

# Interim Assessments

Marianne Perie

Brian Gong

Scott Marion

Presentation for the Reidy Interactive Lecture Series (RILS)

Nashua, NH

October 5, 2006



# New Assessments

- States want more information on student performance at a time when they still have a chance to intervene.
- Led to
  - A call for formative assessment from within states
  - An influx of new programs by vendors



# Our Goal

- Define formative assessment and uses
  - Is it the panacea?
  - Is that what the vendors are selling?
- Broaden our perspective to include interim assessments
  - What other uses can these assessments have?
  - What makes sense?
- Develop a framework for evaluating an interim assessment system.



# Formative Assessment

- Black and Wiliam (1998) wrote the seminal piece describing how formative assessment can improve student learning
- Newly suggested definition:
  - *An assessment is formative to the extent that information from the assessment is used, during the instructional segment in which the assessment occurred, to adjust instruction with the intent of better meeting the needs of the students assessed.*
- Key components
  - Integrate seamlessly with curriculum
  - Provide corrective feedback
  - Include diagnostic and prescriptive information (i.e., what should teachers do next)



# Assessment with Formative Uses

- Some assessments may help form instruction without meeting all the criteria for formative assessment
- Meet some requirements by
  - Providing *qualitative* insights about understandings and misconceptions not just a numeric score
  - Giving timely feedback on what to do besides re-teaching every missed item



# Interim Assessment

- Definition
  - A shorter assessment often given multiple times during the school year
- Focus on uses
  - Instructional (could be formative)
  - Predictive
  - Program evaluation



# Consider these uses...

- Predict student achievement on summative test (e.g., early warning)
- Diagnose gaps between student knowledge and intended curriculum
- Provide information on how best to target curriculum to meet student needs
- Provide aggregate information on how students in a school/district/state are doing and where areas of weakness are
- Reinforce curricular pacing
- Evaluate the effectiveness of various curricular and/or instructional practices
- Determine students' ability levels to group them for instruction
- Enrich curriculum
- Encourage students to evaluate their own knowledge and discover the areas in which they need to learn more.
- Practice for summative test
- Increase teacher knowledge of assessment, content domain, and student learning



# Varied Uses and Purposes

- All of these purposes may be worthwhile even if they are not formative
- However, we need to establish clear links between what questions the policymakers and educational leaders want to answer and the tools they are using to do so



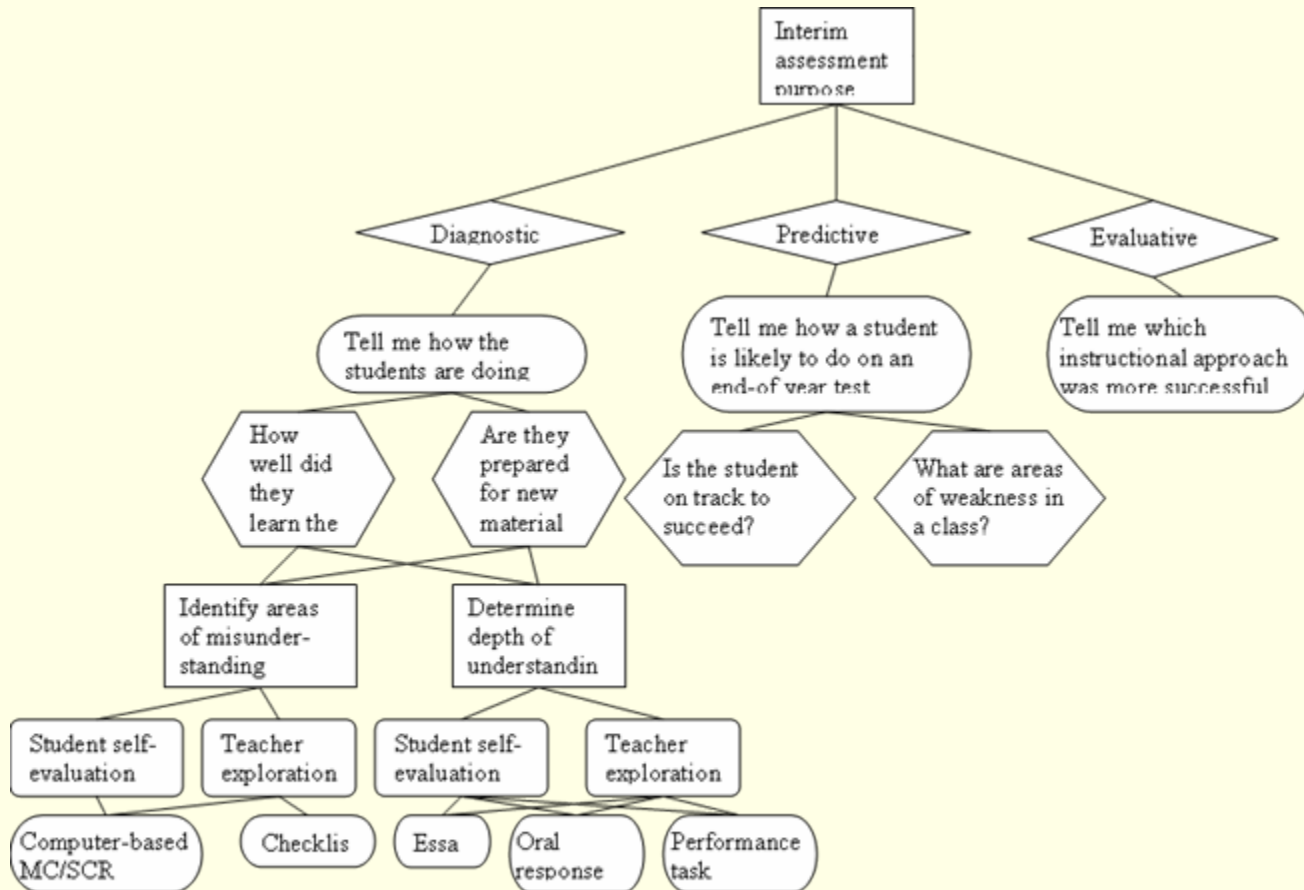


# Consider these questions...

- What do I want to learn from this assessment?
- Who will use the information gathered from this assessment?
- What action steps will be taken as a result of this assessment?
- What professional development or support structures should be in place to ensure the action steps are taken?



# Decision Tree



# Characteristics of a Good Interim Assessment System

- Provides valid and reliable results that are easy to interpret and provide information on next steps
- Includes a rich representation of content with items linked directly to the content standards and specific teaching units.
- Fits within the curriculum so that the test is an extension of the learning rather than a time-out from learning
- Three main elements
  - Reporting Elements
  - Assessment Design
  - Administration Guidelines



# Reporting

- Policymakers should consider carefully the reporting component
  - Thinking about the end result helps to conceptualize the design
  - Reporting translates data into action
- Consider all pieces of reporting
  - Qualitative as well as quantitative information



# Reporting Elements

- Type of data summary
  - Include normative reference
  - Compare against criterion reference
  - Aggregate across classroom/school/district
- Type of qualitative feedback
  - Information on correct/incorrect responses by content area
  - Feedback on what an incorrect answer implies
  - Suggestions for next steps



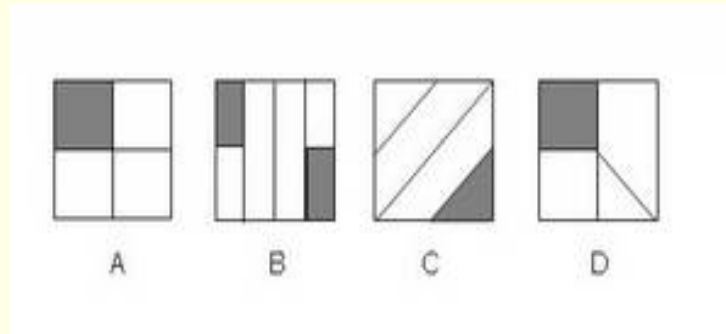
# Assessment Design

- Match item type to purposes
  - Predictive: Match type to what you are predicting, perhaps with additional probes
  - Instructional: More open-ended, probes, performance tasks
  - Evaluative: Can use a combination of multiple-choice and short-answer items with “why” probes
- Consider number and length of items



# Sample MC Item

*Consider the four diagrams shown below. In which of the following diagrams, is one quarter of the area shaded?*



*Diagram A is the obvious answer, but B is also correct. However, some students do not believe that one quarter of B is shaded because of a belief that the shaded parts have to be contiguous. Students who believe that one quarter of C is shaded have not understood that one region shaded out of four is not necessarily a quarter. Diagram D is perhaps the most interesting here. One quarter of this diagram is shaded, although the pieces are not all equal; students who rely too literally on the “equal areas” definition of fractions will say that D is not a correct response.*



# Sample Performance Task

## Can It Reflect Light?

What types of objects or materials can reflect light? Put an X next to the things you think can reflect light.

- |   |  |
|---|--|
| <input type="checkbox"/> water          | <input type="checkbox"/> red apple                     |
| <input type="checkbox"/> gray rock      | <input type="checkbox"/> rough cardboard               |
| <input type="checkbox"/> leaf           | <input type="checkbox"/> the Moon                      |
| <input type="checkbox"/> mirror         | <input type="checkbox"/> rusty nail                    |
| <input type="checkbox"/> glass          | <input type="checkbox"/> clouds                        |
| <input type="checkbox"/> sand           | <input type="checkbox"/> soil                          |
| <input type="checkbox"/> potato skin    | <input type="checkbox"/> wood                          |
| <input type="checkbox"/> wax paper      | <input type="checkbox"/> milk                          |
| <input type="checkbox"/> tomato soup    | <input type="checkbox"/> bedsheet                      |
| <input type="checkbox"/> crumpled paper | <input type="checkbox"/> brand new penny               |
| <input type="checkbox"/> shiny metal    | <input type="checkbox"/> old tarnished penny           |
| <input type="checkbox"/> dull metal     | <input type="checkbox"/> smooth sheet of aluminum foil |

Explain your thinking. Describe the “rule” or the reasoning you used to decide if something can reflect light.

---

---

---

---





# Administration Guidelines

- Flexibility in creating forms
- Administered within instruction or separate from instruction
- Flexibility in when/where the assessment is given
  - Computer-based
  - Web-based
  - Paper-and-pencil
- Turnaround time for results
- Adaptive or not



# Administration Needs x Purpose

Purpose	Flexibility of test administration	Customization of test form	Adaptive	Speed of results
Predict student achievement on EOY test	Could be given flexibly or at pre-set times each year	Not necessary	Nice feature, but not essential	Days
Provide information on strengths and weakness of a particular group of students	Needs to be given at teacher discretion	Helpful	Nice feature, but not essential	Within one day
Determine students' readiness for next topic	Essential	Helpful	Helpful	Within one day
Enrich the curriculum	Essential	Essential	Not necessary	Within several days
Evaluate the effectiveness of an instructional program	Not necessary	Not necessary	Nice feature, but not essential	Weeks
Provide PD for the teacher	Not necessary	Not necessary	Not necessary	Weeks



# Currently Available Systems

- So how does what we want match what already exists?
- What other options are available?
  - Customized assessment
  - Locally-designed assessment



# Features of Many Current Systems

- What these systems can do:
  - Provide an item bank linked directly to state content standards
  - Assess students on a flexible time schedule wherever a computer and internet connection are available
  - Provide immediate results
  - Highlight content standards in which more items were answered incorrectly
  - Link scores of these assessments to the scores of the end-of-year assessments to predict results on the end-of-year assessments
- Questions these systems can answer:
  - Is this student on track to score Proficient on the end-of-year NCLB tests?
  - Which students are on track to score Proficient on the end-of-year NCLB tests?
  - Which content standards are the students Proficient in and which content standards show the weakest student performance (for a student, classroom, school, district, state)?
  - How does this student's performance compare to the performance of other students in the class?



# What These Systems Lack

- What these systems cannot do:
  - Provide rich detail about the curriculum assessed
  - Provide a qualitative understanding of a student's misconception(s)
  - Provide full information on the student's depth of knowledge on a particular topic
  - Further a student's understand through the type of assessment task
  - Give teachers the information on how to implement an instructional remedy
- Questions these systems cannot answer:
  - Why did a student answer an item incorrectly?
  - What are possible strategies for improving performance in this content area?
  - What did the student learn from this assessment?
  - What depth of knowledge does a student display in this content area?
  - What type of thinking process is this student using to complete this task?



# How do I decide?

- List purposes and desired elements
- Compare to commercially-available assessments
  - Do they contain all the elements you need?
  - Would you be paying for elements you do not need?
- If no current system matches well
  - Determine whether it makes more sense to allow districts to develop their own system
  - Design a customized system to be available throughout the state



# An Example

- Consider a district with the goals of
  - Implementing an early-warning system to identify which students, classrooms, and schools are on track to perform well on the end-of-year assessment
  - Identifying areas of weakness both at the student level and aggregated to the classroom and school level for those not on track
  - Providing additional tools for improving performance on those areas identified as weak
  - Administering this test 3–4 times over the year to track student progress toward the goal.



# Example—continued

- Reporting criteria
  - Report “on-track to succeed”
  - Identify areas of weakness
  - Aggregate across classrooms, school, and the district
  - Disaggregate results by the same reporting categories used in the end-of-year reports (racial/ethnic group, disability status, LEP)
  - Illustrate progress over time
    - How the progress relates to where they should be by the end of the year.
  - Provide information that can be used to determine next steps





# Example—continued

- Assessment design
  - Items map directly to the content standards and be similar in type to the items on the end-of-year test
    - May include further probes to help identify misunderstandings or schema problems
  - Items also link to teaching units and text books specific to that district.
    - Using items that link directly to instructional materials will help provide the connection between any weaknesses found and instructional interventions
  - Each test should only assess what's been taught to date
    - Do not give a series of parallel mini-summative assessments
    - May be some overlap between assessments



# Example—continued

- Administration requirements
  - Results should be available within a week or two to allow time for intervention
  - Either computer-based testing or a pencil-and-paper test would serve this district's purpose
  - Flexibility is not a requirement for this system
    - Standardization in the items administered would be necessary to aggregate results across the district



# How do I know I'm getting my money's worth?

- Validating the evidence will be important to do over the next couple of years
  - If the test is used for predictive purposes, do a follow up study to determine that the predictive link is reasonably accurate and that the use of the test contributes to improving criterion (e.g., end of year) scores
  - If the test is used for instructional purposes, follow up with teachers to determine how the data were used and whether there was evidence of improved student learning for current students
  - If the test is used for evaluative purposes, gather data from other sources to triangulate results of interim assessment and follow up to monitor if evaluation decisions are supported



# Our Recommendations

- Avoid mini-summative assessments
- Focus on the reporting elements to clarify assessment design
- Ensure the instructional supports are in place to allow teachers to use the results effectively
  - This includes substantial professional development on assessment literacy, score interpretation, and necessary instructional actions
- Validate the use of the assessment



# Conclusion

- There are valid purposes for giving interim assessments beyond informing instruction at that point
- Examine the purpose of the assessment and what it can and cannot do
- Match the features of the assessment to the purpose of using it
- Further research is needed linking the use of interim assessments with improved student performance

