her research on social inequality rooted in race, class, and residential location, in addition to co-teaching the Introduction to Sociology course. This course employs several graduate students who share the grading responsibilities with two faculty.

### Dr. Andy Cherlin
Professor & Chair of Sociology



Andy Cherlin is professor & chair of Sociology and the Benjamin H. Griswold III Professor of Public Policy. He teaches courses about the family and statistical methods in addition to co-teaching the Introduction to Sociology course.

### Mike J. Reese
5th Year Graduate Student in Sociology

Mike Reese is a 5th year graduate student in Sociology assigned as a TA in the Introduction to Sociology Course in Fall 2009. His research explores how teaching innovations spread through higher education. He is also the assistant director of the Center for Educational Resources.