



**School Improvement
Specialist Training Materials:
Performance Standards,
Improving Schools, and
Literature Review**

**Module 7—
Purposeful Student
Assessment**

December 2005

Appalachia Educational Laboratory (AEL)
at

EDVANTIA™

Partners in education. Focused on results.

Edvantia was founded in 1966 as the Appalachia Educational Laboratory, Inc. (AEL); on September 1, 2005, AEL became Edvantia, Inc. The Regional Educational Laboratory for the Appalachian region is known as the Appalachia Educational Laboratory at Edvantia.

Edvantia is a nonprofit education research and development corporation, founded in 1966, that partners with practitioners, education agencies, publishers, and service providers to improve learning and advance student success. Edvantia provides clients with a range of services, including research, evaluation, professional development, and consulting.

For information about Edvantia research, products, or services, contact



P.O. Box 1348, Charleston, WV 25325 • 304.347.0400 • 800.624.9120 • fax 304.347.0487
One Vantage Way, Suite D-210, Nashville, TN 37228 • 615.565.0101 • fax 615.565.0112
info@edvantia.org • www.edvantia.org

© 2005 by Edvantia

All rights reserved. Except as permitted under the United States Copyright Act of 1976, no part of this publication may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without the prior written permission of the publisher.

Edvantia maintains policies and procedures that govern the Institutional Review Board in ensuring the welfare and protecting the rights of human subjects who choose to participate in research and evaluation activities.

This publication is based on work sponsored wholly or in part by the Institute of Education Sciences (IES), U.S. Department of Education, under contract number ED-01-CO-0016. Its contents do not necessarily reflect the positions or policies of IES, the Department, or any other agency of the U.S. government.

Edvantia is an Equal Employment Opportunity/Affirmative Action Employer.

The School Improvement Specialist Project prepared seven modules. School improvement specialists, as defined by the Appalachia Educational Laboratory at Edvantia, are change agents who work with schools to help them improve in the following areas so as to increase student achievement. These modules are intended to provide training materials for educators seeking professional development to prepare them for a new level of work.

Module 1—Shared Leadership

Module 2—Learning Culture

Module 3—School-Family-Community Connections

Module 4—Effective Teaching

Module 5—Shared Goals for Learning

Module 6—Aligned and Balanced Curriculum

Module 7—Purposeful Student Assessment

Each module has three sections:

1. Standards: Each set of content standards and performance indicators helps school improvement specialists assess their skills and knowledge related to each topic. The rubric format provides both a measurement for self-assessment and goals for self-improvement.
2. Improving Schools: These briefs provide research- and practice-based information to help school improvement specialists consider how they might address strengths and weaknesses in the schools where they work. The information contained in the briefs is often appropriate for sharing with teachers and principals; each includes information about strategies and practices that can be implemented in schools, resources to be consulted for more information, tools for facilitating thinking about and working on school issues, and real-life stories from school improvement specialists who offer their advice and experiences.
3. Literature Review: The reviews of research literature summarize the best available information about the topic of each module. They can be used by school improvement specialists to expand their knowledge base and shared with school staffs as part of professional development activities.

Purposeful Assessment
Content Standards and Performance Indicators for School Improvement Specialists
Self-Assessment Tool

Purposeful Assessment: This matrix measures the extent to which a school improvement specialist has the knowledge and skills to assist a school in developing and sustaining purposeful assessment as reflected by the following characteristics: (1) interpreting and using data, (2) understanding and use of best assessment practices, (3) knowledge of student assessment, (4) articulation of assessment for learning, and (5) knowledge of mandates and statutes regarding assessment.

Knowledge or Skill	Advanced	Proficient	Basic	Novice
1. Interpreting and using data	<p>The school improvement specialist</p> <ul style="list-style-type: none"> a. helps institutionalize the practice among the entire school community of using multiple sources of student performance data, as well as nonacademic data, to analyze the curriculum or instructional program b. teaches faculty to disaggregate data extensively c. facilitates strategizing sessions for using the data to set goals for student achievement d. facilitates school personnel in the use of data to monitor and adjust curriculum and instruction for the purpose of continuous improvement e. facilitates ongoing professional dialogue around student data and its relation to student achievement f. facilitates discussion around resources needed to reach goals; models ways to identify, evaluate, and obtain those resources 	<p>The school improvement specialist</p> <ul style="list-style-type: none"> a. teaches faculty and administration to interpret multiple sources of school data, including both academic and non-academic measures b. works with faculty to disaggregate data c. teaches faculty how to use assessment results to set goals for student achievement d. facilitates discussion of strengths and needs, and suggests ways to improve curriculum or instruction e. facilitates discussion of data analysis several times throughout the school year f. models ways to identify, evaluate, and obtain resources for schools to use in reaching their goals 	<p>The school improvement specialist</p> <ul style="list-style-type: none"> a. interprets multiple sources of student performance data and presents summaries to the entire school staff b. completes an extensive disaggregation of data and shares this information with the entire school staff c. suggests goals for student achievement based on data d. shares strengths and needs based on data and suggests ways to improve curriculum or instruction e. examines data several times throughout the school year, presenting summaries to the school staff f. identifies, evaluates, and obtains resources for schools to use in reaching their goals 	<p>The school improvement specialist</p> <ul style="list-style-type: none"> a. interprets school data from standardized assessments and presents summaries to the school administration b. completes a limited disaggregation of data and shares with the administration c. shares strengths and needs identified by the data d. examines school data at the beginning of the school year and at the end of the school year e. identifies some resources for schools to use in addressing needs

Knowledge or Skill	Advanced	Proficient	Basic	Novice
2. Understanding and use of best assessment practices	<p>The school improvement specialist</p> <ul style="list-style-type: none"> a. has a deep understanding of research connecting high-quality student assessment to student achievement and strategically shares this information with the school community b. has extensive knowledge of multiple assessment types and conducts ongoing dialogue with faculty about ways to match assessments with content c. demonstrates knowledge of a wide variety of formative assessments and teaches faculty to use these d. has a deep understanding of best-practice methods of classroom assessment and consistently shares and demonstrates these methods with faculty in an organized, strategic manner e. facilitates teachers in development and common grading of benchmark assessments related to specific standards; facilitates dialogue about the meaning of the results and the desired next steps f. thoroughly understands several cognitive models and facilitates various opportunities for teachers to learn to incorporate these into classroom assessment practices g. consistently models ways to provide timely and appropriate feedback to students and teachers on their progress 	<p>The school improvement specialist</p> <ul style="list-style-type: none"> a. possesses a broad understanding of research connecting high-quality student assessment to student achievement and explains this connection to the school community b. knows multiple assessment types and encourages faculty to vary assessment strategies c. models a variety of different formative assessments to teachers d. understands best-practice methods for classroom assessments and shares these methods with faculty in an organized, strategic manner e. models and demonstrates development and common grading of benchmark assessments related to specific standards f. understands and articulates several cognitive models and demonstrates ways to incorporate these levels of cognition into everyday assessment practices g. models and demonstrates methods for providing students with timely and appropriate feedback on their progress 	<p>The school improvement specialist</p> <ul style="list-style-type: none"> a. understands the research connecting high-quality student assessment to student achievement and shares this research with school administration b. can identify different kinds of assessments for faculty and staff, but continues to focus primarily on traditional methodologies c. encourages faculty to use frequent short-term formative assessments d. knows some best-practice methods of classroom assessments and shares these methods with teachers, as requested e. encourages the use of benchmark testing at set intervals throughout the school year, using the test data to modify instruction f. understands and articulates at least one model of cognition and encourages teachers to use assessments requiring higher cognitive levels of response g. shares some methods for providing students with timely feedback on their progress 	<p>The school improvement specialist</p> <ul style="list-style-type: none"> a. has minimal knowledge of research connecting high-quality student assessment to student achievement b. has some knowledge of assessment practices, but focuses primarily on summative assessments and traditional methodologies c. can locate sample assessment items from state and local sources and from the Internet d. has a rudimentary knowledge about ways to create classroom assessments e. is knowledgeable about benchmark assessments, but recommends only basic and easily located sources f. understands and can articulate at least one model of cognition, such as Bloom's taxonomy g. encourages teachers to provide students with timely feedback on their progress

Knowledge or Skill	Advanced	Proficient	Basic	Novice
3. Knowledge of student self-assessment	<p>The school improvement specialist</p> <ul style="list-style-type: none"> a. has a detailed comprehension of the value of student self-assessment and clearly articulates this value to the entire school community b. models and shares different self-assessment methods with the school community c. demonstrates extensive self-reflection methods, including both academic and behavioral indicators, and facilitates dialogue about the power of such instruments in increasing student achievement d. demonstrates motivational strategies that enable students to take responsibility for their own work and consistently facilitates faculty reflection on and development of further strategies 	<p>The school improvement specialist</p> <ul style="list-style-type: none"> a. can clearly articulate the value of student self-assessment b. demonstrates use of different self-assessment methods c. models and demonstrates several self-reflection methods, incorporating both academic and behavioral indicators, with teachers and students d. demonstrates motivational strategies that enable students to take responsibility for their own work 	<p>The school improvement specialist</p> <ul style="list-style-type: none"> a. understands the value of student self-assessment b. encourages faculty to use some student self-assessment methods c. shares self-reflection methods such as learning logs with teachers and students d. encourages teachers to use motivational strategies that enable students to take responsibility for their own work 	<p>The school improvement specialist</p> <ul style="list-style-type: none"> a. has an appreciation for the value of student self-assessment but has limited knowledge in ways to achieve it, concentrating instead on teacher-centered assessment strategies b. uses limited reflection techniques for self-assessment with teachers or with students

Knowledge or Skill	Advanced	Proficient	Basic	Novice
4. Articulation of assessment for learning	<p>The school improvement specialist</p> <ul style="list-style-type: none"> a. has a deep understanding of the concept of assessment for learning and mentors school leaders and others in creating a culture that sustains this philosophy b. facilitates the school community in understanding assessment as a key component of instruction c. facilitates faculty in understanding and using protocols for examining student work and leads reflective dialogue in connecting assessment to instruction d. models and demonstrates an in-depth use of student-centered assessment, such as student involvement in developing units of study and assessment strategies, along with ongoing collection of data on student learning; engages the school community in developing and discussing student-centered assessment strategies e. has a deep understanding of the barriers to changing assessment practices; works in groups and one-on-one to help overcome these barriers through information and consensus building 	<p>The school improvement specialist</p> <ul style="list-style-type: none"> a. understands the concept of assessment for learning and leads discussions within the school community about creating a culture of such assessment b. demonstrates principles and methods of assessing for learning to the entire school community in a strategic manner c. facilitates faculty in understanding and using protocols for examining student work and creates opportunities for collaborative discussion of student work d. demonstrates the use of student-centered assessment, such as student feedback on classroom experiences and student reflective logs on their learning styles e. has knowledge of barriers to changing assessment and works with the school to overcome these barriers 	<p>The school improvement specialist</p> <ul style="list-style-type: none"> a. understands the concept of assessment for learning and shares this concept with the entire school community b. creates and shares assessments for learning with faculty c. teaches faculty one or more strategies for examining student work and encourages this practice d. articulates the concept of student-centered assessment to faculty e. has a basic understanding of the barriers to changing assessment practice 	<p>The school improvement specialist</p> <ul style="list-style-type: none"> a. understands the concept of assessment for learning on a limited basis and shares this concept with administrators and some teachers b. locates examples of assessments that demonstrate assessment for learning and shares them with the faculty c. encourages faculty to examine student work collaboratively d. encourages faculty to design assessments prior to instruction and to get feedback from students

Knowledge or Skill	Advanced	Proficient	Basic	Novice
5. Knowledge of mandates and statutes regarding assessment	<p>The school improvement specialist</p> <ul style="list-style-type: none"> a. has extensive knowledge of federal, state, and local regulations and standards concerning student assessment b. can clearly articulate the rationale and history behind the various regulations and standards c. facilitates the school community in conversations toward understanding and respect for the regulations and standards d. engages in reflective dialogue with a professional network to understand, clarify, and possibly refine regulations and standards 	<p>The school improvement specialist</p> <ul style="list-style-type: none"> a. has extensive knowledge and understanding of federal, state, and local regulations and standards concerning student assessment b. can articulate some of the rationale behind the regulations and standards c. works with the school community to explain and clarify regulations and standards 	<p>The school improvement specialist</p> <ul style="list-style-type: none"> a. has knowledge of federal, state, and local regulations and standards concerning student assessment b. explains and clarifies regulations and standards to administration and faculty 	<p>The school improvement specialist</p> <ul style="list-style-type: none"> a. has knowledge of federal and state regulations and standards concerning student assessment, but may have limited understanding of local regulations b. explains regulations to school administration

Assessment for Learning

As long as we have had school, we have had testing, because educators need to know how well their students are performing at the tasks they are learning. For years, teachers have instructed, tested, and then issued student grades after the instruction was complete. This mode of testing represents assessment *of* learning. Summative assessments measure mastery and represent accountability for the teacher and the student. They are a way to determine the competence of any given student or group of students.

While assessment of learning is extremely important, it gives only part of the information needed to modify instruction and increase student learning. Students need to see that the purpose of assessment is more than telling them whether they succeeded or failed. Assessment can be a tool for navigating the voyage from not knowing to understanding. Formative assessment is also a tool for refining curriculum—the plan for instruction.

Formative assessment can help teachers and students make better judgments about what to do next. It asks teachers and students to become more explicit about the scope and sequence of instructional situations, to arrive at valid and reliable judgments about situations, and to make decisions for re-teaching students, changing instructional design, or improving curriculum design. By making formative assessment part of instructional design and curriculum development, districts can help teachers bring subconscious thoughts about situations, judgments, and next steps to the level where consensus can be reached about the best way to proceed (Black & Wiliam, October 1998).

British researchers Paul Black and Dylan Wiliam (March 1998) conducted an extensive review of the research on assessment for the previous nine years. The more than 250 journal articles and books published about the research reveal that using formative assessments in classrooms can significantly increase student learning as measured by end-of-year standardized assessments. These findings also show that the increase is greater with low-achieving students than with other groups.

Although research shows the value of formative classroom assessments, it also reveals some serious problems with the way formative assessments are used. Some tests encourage rote memorization. Too often, the primary purpose of assessments is to assign grades, which can encourage student competition rather than improvement. According to Black and Wiliam (October 1998), formative assessments are those used by teachers to revise their lessons to meet student needs.

This view of formative assessment appears in the work of Rick Stiggins and colleagues (2004) and Jackie Walsh and Beth Sattes (2005). Educators are addressing the value of formative assessment and discovering the best ways to implement it in the classroom. The use of formative assessment to provide consistent, specific feedback to teachers and students is termed assessment *for* learning. The educators mentioned above identify strategies that can be used in assessment for learning. For example, effective questioning: When teachers provide “wait time” of 3 to 5 seconds after posing a question

and after a student answers, the quality of student response increases dramatically (Barnette et al., 1995; Hunkins, 1995; Rowe, 1986; Tobin, 1987).

Other strategies for improving student performance include writing comments on student work. Comments show strengths and areas to improve, and help students to complete and refine their work. Likewise, when students have a clear rubric or scoring guide, they can effectively score their own and others' work. Such collaborative work among students can also increase learning and achievement.

In assessment for learning, students receive specific feedback about their own performance; they are not compared to other students. This feedback focuses on specific learning objectives and occurs throughout the learning period. Learning and assessment are integrally related one to another. The feedback helps students know where they need to increase their knowledge or skill, it helps teachers know what to emphasize in their lessons, and it demonstrates progress over time. With assessment for learning, students frequently engage in peer assessment or self-assessment. They are often able to lead conferences about their progress with their teachers and their parents.

A challenge for educators is to consider and embrace the concept of assessment for learning. We need to ask ourselves some tough questions. Do students have assessments that relate to real life and that require multiple tasks of higher order skills? Do we plan assessment as part of the lesson? Do we revise either or both if the assessment results tell us students do not understand the content? Do we provide scoring guides in the form of rubrics so students know what to aim for? Do we give specific feedback on performance so students can set their learning goals and know where they are going?

We do not want to abandon the concept of assessment of learning but we certainly want to balance it with a strong assessment for learning.

References

- Barnette, J., Orletsky, S., Sattes, B., & Walsh, J. (1995, April). Wait-time: Effective and trainable. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA (ERIC Document Reproduction Service No. ED383706)
- Black, P. & Wiliam, D. (October 1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139-48. Also available at <http://www.pdkintl.org/kappan/kbla9810.htm>
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (September 2004). Working inside the black box: Assessment for learning in the classroom. *Phi Delta Kappan*, 86(1), 8-21.
- Black, P. & Wiliam, D. (March 1998). Assessment and classroom learning. *Assessment in Education*, 7-74.
- Hunkins, F. P. (1995). *Teaching thinking through effective questioning* (2nd ed.). Boston: Christopher-Gordon Publishers.

Rowe, M. B. (1986, Jan.-Feb.) Wait time: Slowing down may be a way of speeding up! *Journal of Teacher Education*, 37(1), 43-50.

Stiggins, R. J., Arter, J., Chappuis, J., & Chappuis, S. (2004). *Classroom assessment for student learning: Doing it right—using it well*. Portland, OR: Assessment Training Institute.

Tobin, K. (1987) The role of wait time in higher cognitive level learning. *Review of Educational Research*, 57, 69-95.

Walsh, J. A. & Sattes, B. D. (2005). *Quality questioning: Research-based practice to engage every learner*. Thousand Oaks, CA: Corwin Press.

A New Look at Bloom’s Taxonomy

Today there is great emphasis on designing instruction that pushes students to go beyond simple recall of facts and develop their abilities to think analytically, write critically, and use knowledge to create something new. As we push students toward higher levels in the classroom, we should also help teachers use assessments that are at higher cognitive levels.

Bloom’s Taxonomy was created in 1956 as a tool for differentiating among levels of thinking, and most teachers are acquainted with it. In it, Benjamin Bloom defined six cognitive levels: knowledge, comprehension, application, analysis, synthesis, and evaluation. In 2001, Lorin Anderson and David Krathwohl published a book that presents the results of many years of collaboration among psychologists, curriculum experts, and researchers who contributed to refining and elaborating the original Bloom’s Taxonomy.

This revised version of Bloom’s Taxonomy has several features useful for designing and assessing classroom instruction, and long-time teachers may appreciate making its acquaintance. First, the terminology is slightly different and a bit easier to understand. The newly-defined levels—remember, understand, apply, analyze, evaluate, and create—are expressed as verbs rather than nouns. This is appropriate to talk about thinking, which is, after all, an action. Also, as shown in The Taxonomy Table, the two highest levels now appear in reverse order from the original: create (synthesis) is now at level 6 and evaluate (evaluation) is now level 5.

The Taxonomy Table

The Knowledge Dimension	The Cognitive Process Dimension					
	1 Remember	2 Understand	3 Apply	4 Analyze	5 Evaluate	6 Create
A. Factual Knowledge						
B. Conceptual Knowledge						
C. Procedural Knowledge						
D. Metacognitive Knowledge						

Reproduced, with permission, from Lorin W. Anderson, and David R. Krathwohl, *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom’s Taxonomy of Educational Objectives*. (Boston: Allyn and Bacon, 2001), p. 28.

As the Cognitive Process Dimension of the Taxonomy Table shows, another change incorporates a completely new feature: Now the cognitive processes are accompanied by four dimensions of knowledge—factual, conceptual, procedural, and metacognitive. When the cognitive processes are placed in a matrix with the dimensions of knowledge, teachers can classify instruction and assessment relative to both. For example, a question or test item that asks students to identify a species as vertebrate or invertebrate would fall on the matrix at cognitive level 2, Understand, and knowledge level B, Conceptual. A task that asks students to decide whether a vertebrate or an invertebrate could survive better in a particular environment moves up to cognitive level 4, Analyze, but stays at knowledge level B, Conceptual. Using a matrix such as this enables teachers to plan their assignments and assessments to take students beyond the levels of Remember and Factual Knowledge.

The Cognitive Process Dimension

Categories & Cognitive Processes	Alternative Names	Definitions and Examples
1. Remember—Retrieve relevant knowledge from long-term memory		
1.1 Recognizing	Identifying	Locating knowledge in long-term memory that is consistent with presented material (e.g., Recognize the dates of important events in U.S. history)
1.2 Recalling	Retrieving	Retrieving relevant knowledge from long-term memory (e.g., Recall the dates of important events in U.S. history)
2. Understand—Construct meaning from instructional messages, including oral, written, and graphic communication		
2.1 Interpreting	Clarifying Paraphrasing Representing Translating	Changing from one form of representation (e.g., numerical) to another (e.g., verbal) (e.g., Paraphrase important speeches and documents)
2.2 Exemplifying	Illustrating Instantiating	Finding a specific example or illustration of a concept or principle (e.g., Give examples of various artistic painting styles)
2.3 Classifying	Categorizing Subsuming	Determining that something belongs to a category (e.g., Classify observed or described cases of mental disorders)
2.4 Summarizing	Abstracting Generalizing	Abstracting a general theme or major point(s) (e.g., Write a short summary of events portrayed on a videotape)
2.5 Inferring	Concluding Extrapolating Interpolating Predicting	Drawing a logical conclusion from presented information (e.g., In learning a foreign language, infer grammatical principles from examples)
2.6 Comparing	Contrasting Mapping Matching	Detecting correspondences between two ideas, objects, and the like (e.g., Compare historical events to contemporary situations)
2.7 Explaining	Constructing	Constructing a cause-and-effect model of a system (e.g., Explain the causes of important 18th-century events in France)
3. Apply—Carry out or use a procedure in a given situation		
3.1 Executing	Carrying out	Applying a procedure to a familiar task (e.g., Divide one whole number by another whole number, both with multiple digits)

3.2 Implementing	Using	Applying a procedure to an unfamiliar task (e.g., Use Newton's Second Law in situations in which it is appropriate)
4. Analyze —Break material into its constituent parts and determine how the parts relate to one another and to an overall structure or purpose		
4.1 Differentiating	Discriminating Distinguishing Focusing Selecting	Distinguishing relevant from irrelevant parts or important from unimportant parts of presented material (e.g., Distinguish between relevant and irrelevant numbers in a mathematical word problem)
4.2 Organizing	Finding coherence Integrating Outlining Parsing Structuring	Determine how elements fit or function within a structure (e.g., Structure evidence in a historical description into evidence for and against a particular historical explanation)
4.3 Attributing	Deconstructing	Determine a point of view, a bias, values, or intent underlying presented material (e.g., Determine the point of view of the author of an essay in terms of his or her political perspective)
5. Evaluate —Make judgments based on criteria and standards		
5.1 Checking	Coordinating Detecting Monitoring Testing	Detecting inconsistencies or fallacies within a process or product; determining whether a process or product has external consistency; determining the effectiveness of a procedure as it is being implemented (e.g., Determine if a scientist's conclusions follow from the raw data)
5.2 Critiquing	Judging	Determining inconsistencies between a product and external criteria; determining whether a product has external consistency; determining the appropriateness of a procedure for a given problem (e.g., Judge which of two methods is the best way to solve a given problem)
6. Create —Put elements together to form a coherent or functional whole; reorganize elements into a new pattern or structure		
6.1 Generating	Hypothesizing	Coming up with alternative hypotheses based on criteria (e.g., Generate hypotheses to account for an observed phenomenon)
6.2. Planning	Designing	Devising a procedure for accomplishing some task (e.g., Plan a research paper on a given historical topic)
6.3. Producing	Constructing	Inventing a product (e.g., Build habitats for a specific purpose)

Reproduced, with permission, from Lorin W. Anderson and David R. Krathwohl, *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives* (Boston: Allyn and Bacon, 2001), p. 31.

A third feature of the new model is the addition of specific processes of thinking that occur within each of the six cognitive levels. For example, level 4, Analyze, has these cognitive processes: differentiating, organizing, and attributing. In addition, the taxonomy gives definitions and examples. The Cognitive Process Dimension Talbe provides clues about ways to scaffold learning for students and to build assessments that require various levels of thinking skills.

Individual teachers can use the revised taxonomy, but its greatest potential for power may come with systematic, collaborative use. Teachers working in grade-level or subject-area teams could identify the key standards for a subject. Then, by standard, they could develop a bank of questions or tasks at each of the six levels. Although this would not be an overnight task, the time spent could result in a bank of quality assessment items at many different levels. Teachers would then have a tool for creating common assessments and for differentiating instruction and assessment for all students.

When students become aware of the taxonomy, they also benefit. They learn a common language of learning and begin the invaluable process of metacognition—thinking about how they are learning.

Using a taxonomy to enhance teaching and learning can be a powerful tool. Walsh and Sattes (2005) point out four important considerations to using this or another taxonomy. First, actual levels of cognition or dimensions of knowledge depend on the context of the question and the student's knowledge and experience. Second, as Kathleen Cotton (1988) discovered, although teachers may pose higher order questions, the students' responses do not match the cognitive level of the question approximately 50 percent of the time. One way to attack this problem is to teach the taxonomy to the students. The third consideration is that most textbooks and teachers' manuals contain questions at the lower end of Bloom's Taxonomy. Teachers need to create their own test banks of higher order questions. Finally, there is an assumption that lower ability students cannot answer higher level questions. However, research tells us that all students can think at higher levels if they have adequate support and instruction (Bulgren, et al., 2002).

References

- Anderson, L. W., & Krathwohl, D. R. (Eds.), (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's Taxonomy of Educational Objectives*. Boston: Allyn & Bacon.
- Bulgren, J. A., Lenz, B. K., Marquis, J., Schumaker, J. B., & Deshler, D. D. (2002). *The effects of the use of the question exploration routine on student performance in secondary content classrooms* (Research Report No. 20). Lawrence: Kansas University, Institute for Academic Access. (ERIC Document Reproduction Service No. ED469289)
- Cotton, K. (1988). *Classroom questioning* (School Improvement Research Series: Research You Can Use, Close-up #5). Portland, OR: Northwest Regional Educational Laboratory.
- Walsh, J. A., & Sattes, B. D. (2005). *Quality questioning: Research-based practice to engage every learner*. Thousand Oaks, CA: Corwin Press.

A Primer of Common Assessment Terms

Every educator needs to be familiar with common assessment. The following are some terms and their definitions:

Reliability refers to consistency. A reliable test is one in which students obtain the same or very similar scores each time they take a test or an alternate form of the test proving that (1) relevant instruction or learning has NOT occurred between test administrations and (2) the student does not remember the questions or his/her responses from one test administration to the next.

Validity means the test measures what it is supposed to measure. With a valid test, educators can draw reasonably accurate conclusions about student learning.

Bias occurs in test items that offend or penalize a group of students. Tests can have gender, ethnic, disability, or other biases. The ideal test has no biased items.

Norm-referenced tests measure students against a norm (average). In standardized assessments, large groups of students participate in a test administration and the average score for that group becomes the norm (50th percentile). That is, half the population taking the test will be above the 50th percentile and half will be below.

Criterion-referenced tests measure students against a set of standards, often called “performance standards.” Students are tested to determine whether they master specific standards.

Constructed-response items on a test require students to create a response to a question. An example would be “Name three major reasons for the U.S. Civil War.”

Selected-response items on a test require students to choose from a selection of possible answers. Examples are multiple-choice or true-false questions.

Formative assessment occurs during instruction as a progress check. It is often used to evaluate the teaching and to give feedback to students.

Summative assessment is given at the end of a lesson or a unit to determine what the student learned as a result of the instruction.

Backward Design

Begin with the end in mind. This adage, popularized by Steven Covey in his book *Seven Habits of Highly Effective People*, is at the heart of the principle of backward design advocated by Jay McTighe and Grant Wiggins. In their groundbreaking book, *Understanding By Design*, McTighe and Wiggins submit that “teachers are designers. An essential act of our profession is the design of curriculum and learning experiences to meet specified purposes” (1998, p. 7). Unfortunately, according to the handbook that

accompanies their work, teachers often begin this design task by going to “textbooks, favored lessons, and time-honored activities rather than deriving them from targeted goals or standards” (1999, p. 37).

Backward design involves teachers in a three-stage process that begins with identifying desired results or outcomes. In the first stage, teachers look to national, state, and district standards as well as to the needs and interests of their students in order to answer the following kinds of questions: What should our students know and be able to do at the end of this unit of study? What are the enduring understandings or essential questions that will drive content selection?

Stage two in the process calls for teachers to decide what evidence is required to demonstrate that students have indeed attained these outcomes. During this stage, teachers select and design assessments. They answer such questions as: How will we know if a student has the knowledge, skills, and understandings associated with the desired results? How will we measure these student learnings? McTighe and Wiggins (1998, 1999) suggest the use of a variety of assessment measures at different points in the learning task. Again, the Understanding by Design (UBD) framework deviates from the traditional approach in which teachers typically design assessments as a last task in the instructional process.

The culminating stage of the UBD is the planning of learning experiences and instruction. Here, teachers answer the following kinds of questions: What enabling knowledge (i.e., facts, concepts, and principles) and skills (procedures) will students need to perform effectively and achieve desired results? (1999, p. 40) What activities will engage learners in experiences that will lead to attainment of desired results? What materials and resources will we need to provide? According to the authors, teachers should be concerned with “uncoverage of concepts” rather than just coverage of material during this stage. Uncoverage involves finding out students’ prior understandings of a particular concept (e.g., gravity) and determining whether the student may be entering the unit of study with misconceptions. Directly addressing misconceptions is an important part of instruction.

Many schools across the country are using UBD. The Association for Supervision and Curriculum Development (ASCD) offers a range of products and services to support educators who are adopting this approach. These are available on the ASCD Web site at www.ascd.org.

References

- Covey, S. R. (1989). *The seven habits of highly effective people*. New York: Simon & Schuster.
- McTighe, J., & Wiggins, G. (1999). *The Understanding by Design handbook*. Washington, DC: Association for Supervision and Curriculum Development.
- McTighe, J., & Wiggins, G. (1998). *Understanding by Design*. Washington, DC: Association for Supervision and Curriculum Development.

A Six-Step Process for Using Student Data

Because so many data are available these days, many teachers find them overwhelming and view assessment data as additional sets of records to file. Given the other demands on teachers' time, we need to offer a simple and straightforward process for data analysis as we encourage the use of data. The following process, presented in Edvantia's *Data Connections* CD-ROM, provides an easy-to-use template for individual and team use, and it can be a tool for school improvement specialists to use in their work.

- 1. Determine your purposes for analyzing the data.** Before you gather your data, prepare by thinking about what you want to learn from the scores. Do you have a school or district requirement you are addressing? Are you reviewing the effectiveness of your instruction? Or are you concerned about the performance of a student or group of students? Your goal is to derive information that will guide you in tailoring your curriculum and in meeting the specific needs of each student.

For example, you may be analyzing the data to determine

- how well your students performed at the end of the year
- changes you might make to your instruction
- how well classroom assessments predict performance on statewide assessments
- an understanding of new students

- 2. Develop a question list.** After you know your purpose for analyzing the data (the "why"), you need to determine the specific questions you want to answer. What questions do the data raise about the curriculum, your students, and your teaching? For example,
 - In which areas did students perform well, and in which areas do they need more work?
 - Did my students who performed well in class also do well on state assessments?
 - How well are students doing over time?
 - How should I group students for differentiated instruction?

You might be able to think of many questions. Focus on one or two of your most pressing questions because you can't accomplish everything in one year. Think of this as an ongoing process. As you make improvements in your teaching effectiveness, you can look at new issues each year.

- 3. Familiarize yourself with the data.** Gather all the data you need. Consider a variety of classroom data, including gradebook, attendance records, past standardized tests, student portfolios, and qualitative data such as notes from parent conferences, observations of students, and so on.

4. **Organize the data.** Sometimes you'll get data in the form of feedback sheets from the test publishers along with standardized test scores. These feedback sheets might be useful, or you may need to organize the data differently in order to answer your questions.

5. **Answer your questions.** Using your judgment and professional expertise, can you answer the questions you raised in step 2? If you can't answer the questions you raised, what data are you missing that could answer them? Sometimes you need to go beyond your gradebook and state report card.

6. **Develop action steps.** The actions you take should relate directly to your questions. They may be pedagogical in nature or could relate to classroom management or test-taking strategies. Colleagues in your subject area are a good resource for addressing pedagogical concerns. If your students appear to have difficulty with a particular content area or set of standards, find out how other teachers in your subject address this content. Are they successful, or is there a trend that could be addressed by new activities, resources, or even professional development?

The most critical aspect of your action steps is to develop steps that relate directly to your questions and the answers you discovered. If the data identify a problem with classroom assessments, your action steps should address classroom assessments. If the data identify attendance problems for an individual, your action steps should address solving those attendance problems.

<i>If your question is</i>	<i>then you should</i>
In which areas did my students perform well/poorly?	Look for group trends on standards.
Did my students who performed well on classroom assessments also perform well on statewide assessments?	Indicate test scores in your grade book and compare them with state data.
How well are students doing over time?	Collect data, such as state data, from the past three years.
How should I group students as I attempt to differentiate instruction?	Note levels of performance on both classroom and state assessments and determine an effective grouping of students.

As you identify problems, you also need to find solutions. Here are some problems and suggested solutions.

<i>If you find</i>	<i>then you should</i>
High absences correlate to poor performance.	Communicate with parents.
Whole class performance is poor on standardized test.	Review and adjust curriculum and instructional strategies to better align with standards and test.
Half of class did better than other half.	Review class assignments. Examine how activities are structured.
Performance is low across three years.	Meet with prior teachers to align curriculum so that incoming students are better prepared.

Data Connections

Data Connections: Using Assessment to Improve Teaching and Learning is a multimedia CD-ROM that helps educators better analyze, understand, and apply standardized test data in their classrooms. The course also helps teachers make connections between standardized test data and other types of classroom assessment data. Most important, it assists learners in determining how to use these data to modify the curriculum and improve student achievement through a series of lessons on

- test score utilization
- curriculum alignment
- classroom assessment
- test-taking strategies

For more information on *Data Connections*, visit www.edvantia.org.

Book Review

Accountability for Learning: How Teachers and School Leaders Can Take Charge by Douglas Reeves. Alexandria, VA: Association for Supervision and Curriculum Development, 2004.

Douglas Reeves has spent years researching student assessment and teacher accountability. In his new book, Reeves takes a strong stance for student-centered accountability—looking at what students have learned as a result of their instruction. Using this approach, Reeves advocates a holistic evaluation system.

Reeves identifies four components of teacher leadership in accountability required for students to reach their potential. The first component is observation of teaching processes and practices. Teachers must observe with unbiased clarity how often students resubmit work, how carefully teachers themselves follow best practices, how often and how well they collaborate with other teachers, and so on. As they observe, they begin the second component of leadership for accountability: reflecting on what worked, what didn't, and what could be improved. The third component is synthesis, or putting together

conclusions from observations and reflections with information from available research. Finally, the fourth component is commitment to replicate practices that work. For example, Reeves asks, “We know that collaborative scoring of student work is associated with higher levels of fairness and greater levels of student performance. Will we expand it?” (p. 55).

Reeves considers the challenges presented by teaching students from poverty, students from severely broken homes, students who do not speak our language, and a host of other issues. When he examines schools that have high student performance, Reeves finds that performance cannot be explained by differences in student demographics or available funds. He sees improved student performance as being tied to nine key professional practices of teachers and leaders. Among these are several related to assessment:

- strong collaboration in examining student work, collectively defining excellence, and building and collaboratively scoring common assessments
- commitment to frequent feedback to students in the form of coaching, practice, and clear goals
- schools’ consideration of multiple sources of student data and observation of data for groups of students over time
- common assessments that are used consistently and scored collaboratively

Reeves encourages the use of standards-based assessments rather than comparative assessments. It is not enough to know that a group of students performed better than another group if neither group met a specific standard. Reeves differentiates between performance indicators (e.g., adding double-digit numbers) and performance standards (e.g., adding double-digit numbers with 95% accuracy).

How do we reach the standard? Reeves is very clear: “Classroom assessment, created and scored by classroom teachers, is the gold standard in educational accountability” (p. 114). Only through constant formative assessments and collaborative conversation about scoring can a faculty collect sufficient data to alter approaches, to integrate curriculum, and to assist students in meeting the standards.

This is exciting reading. In addition to offering the results of his research and challenging schools to replicate it, Reeves offers an appendix that outlines ways to develop holistic accountability from the school board office to the classroom.

Book Review

Getting Excited About Data, Second Edition: Combining People, Passion, and Proof to Maximize Student Achievement by Edie Holcomb. Thousand Oaks, CA: Corwin Press, 2004.

This second edition of Edie Holcomb’s *Getting Excited About Data* focuses on how educators can connect the data at hand with their passion for teaching and learning.

Acknowledging what the book is not—a comprehensive, bureaucratic, impersonal quick-fix book—Holcomb proclaims that this edition is about the human element.

From how to use data for alignment and achievement to how to engage people, Holcomb suggests ways to involve educators in asking tough questions and finding answers for themselves. Of special interest are the many charts and graphs that offer ideas for collaborating, surveying for interest and knowledge levels of teachers and students, and whole-school planning. The book, however, is not “touchy-feely.” There is plenty of hard data talk, just in user-friendly terms.

School improvement specialists will appreciate Holcomb’s tried-and-true ideas for consulting with a faculty using Carousel Data Analysis, Reporting Out, Round Robin Listing, Prioritizing, Data Day, and Go for the Green. In Chapter 11: Drilling Down the Priority Data, she shares the how-tos for planning forward with student-specific data and planning backward with skill-specific data. Step-by-step reporting of the author’s consulting ventures with reluctant-then-appreciative participants hammers home her message that applying data to improve schools is everybody’s business.

Chapters 14 and 15 offer a detailed look at how to convert data and priorities to a school action plan; they lead and end with realistic ideas for “Sustaining the Struggle” and “Relentless Resilience.” In the author’s relaxed, first-person commentaries detailing what works in schools (and sometimes what doesn’t), readers will find practical, insightful ideas that reveal Holcomb’s zeal for combining people, passion, and proof.

A Portfolio Saved the Day

The school improvement specialist stories that appear in Improving Schools come from real life. The names have been changed or removed to preserve confidentiality.

The young man, Fred, arrived as the school’s new Spanish teacher in the middle of the year. The previous teacher had resigned after a long illness and many administrative problems. Fred’s directions were to finish the year and teach as much Spanish as he could in the remaining 14 weeks. Noncertified and completely unprepared for teaching adolescents, he came to me, the school improvement specialist, in a panic. “They know nothing! I gave a quiz and they failed it miserably. Most of them have an F for the first semester’s grade. Their lack of vocabulary is amazing. How can I ever teach them something in such a short time?”

Fred and I talked at length and decided to gamble on some strategies these students had never used before. Given the students’ apathy and history of failure, we decided to give the students choice and responsibility by having them build portfolios.

As Fred taught his first unit, he had students grade their own and one another’s work. He kept student portfolios in his class file cabinet, and students put their work in the portfolios. At the end of the unit, Fred gave these instructions: “Choose your best

work from the areas of homework, quizzes, and translations. Take time to correct the work to make it the best you can. Then write an explanation of why this work is excellent and what you learned from it. Your grade will depend on the quality of the three pieces you turn in and your explanations.” With my help, he crafted a rubric to guide the written explanations.

Students were amazed that they were only going to have to turn in three pieces for a grade. We heard comments like, “No fair! Why did I do all those other things if he wasn’t going to use them?” We also heard, “Gosh, you mean we can make it better—won’t that mean a better grade for us?”

Fred and I held our breath, but the students actually went to work on the task. Even students who had spent much time with their heads down on their desks came to life and tried to resurrect some assignments. The results were more than credible. The three pieces the students turned in showed students that they were capable of producing quality work. The companion reflections made the students consider what they were learning and how they learned it.

We tried another unit. Then another. Student learning improved gradually, with some students making remarkable progress. Fred decided he needed to collect what he now called “samples of excellence” more frequently than at the end of a unit. He got feedback from the students on refining the rubric. When he became discouraged at the quality of some of the work, he learned to give more powerful assignments. At the end of the year, Fred asked the students to prepare their portfolios to present to their Spanish teacher next year. Fred’s students spent their time perfecting recent work, writing messages to the new teacher about their learning, and making suggestions for next year’s class. Some students went all out with artwork and audiotapes.

Did these students who had been deprived of good Spanish instruction for most of the year make up their deficits? Fred and I don’t think so. But we do know that the practice of peer and self-assessment, combined with reflection and the opportunity to perfect a few assignments, made a big difference. Fred created motivation and opportunity to learn. The student portfolios gave students confidence that they could learn.

Student Self-Assessment Instruments

- Weekly learning logs
- Reflective responses (The most interesting thing I learned today was _____. The thing I still have a question about is _____.)
- Self-assessment checklists
- Rubrics
- Teacher-student interview
- Response to this question: The reason this work is high quality is _____.
- “Traffic light” icons to label a level of understanding: good, a little, or no understanding.

Book Review

Using DATA to Improve Student Achievement: A Handbook for Collecting, Organizing, Analyzing and Using Data by Deborah Wahlstrom. Suffolk, VA: Successline, Inc., 2002.

Deborah Wahlstrom's book opens with a rationale for data collection, then moves into user-friendly types and sources of data in three categories: outcome, demographic, and process. She presents a four-step process for looking at data: collect, organize, analyze, and use.

Chapter 3 defines and illustrates terms for "talking the talk" of data. Chapters 4 and 5 "walk the talk" of data disaggregation and the construction of useful tables. Throughout these chapters are practical, relevant ways of looking at data, presented in readable, informative charts, graphs, and definitions.

How-to sections (Chapters 6 and 7) lead readers into creating visual tables and pose questions for "thinking deep." From discussions about criterion-referenced questions to attendance and various forms of assessment, Wahlstrom provides factors to analyze for a comprehensive school study. Quick tips, ready-to-use templates, and data organizers invite hands-on participation.

The final three chapters put Wahlstrom's philosophies to work as readers find uses for data in curriculum alignment, student-centered instruction, and whole-school planning. Everything from charting to implementing and monitoring is discussed in easy-to-read text supported with illustrations. This book could become one of the most-used resources for faculty collaboration.

Rubrics For Assessment

A class has a project due in two weeks. The students ask, "What counts? What do we need in the project? How will I know if my project is good enough?" The teacher receives a set of projects and wonders, "How will I assign a grade to these? What factors are really important? If one has clever artwork and another has no artwork, does that make a difference?"

Thoughtfully creating a rubric prior to making the assignment can make life much easier for the students and the teacher. Spence Rogers tells us that a rubric clarifies "what counts" in an assignment. Most rubrics pick the key criteria of an assignment and detail the different levels of quality for those aspects. Look at the example below.

Rubric for History Timeline				
	Outstanding: 4	Very Good: 3	Could Improve: 2	Needs Help: 1
Accuracy of Information (80%)	All required information is provided, with additional clarifying or interesting information. Events appear in the correct sequence.	All required information is provided. Events appear in the correct sequence.	Some information is missing. Some events are missing or not in the correct sequence.	A significant amount of information is missing and/or not in correct sequence. Timeline is difficult to read because of the errors.
Presentation Qualities (10%)	Easy to read with neat lettering. There are no erasures or illegible writing. The timeline has additional features, such as pictures or graphics.	Easy to read with neat lettering. Writing is large enough to read and evenly spaced. There are no erasures or illegible writing.	The timeline can be read and understood. The events may not be evenly spaced and/or writing may not be large enough. May contain some erasures or illegible writing.	The timeline has erasures and/or illegible writing that interfere with readability and understandability. Spaces and size of writing are erratic/uneven.
Mechanics (10%)	No errors in capitalization, punctuation, or spelling.	Only minor errors in capitalization, punctuation, or spelling.	A few errors in capitalization, punctuation, or spelling.	Significant errors in capitalization, punctuation, or spelling.

In this case, the teacher wants accurate information and also a clean and clear presentation and proper mechanics. If you were a student, wouldn't you know how to create your timeline? And wouldn't the teacher find it easy to assess the quality of your work? In addition to identifying and describing the key criteria for the timeline, this rubric spells out how much each category counts toward the grade.

General Types

Judith Arters defines two general types of rubrics that have different uses and values. A *holistic* rubric provides a single score based on an overall impression of the student's work. Holistic rubrics are good for getting a "snapshot" of the quality of work. They are also fairly quick to grade, making them excellent choices for large-scale assessments, such as state writing assessments. Holistic rubrics are fine for judging work

that has only one trait to score or for judging the overall impact or success of a work. For example, how well did the student persuade the reader?

Analytic trait rubrics divide the task or assignment into areas or traits that can be judged separately. An example might be a writing rubric that contains separate gradations for organization, voice, word choice, and mechanics. Analytic trait rubrics are good for judging complex assignments. These rubrics provide specific information and feedback to students and help the teacher identify individual and class strengths and weaknesses for instructional planning. Analytic trait rubrics are excellent for peer assessment or for coaching students to meet a higher standard. Because they can provide formative feedback to students, these rubrics are most often the choice for general classroom use.

Tips for Creating Sound Rubrics

- Decide on the key criteria for the assignment and create levels of quality for each.
- A 4-point rubric is comfortable for classroom use. There is always the implied “zero,” which means no work was turned in.
- For a 4-point rubric, define “3” really well. A score of “3” meets the standard; anything better is a “4.” Thereafter, it is fairly easy to define a “2” and a “1”.
- Have levels build on one another; for instance, a “3” has everything a “2” has and more (more information or fewer errors).
- When possible, aim for a generic rubric that can be used for several assignments. The rubric can be modified and improved over time, and students can become familiar with it.
- Look for examples of quality work. Use the examples to build your rubric and to share with students.
- Post the rubric in the classroom and give copies to students and parents. Rubrics should not be secret.

(Adapted from *The High Performance Toolbox* by Spence Rogers and Shari Graham.)

Student Involvement

Involve students in creating a rubric. Spence Rogers and Judith Arters suggest that teachers provide exemplars—examples of excellent work. Students then form groups to brainstorm the qualities common to the examples. Those commonalities and their descriptions become the beginnings of a rubric. Building a classroom rubric is a powerful learning experience because the group must analyze and compare the exemplars, pull out key elements, and create definitions. Using these higher order thinking skills gives students insights into completing their own tasks and builds ownership and self-confidence. (Rogers & Graham [1998]; Arters & McTighe [2001]).

When to Use a Rubric

When is it important to have a rubric? Not every assignment needs a rubric, but experts say a class should have a rubric, even a very simple one, for all constructed-response assessments.

Arters and Rogers suggest using rubrics, especially the analytic trait variety, to coach students toward meeting a specified performance goal. If the purpose is to assess for learning by giving students specific feedback, the rubric will do that and grades can be assigned through logic or equivalencies (Rogers & Graham, [1988] p. 201; Arters & McTighe, [2001] p. 81).

Online Resources

<http://rubistar.4teachers.org/index.php>

<http://myt4l.com/index.php>

MyT4L and RubiStar are sites that offer templates for creating rubrics, plus examples.

<http://www.pz.harvard.edu/Research/RubricSelf.htm>

Harvard's Project Zero presents research on the effects of using rubrics in the classroom.

<http://school.discovery.com/schrockguide/assess.html#web>

Kathy Schrock's Guide for Educators offers instructions, sample rubrics, links to other resources, and more.

<http://edtech.kennesaw.edu/intech/rubrics.htm>

The Georgia Education Technology Training Center at Kennesaw State University presents information on making and using rubrics.

References

- Arters, J., & McTighe, J. (2001). *Scoring rubrics in the classroom*. Thousand Oaks, CA: Corwin Press.
- Burke, K. (2001). *How to assess authentic learning*. Arlington Heights, IL: Skylight.
- Marzano, R., Pickering, D., & McTighe, J. (1993). *Assessing student outcomes: Performance assessment using the dimensions of learning model*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Popham, W. J. (2003). *Test better, teach better: The instructional role of assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Rogers, S., & Graham, S. (1998). *The high performance toolbox: Succeeding with performance tasks, projects, and assessments*. Evergreen, CO: Peak Learning Systems.

Solomon, P. G. (2002). *The assessment bridge: Positive ways to line tests to learning, standards, and curriculum improvement*. Thousand Oaks, CA: Corwin Press.

The Dreaded Algebra Test: How Some Teachers Attacked It

The school improvement specialist stories that appear in Improving Schools come from real life. The names have been changed or removed to preserve confidentiality.

Teachers in a large urban high school where I was the school improvement specialist remained in a constant state of worry and concern for their Algebra I students. These students needed to pass a state examination to graduate from high school, and many of them lacked even rudimentary math skills. As they considered strategy after strategy, teachers read Mike Schmoker's work (1999), which advocates a clear and focused goal, collaborative teamwork, and consistent use of student data. They read literature from Douglas Reeves (2004), Rick DuFour (2003), and others who advocate developing and administering common benchmark assessments, then scoring them collaboratively.

The teachers decided to give common assessments a try. To start, the teachers sat down and charted a pace for teaching each of the standards during the year. They decided to give a test every Friday to assess mastery of the standards they taught that week. As they graded the tests on Friday afternoons, they soon realized that the students were not retaining knowledge of the algebraic concepts from week to week. What could be the answer? With students who came to school without the prerequisite background knowledge, the teachers knew they needed repetition and constant review of the key principles.

To include an intensive review of all concepts each Friday seemed an impossible task: the quiz would quickly turn into a two-hour test. Instead, the teachers decided to review all the previous concepts and indicators during the first 10 minutes of every class. Over the course of the year, the teachers developed a class "starter" activity for each day. This activity consisted of 10 math problems to be completed in 10 minutes. Each of the 10 questions reviewed a key algebraic concept. Every week, students completed 50 review problems and graded their own work with the teacher's assistance. As new indicators were taught, they gained a place in the starter activities. When the teachers discovered concepts the students were really struggling with, those concepts earned more places on the starter activities.

Meanwhile the teachers continued their Friday collaborative assessments. Completing these tests and grading them together enabled the teachers to discuss their teaching methods, to talk about students' key needs, to modify their pacing charts, and, of course, to modify their starter activities.

These Algebra I teachers had a clear focus, strong teamwork, and used consistent data to provide a strong math program to their students. The first year of their collaborative efforts yielded impressive results. By their third year of collaborative problem solving and grading assessments, the students in their high school had reached a 93% passing rate on the state exit examination. Both teachers and students reached an

amazing goal. They demonstrated the power of a clear focus in teaching that was backed up by using and scoring common assessments.

References

- DuFour, R. (2003, Summer). Leading edge: "Collaboration lite" puts student achievement on a starvation diet. *Journal of Staff Development*, 24(3), 63-64.
- Reeves, D. (2004). *Accountability for learning: How teachers and school leaders can take charge*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Schmoker, M. (1999). *Results: The key to continuous school improvement*. Alexandria, VA: Association for Supervision and Curriculum Development.

Zeros for Missing Work?

Douglas Reeves, in his online newsletter from the Center for Performance Assessment (<http://www.makingstandardswork.com>), offers an interesting opinion about giving a zero when a student does not turn in an assignment. He points out that grades reflect relatively equal intervals. For example, an A is 90-100, a B is 80-89, a C is 70-79, and a D is 60-69. If so, wouldn't an F be 50-59? Reeves points out that a grade of zero is about six times as bad as an F!

Think about what a zero does to a student's grade. Suppose Charlie is a borderline student with 10 grades that average 71. If he fails to turn in his next assignment and gets a zero, his average falls to 64. If he gets two zeros, the average falls to 59. Three zeros drop the grade to 53. It's easy to see that a couple of zeros can devastate a borderline student's average.

Yes, says Reeves, we need to emphasize timeliness and responsibility, but we also need to teach the Charlies in our classes. Because many students are not bothered by a low grade or a zero, wouldn't a better "punishment" be a requirement to complete the work? If students who do not turn in work must complete it by forfeiting time from lunch, ball practice, homeroom, or the like, they may soon learn to complete work on time.

Reeves advocates giving a grade of 50 for work that is never turned in, so the student can ultimately retrieve his grade average. If Charlie received three grades of 50 for his missing work, his new average would be 66. He's still failing the class, but his chances of bringing the semester average to passing have just increased dramatically. If the teacher assigns Charlie additional work to compensate for the missed work, the teacher has the opportunity to give work that may be more appropriate to his developmental level than the original, missing assignment.

Think about it. Giving a zero is often a punishment rather than an attempt to provide a student with feedback. Doesn't teaching mean offering better alternatives?

When Planning Is Not Enough

The school improvement specialist stories that appear in Improving Schools come from real life. The names have been changed or removed to preserve confidentiality.

It was my second year as a school improvement specialist at a middle school that had never made adequate yearly progress. The school leaders and I were not totally discouraged, but a number of school improvement ideas had not taken hold as we expected.

When the principal returned from a fall retreat, she greeted the leadership team with enthusiasm: “We need to do some formative assessments!” (I smiled approvingly.) “You know,” she said, “we purchased formative assessments from a company last year, but the questions didn’t really match the standards we taught. All of us need to get together and write our own, and then see how the students perform.” (I smiled again, stopping short of a standing ovation and a big shout-out. After all, a good idea from any source is a good idea.)

So the principal wrote an e-mail to the staff: “If you are a teacher of a core academic subject, please turn in four multiple-choice questions that cover the standards you have taught to date.” She assigned a due date one week hence, a drop-off location, and a teacher to coordinate and publish all questions. At the end of the week, however, not a single question had been submitted, and we were within three weeks of doing a whole-school formative assessment. My conversations with teachers made it clear that this was a principal directive with no explanation, no support. They just didn’t know what was expected of them.

I asked the principal if I could facilitate a session on how to write quality questions aligned with standards at the next grade-level meetings. She liked the idea, so I worked up some activities and led teachers in discussion. They assured me they understood what to do. At the end of another week, we still had no questions for the formative assessment.

After a frank discussion with the principal, we both agreed that teachers still were unclear about how to construct good questions, hadn’t bought in to the idea of formative assessments, or had little experience in collaboration. For the next grade-level meetings, we decided on a make-and-take format.

Teachers arrived with pacing guides, texts, and workbooks. We provided sample questions, paper, and pencils. The teachers looked at the standards they had taught, worked in groups, and produced questions in the same format as the state test. Almost without fail, teachers thanked us for the opportunity to work together. The principal and I were gratified by both their productivity and the well-crafted questions they produced.

When the principal and I reflected on the experience, we agreed that planning without clear explanations doesn’t always work. The right atmosphere, appropriate

materials, and enough time for collaboration—along with a lot of team work and a little “gentle persuasion”—took us a long way toward accomplishing our goal.

Contributors to this issue of *Improving Schools* include school improvement specialists Rusha Sams and Susan Hudson and Appalachia Educational Laboratory at Edvantia staff members Jackie A. Walsh and Nancy Balow.

**Purposeful Student Assessment—
A Review of Selected Research Literature**

David A. Squires
Immacula Didier-Gabriel

December 2005

Appalachia Educational Laboratory (AEL) at Edvantia

Introduction

The accountability demanded by the No Child Left Behind (NCLB) Act of 2001 has focused educators' attention on assessment at multiple levels (grade, course, school, and district) and on the results of various groups of students within the overall student population (racial and ethnic groups, economically disadvantaged, students with disabilities, and limited English proficient). This heightened awareness has created a need for classroom assessment strategies to ensure the success of all students.

Review of the Literature

Assessing or evaluating student performance is not new. Tyler (1966; cited in Jones, 2001) describes three purposes of assessment: (1) to assign grades or classify students, (2) to plan for subsequent teaching, and (3) to evaluate the effectiveness of a change in curriculum or methods.

Black and Wiliam (1998a), however, write about only two functions of assessment—*formative* and *summative*. Summative assessments are very much a part of teachers' work. Teachers use the results of summative assessments to report to parents through periodic report card grades and/or teacher-parent conferences, and these assessments determine whether students pass the course and are promoted to the next grade. Summative assessments can also serve purposes external to the classroom. According to Black and Wiliam, "Such assessment becomes *formative assessment* when the evidence is actually used to adapt the teaching to meet student needs" (pp. 1-2).

The use of formative assessment reflects the behaviorist (stimulus-response) theory articulated by Skinner (1954). Building on behaviorist theory, Bloom (1976) proposed a theory of schooling formulated on three constructs: (1) student characteristics, (2) quality of instruction, and (3) learning outcomes. Student characteristics include cognitive entry behaviors and affective entry characteristics. Quality of instruction, incorporated in learning tasks, includes the extent to which the cues, practice, and reinforcement of the learning are appropriate to the needs of the learner. Three learning outcomes are defined: the level and type of achievement, the rate of learning, and affective outcomes.

According to Bloom, when student entry characteristics are favorable and the quality of instruction is optimal, learning outcomes will be high. Further, it is possible to modify both students' entry characteristics and the quality of instruction so that learning outcomes are high for 90 to 95% of the students.

Bloom set out to confirm his theory through a series of experiments. Testing students after each learning task, Bloom (1976) found that, if students who have not met the criteria for mastery are then presented with instructional content in a different way, they will not only possess the affective disposition for the next learning task but will also have the cognitive skills to learn the next task (mastery conditions). However, if students receive no feedback or instruction after an assessment, their performance continues to decrease on subsequent tasks

and their affective disposition toward learning also declines (nonmastery conditions). Further, the achievement gap between mastery and nonmastery students widens as instructional tasks continue to be given. This is because mastery students receive feedback and instruction, while nonmastery students do not. Students who receive such mastery treatment over three or more learning tasks score higher than the nonmastery students who receive no feedback after initial testing for each of the learning tasks.

Moreover, Bloom's research indicates that if students who initially did not master a task are provided with feedback and additional instruction, they quickly score better than the group of students provided with no feedback or additional instruction. This experiment was repeated using the same design in a variety of subjects: second language, imaginary science, matrix algebra, and elementary probability.

Bloom suggests that achievement within each learning task is attainable for all students if feedback and additional instruction are provided. This information counters claims that lack of achievement is a product of factors over which schools have no control (such as socioeconomic status). Bloom also posits other variables and shows their association with learning outcomes. These variables include "cues, participation of the learner in the learning activity, and reinforcement which the learner secures in relation to the learning" (1976, p. 115).

Based on his research, Bloom (1976) concludes that feedback and corrective procedures account for 25% of the variation in student learning. Although Bloom focuses on individual students instead of classrooms, schools, or school systems, his theory suggests that schools and classroom teachers who provide consistent feedback on instruction to lower performers could achieve better results.

Implemented in schools as mastery learning or outcomes-based education, Bloom's approach has been studied in many different settings. In all, Block, Efthim, and Burns (1989) reviewed 48 studies and conclude:

Mastery learning approaches seem to work comparatively well almost all of the time. That is, they typically produce effects that are greater than or equal to a non-mastery approach. . . . Mastery learning approaches have comparatively strong effects on general student achievement. Assuming normality of scores, a median effect size of .76 [from the 48 studies] means that the typical application should move the average 50th percentile student to about the 77th percentile in achievement. (p. 28)

Nearly 10 years later, Black and Wiliam (1998a) reported the results of an extensive survey of the research literature on formative assessment. Their criteria for inclusion in the review were that "quantitative evidence of learning gains was obtained for those involved and for a similar group not so involved" (p. 2). Out of 580 journal articles and book chapters, they found 250 that met their criteria and were also responsive to this question: Is there evidence that improving formative assessment raises standards?

In another analysis, Black and Wiliam (1998b) selected 23 studies from a research review published by Fuchs and Fuchs (1986) that concentrated on classroom assessment for children with mild handicaps and 20 more studies conducted after the review by Fuchs and Fuchs. All together, the studies investigated kindergarten through undergraduate students, several content areas, and several countries.

Studies examined by Black and Wiliam (1998b) showed that strengthening the practice of formative assessment produced significant and often substantial gains. The learning gains were measured using *effect size*, which is calculated by comparing the average improvement in the test scores of pupils involved in a treatment with the range of scores that are found for typical groups of pupils on the same test. The ratio of the former divided by the latter is known as the effect size. Typical effect sizes found in the formative assessment experiments were between 0.4 and 0.7. Another important finding was that formative assessment helps low achievers more than other students and so reduces the range of achievement (achievement gap) while raising achievement overall.

More recently, Wiliam, Lee, Harrison, and Black (2004) conducted a study of secondary school students whose teachers were asked to dedicate time to formative assessment and instruction. Twenty-four teachers (two math and two science teachers from each of six schools located in two school districts in England) incorporated formative assessment in their classrooms. Classes of teachers who participated in the study were matched with equivalent classes, and student performance was compared. Effect sizes were calculated for each pair of classrooms. The differences in achievement between the treatment groups and the comparison groups resulted in a median effect size of .27 and a mean effect size of .34. This study is noteworthy because the intervention was conducted with teachers, so the effects on students were indirect.

Guterman (2002) examined the use of metacognitive awareness guidance to provide feedback on formative assessments to see if such guidance is associated with a difference in the learner's level of performance and achievement and also the learner's chances of internalizing the guidance component. Guterman describes the use of self-talk, a process of developing instruction material that helps students and teachers perform activities. Guterman observed 300 students from four schools in Israel (two schools in upper-class neighborhoods and two schools from lower-middle-class neighborhoods). Students were randomly assigned by classrooms to one of three groups: control, placebo, and experimental. Three different versions of reading assessment tasks were produced:

- The control group was asked to read specific text and answer a set of questions that followed.
- In the experimental group, written metacognitive guidance preceded reading of the passage and helped students engage in some of the larger issues suggested in the passage.
- The placebo group was asked to read the content instruction and the text only. When students finished reading, they were asked to raise their hands to signal that they were ready to proceed with the reading assessment tasks. They were then given special instructions on how to proceed, similar to the experimental group.

All participants were given a standardized reading comprehension test developed by the Israeli Ministry of Education; grades in reading and writing were obtained from school files. A one-way analysis of variance performed on the reading ability scores demonstrated that random assignment resulted in three groups that prior to the intervention were essentially equal in reading ability. All students were also given the Metacognitive Strategy Index, a questionnaire that measured their awareness of metacognitive reading strategies. All reading assessments were graded in a standard way. The analysis demonstrated that following the intervention, “learners who were given written metacognitive awareness guidance built on their prior knowledge exhibited higher achievement on the task” (Guterman, 2002, p. 296).

Stecker and Fuchs (2000) used a Curriculum-Based Measurement (CBM) system to provide individual feedback on the results of curriculum-embedded assessments to a group of 42 special education students in Grades 2 through 8. Each of the 42 students in the experimental group had a partner, and the two students received the same feedback rather than feedback based on their own results. Stecker and Fuchs found that “students whose teachers tailored instructional adjustment based on those students’ own CBM data performed significantly better on a global achievement test than did their partners whose instructional adjustments were not based on their own assessment data” (p. 128). The authors suggest that the study shows the importance of making instructional adjustments with the child’s assessment results in mind.

Rea-Dickins (2001) conducted a case study of assessment use by two language-support teachers and one mainstream class teacher in inner-city schools that had a high density of English language learners (up to 98%). Data collected included classroom observations and interviews with classroom teachers and language support teams. Rea-Dickins identified three different characteristics of classroom-based assessment: bureaucratic, pedagogic, and learning. The bureaucratic demands of assessment serve the obligation that schools must provide feedback to parents and external agencies. The pedagogic phase refers to instructional decisions based on group and individual learners’ performance. The learning phase is embedded within instruction. Rather than measuring learning, assessment helps the learner monitor his or her own learning. This study demonstrates that classroom assessment practice is related to determining what the learner knows and the amount of progress made. This determination then helps to guide teachers’ decisions about what and how to teach the next lesson.

Brookhart and Durkin (2003) collected data from 96 students in 12 social studies classrooms in an urban high school, including pre- and post-survey results, anecdotal records from students’ assessments, interviews, and classroom observations. In addition, the researchers interviewed two students per class, one low achiever and one high achiever. Brookhart and Durkin found that classroom assessment appears to be associated with three types of student motivation: (1) wanting to learn for learning’s sake, (2) wanting to show what they learned, and (3) wanting to help others learn or to learn from others. The researchers conclude that students want to learn more when the content of the assessment is connected with “a conscious, expressed interest” (p. 43).

While experimental research indicates that appropriately designed and implemented mastery learning models and other types of formative assessment can increase student achievement, formative assessment has proven difficult to implement in practice because students are no longer progressing in unison (Jones & Spady, 1985). The use of computer-based formative assessment may hold promise for overcoming student instructional grouping and scheduling difficulties. Ediger (2000) asserts that because computer packages involve text with read, respond, and check sequences, students take small steps in responding to multiple-choice test items that cover the content of the assignment and get instant feedback without teacher involvement. The computer software program keeps track of the correct responses and diagnoses errors within each sequential step of learning so the teacher also receives immediate feedback from the software program.

Chudowsky & Pellegrino (2003) suggest that the use of technology may help remove some of the constraints associated with current large-scale assessment practice. Computer-based assessments have the capability to assess problem-solving skills and also to identify the sequence of actions taken by the learner in solving the problem. Technology holds the promise for creating assessments that indicate how students think and reason while they are engaged in important learning activities. This information provides teachers with a profile of student learning. In addition, computer-based assessment information can be collected during the normal instructional process rather than interrupting the learning process to administer external tests.

In order for technology to become an effective tool for assessment, many issues need to be addressed. For example, McFarlane, Williams, and Bonnet (2000) point out that a framework for assessing the quality of the content of the technological tools and the procedure for learning needs to be established. Media genres, learner progression, and teaching goals and approaches must also be addressed. There should be a common terminology for addressing multimedia. If used properly, technology is powerful as a diagnostic tool. It can be used to support judgments about student performance and to conduct self-assessment of strengths and needs. To get an up-to-date overview of the issue, see reports from IAETE national conferences on technology for assessment at <http://www.edvantia.org/corporate/index1.cfm?§ion=about&area=iaete>.

Peat and Franklin (2002) surveyed 700 first-year biology students at the University of Sydney to determine students' perceptions of computer-delivered quizzes. A shortage of qualified faculty and increased enrollment at the university created a need to find new ways to improve student learning. One strategy was computer-based learning. Students took weekly quizzes, mock examinations, and special self-assessment modules using the computer. The students who used the computer program reported that it helped them to sort out details in content areas where neither the textbook nor the professor were very useful. Most of them had used the computer as a learning tool and self-assessment before the final exam. Peat and Franklin found evidence that student learning outcomes improved with the use of computer-based assessments. This increase in achievement appears to be associated with improved student attitudes toward learning. Students who used the computer program reported that they were encouraged by the instant feedback received. While the preliminary

findings are promising, there is a need for experimental studies to test the effectiveness of computer-based assessment.

Summary

Classroom assessment is used for many purposes. A substantial amount of scientifically based research has shown that classroom assessment can help to improve student achievement if the results are used to plan subsequent instructional activities. Students who don't initially learn a concept will need feedback, more time, and additional instruction. If these students do not receive appropriate feedback and additional instruction, they will fall further and further behind. As promising as the use of formative classroom assessment is, it is difficult for teachers to manage the differences in individual student progress. Computer-assisted instruction and assessment show promise for helping classroom teachers overcome these management issues.

References

- Black, P. J., & Wiliam, D. (1998a). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2): 139-148. Retrieved October 5, 2005, from <http://www.pdkintl.org/kappan/kbla9810.htm>
- Black, P. J., & Wiliam, D. (1998b). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-73.
- Block, J. H., Eftim, H. E., & Burns, R. B. (1989), *Building effective mastery learning schools*. New York: Longman.
- Bloom, B. S. (1976). *Human characteristics and school learning*. New York: McGraw-Hill.
- Brookhart, S. M., & Durkin D. T. (2003). Classroom assessment, student motivation, and achievement in high school social studies classes. *Applied Measurement in Education*, 16(1), 27-54.
- Chudowsky, N., & Pellegrino, J. W. (2003). Large-scale assessments that support learning: What will it take? *Theory into Practice*, 42(1), 75-83.
- Ediger, M. (2000, December). Purposes in learner assessment. *Journal of Instructional Psychology*, 27(4), 244.
- Fuchs, L. S., & Fuchs, D. (1986). Effects of systematic formative evaluation: A meta-analysis. *Exceptional Children*, 53(3), 199-208.
- Guterman, E. (2002). Toward dynamic assessment of reading: Applying metacognitive awareness guidance to reading assessment tasks. *Journal of Research in Reading*, 25(3), 283-298.

- Jones, L. V. (2001). Assessing achievement versus high-stakes testing: A crucial contrast. *Educational Assessment, 7*(1), 21-28.
- Jones, B. F., & Spady, W. G. (1985). Enhanced mastery learning and quality instruction. In D. U. Levine (Ed.), *Improving student achievement through mastery learning programs* (pp. 11-43). San Francisco: Jossey-Bass.
- McFarlane, A. E., Williams, J. M., & Bonnet, M. R. (2000). Assessment and multimedia authoring: A tool for externalizing understanding. *Journal of Computer Assisted Learning, 16*(3), 201-212.
- No Child Left Behind Act of 2001, Pub. L. No. 107-110. (2002).
- Peat, M., & Franklin, S. (2002). Supporting student learning: The use of computer-based formative assessment modules. *British Journal of Educational Technology, 33*(5), 515-526.
- Rea-Dickins, P. (2001). Mirror, mirror on the wall: Identifying processes of classroom assessment. *Language Testing, 18*(4), 429-462.
- Skinner, B. F. (1954). The science of learning and the art of teaching. *Harvard Educational Review, 24*(2), 86-97.
- Stecker, P. M., & Fuchs, L. S. (2000). Effecting superior achievement using curriculum-based measurement: The importance of individual progress monitoring. *Learning Disabilities Research & Practice, 15*(3), 128-135.
- Tyler, R. W. (1966). The objectives and plans for a national assessment of educational progress. *Journal of Educational Measurement, 3*(1), 1-4.
- William, D., Lee, C., Harrison, C., & Black, P. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education, 11*(1), 49-65.