Consensus-Based Assessment

Passport to Great Teaching – Creative Assessment

Timothy S. Brophy, Professor and Director of Institutional Assessment
Goals of Module 6

- Establish the emerging conceptions of individualized assessment
- Describe Consensual Assessment and Consensus Moderation
- Explain and operationalize Consensus Moderation as an assessment process
The assessment quandary posed by creative and open-ended responses
Data collected as evidence of learning is delimited by assessment type.

Consensus – based methods present a non-delimited approach to assessment.
The assessment of creative or open-ended works is difficult to standardize.

There is no ‘standardized student’ nor is there a ‘standardized response’ to an open-ended or creative assessment task.

Even our best attempts at developing standardized rubrics or scoring approaches are limited in their scope, transferability to practice, and still raise questions of validity.

If variability is systematically ignored, individuals become synonymous with statistical averages, and faculty and researchers lose the ability to account for the very processes that underpin the phenomena they seek to explain.
Rose et al. (2013) present an approach to understanding and analyzing human behavior based on the precept that individuals behave, learn, and develop in distinctive ways, showing patterns of variability that are not captured by models based on statistical averages.

The authors ask us to consider human beings as *dynamic systems*, which assumes that behavior is actively organized and context-dependent, variability is expected as a natural outcome.

They argue that learning is not a linear progression through a universal sequence, where the start and endpoint are predetermined.

Creative and open-ended tasks and their responses often fall into this category.
Reconsidering our approach to assessment for individualized responses

Based on our understanding of the individual variability of creative human responses, we should reconsider our approach to assessment data collection and analysis.

Standard assessment methodology: aggregate, then analyze.

A reconsidered methodology: Analyze, then aggregate.
Why consensus-based assessments?
Most assessments are designed to be measured by a predetermined set of criteria using a tool such as a rubric or checklist, or by counting points earned by responding to questions (such as those on an exam or quiz) and placing the total score on a grading scale to determine achievement.

However, there are a number of assessment types for which the development of preset criteria is counterproductive because the criteria actually constrain the response by forcing respondents to conform to the preset criteria.
Why Consensus-based Assessments?

- These assessments include those that examine individual distinctiveness or creativity and therefore result in considerable expected response variability – including, but not limited to: interpretation, creative writing (stories, novels, poems, etc.), artistic creations (musical compositions, improvisations, choreography, paintings, sculptures, ceramics, and other fine arts), development of new theories, logical arguments, etc.

- Setting criteria in advance for these types of assessments could lead to response conformity that conflicts with their intended purpose to express the individuality of the respondent.
Consensual Assessment

Consensus – based assessments
The Concept of *Consensual Assessment* to measure creativity

- The concept of consensual assessment (Amabile, 1996) rests on the belief that validity evidence for assessments of creativity is strongest when experts rate the creative product using their subjective judgments.

- In the consensual assessment process, raters use predetermined criteria (dimensions) as a focal point for measurement, and determine levels of achievement using some form of scale that leaves determination of levels of criterion demonstration to the raters’ subjective judgments.
Consensual Assessment

• Judges measure the creative products of interest in isolation – there is no collaboration or contact among the judges.

• Interrater reliability is critical and has been shown to be in acceptable ranges in some recent studies (Hennessey et al., 2011), but not all (Hickey, 2001).
Consensus Moderation

Rethinking our approach to the measurement of creative learning and open-ended/creative works to accommodate individual variability
A Review: Setting criteria for assessment

The specification of criteria in advance is approached in one of two ways: analytically or holistically.

There are rubric types that correspond to these two approaches:

| Analytic rubrics list criteria and describe levels of achievement | Holistic rubrics specify levels of achievement that address multiple criteria at each level |
Here are some reasons why we do this:

- Students have a right to know how the quality of their work will be judged before they begin constructing responses.
- All student responses to the same task should be assessed according to the same criteria (fairness).
- Criteria provide guidance to students – they can attend to the criteria as they construct their response.
- Fixed criteria add objectivity to the judgment (reduces or eliminates subjectivity).
- Sets of criteria provide a convenient and economical way to provide feedback to students.

Sadler, D. R. (2009, Spring)
Reconsidering the use of preset criteria

When you set criteria in advance for assessment, what happens?

While criteria appear at the outset to be a good idea, in practice they are much more difficult to separate.

Much of our actual feedback to students lies outside of the criteria!

So, the criteria that were specified in advance have really not helped much in the assessment process.
**Consensus**

Consensus: reaching a general or common understanding

**Moderation**

Moderation: by definition, lessening of extremes.

Moderation can be done in different ways: (1) **Averaging** different readings or coded judgments; (2) **removing** the most discordant judgment or divergent judgments (and averaging the remainder); (3) Accepting the **middle** reading; (4) **Discussing until consensus** is reached.

Consensus moderation is a **result** of successful consensus-seeking which reduces discord and by that means, moderates.

Sadler, D. R. (2015, August 27)
Consensus Moderation as an assessment process

Multiple ‘experts’ engaged in the review of an artifact or work

Assessors must be open to the qualities observed in the work; there is no attempt to steer the student toward any particular qualities

Assessors make a holistic judgement about the level of proficiency or competence
Consensus
Moderation
as an assessment process

Reasons follow the judgement – this invokes criteria – but criteria that are salient to the judgement

- Justification is not the same as rationale
- **Justification** sets out the grounds (reasons) for the judgment
- **Rationalization** is a defense of the assessor’s judgment

Students need to be inducted into this process – so they can learn to monitor and control the quality of their own performances and productions while their productions are in progress.
Reliability and Validity

• Consensus-moderation provides strong validity evidence in that there is discussion among the assessors to arrive at consensus on the assessment decision

• Interrater reliability is high when all assessors reach consensus
Pause to think: Consensus-based approaches

• Where in your assessment of student work in your course or in your program is consensual assessment or consensus moderation the best approach?
• How might you triangulate your other assessment data from exams, quizzes, etc. with the results of a consensus-based assessment?


• Rose, L. T.; Rouhani, P.; Fischer, K.W (2013). The science of the individual. *Mind, Brain, and Education 7*(1), 152-158
