Formative and Diagnostic Assessment: Review of Literature and Presentation of the Washington Guides

http://www.uoregon.edu/~stevensj/wera.pdf

Joseph Stevens, Ph.D.
University of Oregon
MESA Associates

A presentation at the WERA Spring Conference,
Seattle, Washington, March 27, 2008

Contact:
jstevens.mesa@comcast.net
4524 Fox Hollow Road, Eugene, OR 97405

WERA Spring Conference Presentation

- Provide overview of the Diagnostic Assessment Project
- Types of assessments and brief literature review
- Overview of the:
  - Washington Diagnostic Assessment Guide
  - Washington Comparative Guide
- Preview of:
  - Resources on the web
  - Summer workshops
The 2007 Washington State Legislature appropriated $4.8 million to school districts to purchase and implement diagnostic assessments:
- For the 2007-08 and 2008-09 school years
- For Grades K-12
- For assessments in reading, writing, mathematics, and science
A second phase of funding will be opened in September 2008 to provide funding to districts for the 2008-09 year.

The Office of Superintendent of Public Instruction (OSPI) will utilize iGrants to collect requests from districts for funding.

Diagnostic assessment — an assessment that helps to:
- improve student learning
- identifies academic weaknesses
- enhances student planning and guidance
- Target instructional strategies to assist students before the HS WASL

(This broad definition of diagnostic assessment includes assessments used for screening and progress monitoring)
- Allows student progress to be compared to other students across the country
- Readily available to parents
- Cost-effective
**Washington State Diagnostic Assessment Project: Background**

Allowable Uses of Funding:
- Purchase of diagnostic assessments
- Costs of administering, scoring, and reporting results
- Training costs

Funds for administering the assessments to students. Funds may not be used for developing diagnostic assessments. However, funds may be used to administer and score diagnostic assessments that were developed prior to the 2007-08 academic year.

If you have any questions regarding this process, please contact Catherine Taylor, Assessment Alternatives and Innovations Director, phone 360/725-6061, TTY 360/664-3631, catherine.taylor@k12.wa.us.

---

**Diagnostic Assessment Project**

As part of the legislation, OSPI is empowered to contract with an outside party to:
- Participate and consult on the development of a diagnostic assessment guide
- Collaborate in the delivery of professional development on diagnostic assessment

MESA Associates has been awarded the contract; participants in the Washington Diagnostic Assessment Project include:
- Joseph Stevens, Ph.D.
- Leanne Ketterlin-Geller, Ph.D.
- Paul Yovanoff, Ph.D.
- Keith Zvoch, Ph.D.
- Rohanna Buchanan, doctoral student
- Maya O’Neil, doctoral student
- Debbie Soul, doctoral student
Our part of the Diagnostic Assessment Project: Products

- Literature review
- Diagnostic Assessment Guide
- Comparative Guide
- Provide material for web site
- Summer Institute workshops

Types of Assessments

- Can be distinguished based on
  - Purpose
  - Test development
  - Timing of assessment
  - Reporting
  - Uses and decisions made
Types of Assessments

- Summative
- Formative
  - Screening
  - Diagnostic
  - Progress Monitoring
- Norm Referenced
- Criterion Referenced
- Standards Based
- Standardized

Types of Assessments: A caveat

- Our part of the Washington Diagnostic Assessment Project focuses only on commercially available instruments
- However, research shows that informal assessments and locally developed classroom assessments can be very effective for formative purposes
  - Q & A
  - Simple prompts and activities
  - Peer and self assessment
  - Concept mapping, etc.
Types of Assessments

- Summative
  - To evaluate, mark attainment of a benchmark, often comparative and normative
  - Test development to compare performances; breadth of content
  - Compares individual performance to other individuals; emphasis on group performance; NCLB
  - Psychometric emphasis: discrimination, classification
  - At end rather than during
  - Reporting emphasizes evaluative criterion (A, B, C’s, “meets” versus “does not meet”)
  - Extrinsic motivation; accountability

- Formative
  - Purpose to guide and motivate learning, to provide feedback especially to student or teacher
  - Test development to support documentation of strengths and weaknesses; depth
  - Compares individual performance to learning targets
  - Psychometric emphasis: tracking change, prediction
  - At beginning and during rather than at end
  - Reporting emphasizes comparison of current performance to learning targets or goals
Types of Assessments

Formative assessments with more specific purposes:

- Screening—to make early identification of needs for classification or intervention; instruments may be characterized by
  - Little depth
  - Less accuracy or detail in information provided
- Diagnostic—specifically to identify student strengths and weaknesses for intervention
- Progress Monitoring—specifically for frequent measurement or observation usually as an adjunct to assess response to intervention

Types of Assessments

- Norm-Referenced Test (NRT)—to compare one individual’s performance to a large group (the norm)
  - Requires representative norm sample
  - Primary interpretation is reference to norm group (GE, PR)
- Criterion-Referenced Test (CRT)– compares individual’s performance to a standard of mastery or competence
- Standards-Based Test (SBT)– In practice may be either a NRT or CRT but test content is aligned to state content standards
Types of Assessments

- Standardized Tests — refers to making test conditions and administration comparable for all
  - Necessary for fairness and interpretability
  - Important if comparing one person’s performance to another’s
  - Note that term “standardized” often confused with NRT

- For formative assessment purposes, different aspects of standardization may be important:
  - Purposefully different from student to student
  - Form equivalence
  - Standard conditions over time to measure growth

Literature Review: Some Seminal Articles

- Scriven (1967) coined terms “summative” and “formative” with respect to program evaluation
- Bloom, et al. (1971) borrowed term “formative evaluation” from Scriven to apply to brief diagnostic tests used to evaluate mastery learning
- Bloom, et al. (1971) developed model consisting of:
  - Diagnosis of learner characteristics
  - Formative evaluation of learning task
  - Feedback and corrections
  - Summative evaluation of attainment
Literature Review: Some Seminal Articles

- Sadler (1983, 1989), three steps in the formative feedback loop:
  - Attending to goals
  - Developing strategies to meet goals
  - Monitoring performance for goal achievement

- Natriello (1987) impact of assessment features on student motivation and achievement:
  - Focus on tasks not comparison of students
  - Clear criteria
  - Challenging standards
  - Differentiated feedback

- Crooks (1988), impact of formative assessment practice on students:
  - Consolidates prior skill and knowledge before new material introduced
  - Focuses attention on most important content
  - Encourages active learning
  - Provides opportunities for practice
  - Provides corrective feedback
  - Allows and develops self-monitoring
  - Guides further instruction
  - Creates feelings of mastery and accomplishment
Literature Review: Some Seminal Articles

- Many reviews on classroom and formative assessment (e.g., Bangert-drowns, et al., 1991; Black & Wiliam, 1998a; Brookhart, 2007; Crooks, 1988; Natriello, 1987; Sadler, 1989; Stiggins, 1999)
- Formative assessment means different things to different people
- Some key findings:
  - Teacher beliefs affect assessment practice
  - Classroom environment affects student perceptions, performance and behavior
  - Effective formative assessment blends assessment and instruction

Literature Review: Some Key Findings

- Effects on student achievement
  - Black and Wiliam (1998b), for studies of the effects of formative assessment on learning, typical effect size was .40 to .70
  - Formative assessment raises performance of lower achieving students more than other students (as well as raising achievement overall)
  - Formative assessment can affect motivation and self-esteem of pupils
Literature Review: Some Key Findings

- The importance of feedback in formative assessment
- Feedback represents:
  ‘…information about the gap between the actual level and the reference level of a system parameter which is used to alter the gap in some way’ (Ramaprasad, 1983, p. 4, emphasis added)
- The information must therefore:
  - Have some degree of prescription about what must be done
  - Be related to a developmental model of growth in the domain (construct-referenced; Messick, 1975)
- Feedback is the critical link in a loop that makes formative assessment effective

Literature Review: Some Seminal Articles

- Crooks (1988), examined whether formative and summative assessment are compatible
  - He argued for separating the two functions
  - Feedback versus grading for example
  - Summative function has been too dominant
- Brookhart (2001) and others argue that each kind of assessment can be seen as parts of the same whole
- Biggs (1998), need to make use of both kinds of assessment; works best if both types are criterion referenced
Compatibility of summative and formative assessment?

- Test design, development, analysis, scoring, and reporting founded in test purpose
- Test adequacy and utility depend on these
- A test designed for one purpose may not function well for another (Popham, 1997)
- Kinds of items, scores, and reports differ
- Level of detail; breadth vs. depth
- Imbedded in instruction?

Literature Review: Some Seminal Articles

Difficulties with current assessment practice:

- Effective learning (Black & Wiliam, 1998b)
  - Tests used emphasize superficial learning and recall
  - Teachers appear to be unaware of the assessment work of colleagues and do not trust or use their assessment results
  - Both in questioning and written work, teachers' assessment focuses on low-level aims, mainly recall (Cizek et al., 1995)
  - Emphasis on quantity and presentation of work rather than on quality in relation to learning
Literature Review: Some Seminal Articles

Difficulties with current assessment practice:

- Impact and consequences of assessment
  - Overemphasis of grading function
  - Under emphasis on feedback and advice for learning
  - Competition rather than personal improvement
  - Approaches that emphasize comparison to other students resulting in “low performers”, “low ability students”

- Managerial role of assessment
  - Teachers feedback often serves social and managerial functions instead of learning functions
  - Teachers may be able to predict student performance but know too little about student learning needs or strategies for student improvement
  - Teachers may not use or may have no information on student performance from previous teachers
  - Grading is given higher priority and importance than analysis of student work for learning intervention
Suggestions from the Literature (based on Black & Wiliam, 1998b)

- Ensure active involvement of students
  - Self assessment and monitoring as interventions
  - Improved engagement
  - Enhance effect of assessment on student motivation and self esteem
- More integrated involvement of teacher
  - Emphasis on feedback useful for learning; may require changes in classroom practice
  - For assessment to function formatively, the results have to be used to adjust teaching and learning

Suggestions from the Literature (Stiggins, 2007)

- Clear learning targets
- High quality assessments that match targets
- Representative sampling of content domains
- Assessments that minimize bias
- Communication during the learning process
- District conditions:
  - Clear achievement expectations
  - Coordinated assessment systems
  - Assure assessment accuracy through literacy and evaluation
  - Communicate results in timely and understandable way
INTRODUCTION AND BACKGROUND
OVERVIEW OF THE WASHINGTON STATE LEGISLATION
PURPOSE OF THIS GUIDE
PURPOSES AND KINDS OF ASSESSMENT
  Summative Assessment
  Formative Assessment
  Diagnostic Assessment
  Progress Monitoring
  Norm Referenced, Criterion Referenced, and Standards Based Assessment
THE IMPORTANCE OF FEEDBACK
ISSUES IN THE USE OF DIAGNOSTIC ASSESSMENTS
  Implementation and Interpretation
  Problems, Pitfalls, and Constraints
  Other Classroom Assessment Strategies
  Technical Adequacy
DIAGNOSTIC ASSESSMENT FOR SPECIAL EDUCATION LEARNERS
DIAGNOSTIC ASSESSMENT FOR ELL LEARNERS
BRIEF BIBLIOGRAPHY

ISSUES IN THE USE OF DIAGNOSTIC ASSESSMENTS
  Implementation & Interpretation
  Problems and pitfalls
  Diagnostic Assessment for:
    - Special education learners
    - English language learners
Implementation and interpretation:

- The importance of refining test purpose
- Linkage of assessment with curriculum and instruction
- Professional development and staff support
- Explicitly design feedback strategies
- Design transfer and generalization assessments
- Analysis at level relevant for instructional intervention (e.g., item analysis)
- Design reports useful for instructional purpose (e.g., growth curves)

Problems and Pitfalls:

- Lack of time
- Assessment literacy
- Delayed access to results
- Narrowing the curriculum and teaching to the test
- Test that does not support purpose (e.g., two test forms when multiple assessment occasions needed)
- Test that measures construct irrelevant skills (e.g., economically disadvantaged students; Amrein & Berliner, 2002)
- Prerequisite knowledge and skills versus targeted content knowledge
- Technical adequacy at level of inferences (e.g., subscore reliability and validity)
### Washington Diagnostic Assessment Guide

- **Diagnostic Assessment for:**
  - ELL Learners
  - Special Education Learners
- **Review and consider time limits for all students (can reduce need for accommodations)**
- **Develop accommodations policy; same as WASL?**
- **Principle of universal design (design test to allow access by largest possible number of students)**
- **Has test gone through bias and sensitivity review?**

---

### Washington Diagnostic Assessment Guide: Assessment for Special Education Learners

- Diagnostic assessments play a critical role in the identification of students with special needs
- Individuals with Disabilities Education Act (IDEA; 2004) recognized RTI as a means for identification
- RTI requires diagnostic assessment and the integration of instruction and assessment
- Screening tests followed by diagnostic assessment (Fuchs, et al., 2003)
- Progress monitoring to determine if instruction results in growth
- Separate target skills from prerequisite and ancillary skills
- Provide accommodations when they will allow more valid inferences (IEP team)
Washington Diagnostic Assessment Guide: Assessment for English Language Learners

- Universal design applied to assessments for ELL students
- Language ability should not impede assessment on content measuring something other than language ability
- Need clear specification of the target skills being assessed
- For example, in a written science assessment with open-ended responses, is writing a target skill or an ancillary skill?
- Tests should use “simplified” or “plain” language; benefits all learners
- Train raters, scorers, graders in error patterns of ELLs or adjust rubrics (e.g., spelling errors on a science test)

Comparative Guide to Diagnostic Assessments

- Provides information on commercially available instruments in mathematics, reading, science, and writing
- Our review is based on principles from the Standards for Educational and Psychological Testing
- The guide provides information on important components and features of each assessment
**How to Choose an Assessment**
- Finding and reviewing instruments
- Match of assessment to purpose
- Content match and alignment, depth versus breadth; caveat about our work
- Technical properties
- Cost; state negotiated in future?

**Finding and Reviewing Instruments**
- Hopefully we’ve done most of the work
- Publisher’s web sites
- Importance of independent reviews and research literature
Comparative Guide to Diagnostic Assessments

Finding and Reviewing Instruments

- Buros Institute of Mental Measurements
  http://www.unl.edu/buros/bimm/index.html

- Listing of Buros test reviews:
  http://buros.unl.edu/buros/jsp/search.jsp

Comparative Guide to Diagnostic Assessments

Finding and Reviewing Instruments

- ETS Test Collection:
  http://www.ets.org/portal/site/ets/menuitem.1488512ecfd5b8849a77b13bc3921509/?vgnextoid=ed462d3631df4010VgnVCM10000022f95190RCRD&vgnextchannel=85af197a484f4010VgnVCM10000022f95190RCRD
Comparative Guide to Diagnostic Assessments

**TABLES IN GUIDE PROVIDE INFORMATION ON:**
- Test name, publisher, contact information
- Test purpose
- Descriptive information about test
- Details on test administration
- Content covered
- Scoring and reporting
- Technical adequacy including reviews

Example from the Comparative Guide
The Guide and the tables of assessments will also be made available to users on the OSPI web site.

Hope to provide easy accessibility and interactivity in obtaining information about formative and diagnostic assessment.

Examples of possible web page designs follow.

For each assessment, there will be links that provide an overview description of the assessment as well as a link to the publisher’s web site.

Other columns on the web page will provide access to more detailed information on each assessment including links to independent reviews.
Another design we have drafted is a “consumer reports” style presentation that summarizes information more briefly.

One possibility is to present this table first, then user could link to more detailed table when desired.

Based on NCTM standards; Number sense, Algebraic sense, Geometric sense, Measurement, Probability, Statistics and data

<table>
<thead>
<tr>
<th>Test Name</th>
<th>Description</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnostic Mathematics Test, Second Edition</td>
<td>Measures competence in the basic concepts and skills that are prerequisite to success in mathematics, while emphasizing problem-solving and problem-solving strategies. Test identifies specific areas of difficulty for each student so that teachers can plan appropriate intervention. Designed to be group administered, test provides both multiple-choice and optional free response assessment formats.</td>
<td>Formative, Diagnostic</td>
</tr>
</tbody>
</table>

Another possibility is to present this table first, then user could link to more detailed table when desired.

<table>
<thead>
<tr>
<th></th>
<th>◊</th>
<th>◊◊◊</th>
<th>◊◊</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exceeds minimum criteria</td>
<td>Meets minimum criteria</td>
<td>Does not meet or marginally meets minimum criteria</td>
<td></td>
</tr>
</tbody>
</table>

Summer Institute Workshops

- Three workshops on the Diagnostic Assessment Project are planned as part of the OSPI Summer Institute series
- In each workshop we will:
  - Introduce users to the Web site and its use
  - Introduce and discuss the Diagnostic Assessment Guide
  - Present the Comparative Guide of Assessments and provide guided practice on its use
References


---

Discussion

Thank you for your participation!

Contact:
jstevens.mesa@comcast.net
4524 Fox Hollow Road, Eugene, OR 97405