



# Integrating Measurement Principles into Formative Assessment

*Randy Bennett  
Educational Testing Service  
Princeton, NJ 08541  
rbennett@ets.org*

Presented at the annual meeting of the International Association for Educational Assessment, Baku, Azerbaijan, September 2019

---

# HANDBOOK OF FORMATIVE ASSESSMENT IN THE DISCIPLINES

---

Edited by  
Heidi L. Andrade  
Randy E. Bennett  
Gregory J. Cizek



# Overview

- Premise
- Epistemic frame
- Assessment as four fundamental acts
- Conclusion

# Premise

- Fundamental measurement principles apply to formative assessment
  - *Doesn't* mean formative assessment should be standardized, quantified, or held to standards of technical quality required for high-stakes summative tests
  - *Does* mean there are basic precepts which, if ignored, can render formative judgments of limited value for moving learning forward

# Epistemic Frame

- Use frame of evidentiary reasoning
  - Mitlevy et al. (2003)'s ECD
  - Offers a powerful perspective for making meaning and taking action from our observations of student behavior
- The theory and methodology of ECD provide:
  - A way of reasoning about assessment design
  - A way of reasoning about observable behavior

# Assessment as Evidentiary Reasoning

- Four fundamental acts:
  - Engineering opportunities to observe evidence of the competencies we wish to make claims about, and then making the relevant observations
  - Inferentially connecting that evidence to meaningful characterizations of individuals, groups, or institutions
  - Acting on those characterizations (e.g., making instructional adjustments, making an admissions decision)
  - Evaluating the quality and impact of the above
- Applies *regardless of assessment purpose*, though the rigor, formality, and methodology will differ significantly

# Engineering Opportunities to Observe Evidence

- *Design* situations, activities, tasks, or questions that generate observable *evidence*
  - Design
    - Working intentionally from claims, to the types of evidence required, to the opportunities that will provide that evidence
    - Intentionality should be the usual practice
      - Unanticipated situations occur
  - Evidence
    - Presumes some disciplinary framework—content standards, domain theory—without which it's hard to know what might be an indication of competency

# Inferentially Connecting Evidence to Characterizations

- Use that evidence to make judgments, or characterizations, about what the student knows and can do so that instructional next-steps can be taken
- Our characterizations of competency always have some degree of uncertainty
  - Students sometimes misunderstand questions or are distracted such that their response implies a deficiency when there is none
  - Students sometimes answer correctly for the wrong reasons, suggesting they understand something they really do not
- Uncertainty is noise that clouds the underlying signal, possibly leading to inappropriate adjustments



# Reducing Uncertainty

- Integrate the observed evidence with what is known from past behavior
  - Does the response agree with what was observed in homework, quizzes, tests, or other classwork on the same topic?
  - How does the response square with what is already believed about the student's prerequisite, as well as more advanced, knowledge?
- Gather more evidence using, as appropriate:
  - Different task formats
  - Other topical contexts
  - Other framing or phrasing

# Systematic Bias

- Biases may be due to unfamiliarity with the culture, language, or behavior of students from particular demographic groups
- Research suggests teachers' judgments of students' academic competencies may be influenced by:
  - Race/ethnicity
  - Social class
  - Gender
  - Language
  - Disability status
- Bias may contribute to consistent under- or over-estimates of what students know and can do

# Systematic Bias

- Problematic if it leads to misconstruing a poor response as lack of competency when the cause of incorrect responding is an irrelevant factor
- Repeated misconstrual may:
  - Lead teacher to lower expectations for some students
  - Suggest to those students that they are not progressing as well as they really are
  - Cause instructional time to be spent in less productive ways

11 BB 15



A B C



# Reducing Bias

- Teachers can take several steps to reduce the possibility of bias
  - Recognize that most individuals have preconceptions about other groups and that these preconceptions can affect how the behavior of group members is perceived
  - Develop knowledge of the student groups present in one's classes, especially if they are different from one's own demographic group
  - Routinely consult other information sources that might not be subject to the same biases
    - Colleagues with suitable background and experience with unfamiliar groups can react to:
      - Samples of student work
      - Descriptions of student classroom behavior
    - Colleagues' reactions:
      - May suggest need for further data collection
      - Can be integrated into teacher judgments of student competency

# Acting on Characterizations

- Action often takes the form of next instructional steps
- A next instructional step also can reduce uncertainty as the student's response may strengthen or weaken a characterization
- Characterization as a *formative hypothesis* that can guide an iterative cycle:
  - Observe behavior
  - Generates hypotheses about what the student knows and the causes of incorrect responding
  - Take further action (including the making of instructional adjustments and gathering of new observations)
  - Update the starting hypothesis

# Evaluating Quality and Impact

- The quality of formative assessment is a function of the:
  - Situations teachers (or students) engineer
  - Characterizations of competency they make about behavior in those types of situations
  - Actions they take based on the characterizations
  - Impact of their actions on learning



# Formative Assessment and Learning

- Formative assessment as a chained activity, from which learning is more likely to occur if:
  - Its constituent situations reveal suitable evidence
  - That evidence is used to meaningfully characterize what students know and can do
  - Instructional next-steps are taken that sensibly follow from those characterizations
- To the extent that one or more of these links is inadequate, it becomes more difficult to logically ascribe any observed learning to formative assessment

# Improving Quality and Impact

- Self-reflection about:
  - Which situations produced useful evidence (as well as how to change those situations that didn't work)
  - How one's characterizations of proficiency compare to characterizations from other sources
  - How sensibly one's instructional decisions follow from those characterizations (and from cognitive-domain theory, content standards, or curriculum objectives)
  - Whether learning occurred and if that learning could be reasonably connected to the chain of formative actions

# Conclusion

- Principles from educational measurement and the practice of formative assessment can be brought together through evidentiary reasoning
- Evidentiary reasoning is a way of thinking (or habit of mind) for teachers (and students) to use in:
  - Designing situations that allow for observing evidence of target competencies
  - Reasoning backward from that evidence to characterizations of proficiency
  - Taking action in keeping with those characterizations
  - Reflecting upon quality and impact (and improving formative practice)

# Conclusion

- Within this reasoning framework, behavior provides evidence for characterizing what students know and can do
  - With such characterizations come:
    - Uncertainty
    - Systematic biases
  - Uncertainty and bias can reduce the appropriateness and effectiveness of next instructional steps

# Conclusion

- To minimize uncertainty, one can:
  - Integrate observations with knowledge of past student behavior and evidence from other sources
  - Gather more evidence using a variety of task formats and topical contexts
  - Use an iterative cycle of observing behavior, formulating hypotheses, taking action, and updating hypotheses

# Conclusion

- To minimize bias, one can:
  - Recognize that most individuals have preconceptions that can affect how behavior is perceived
  - Develop knowledge of the (unfamiliar) student groups in one's classes
  - Routinely consult other information sources that might not be subject to the same preconceptions

# Conclusion

- Continual self-reflection on the quality and impact of one's formative practice would seem important
- Ideally, teachers should cultivate in themselves the same types of self-reflective and self-regulatory behavior students are expected to develop

# Citations on Bias

- Bennett, R. E., Gottesman, R. L., Rock, D. A., & Cerullo, F. M. (1993). The influence of behavior and gender on teachers' judgments of students' academic skill. *Journal of Educational Psychology*, 85, 347-356.
- Hurwitz, J. T., Elliott, S. N., & Braden, J. P. (2007). The influence of test familiarity and student disability status upon teachers' judgments of students' test performance. *School Psychology Quarterly*, 22, 115-144. doi:10.1037/1045-3830.22.2.115
- Meissel, K., Meyer, F., Yao, E. S., & Rubie-Davies, C. M. (2017). Subjectivity of teacher judgments: Exploring student characteristics that influence teacher judgments of student ability. *Teaching and Teacher Education*, 65, 48-60. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0742051X17303475#bib64>
- Ready, D. D., & Wright, D. L. (2011). Accuracy and inaccuracy in teachers' perceptions of young children's cognitive abilities: The role of child background and classroom context. *American Educational Research Journal*, 48, 335-360. Doi: [10.3102/0002831210374874](http://dx.doi.org/10.3102/0002831210374874)



# Citations on Measurement Principles in Formative Assessment

- Bennett, R. E. (2019). Integrating measurement principles into formative assessment. In H. L. Andrade, R. E. Bennett, & G. J. Cizek (Eds.), *Handbook of formative assessment in the disciplines* (pp. 20-31). New York: Routledge.
- Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy and Practice* 18, 5-25.  
<https://www.tandfonline.com/doi/full/10.1080/0969594X.2010.513678>
- **rbennett@ets.org**

---

# HANDBOOK OF FORMATIVE ASSESSMENT IN THE DISCIPLINES

---

Edited by  
Heidi L. Andrade  
Randy E. Bennett  
Gregory J. Cizek

