

Guide Book

You can view this guide book
on your screen or print a copy.

▶ UCLA Center for the
Study of Evaluation

in collaboration with:

- ▶ University of Colorado
- ▶ NORC, University
of Chicago
- ▶ LRDC, University
of Pittsburgh
- ▶ The RAND
Corporation

Guidebook
for the Video Program

ASSESSING THE WHOLE CHILD

Narrated by Patrick Stewart

National Center for Research on Evaluation,
Standards, and Student Testing

Guidebook for the Video Program Assessing the Whole Child

Table of Contents

	<i>About Our Centers</i>	<i>iii</i>
Part I:	<i>Introduction to Performance Assessment.....</i>	<i>1</i>
Part II:	<i>Getting Started</i>	<i>6</i>
Part III:	<i>Criteria and Scoring for Performance Assessment.....</i>	<i>9</i>
Part IV:	<i>Involving Parents in Performance Assessment.....</i>	<i>12</i>
Part V:	<i>Am I a Teacher Researcher?.....</i>	<i>14</i>
Part VI:	<i>Resources for Teachers, Schools and Districts.....</i>	<i>15</i>
Part VII:	<i>Glossary of Useful Assessment Terms.....</i>	<i>19</i>

About Our Centers...

UCLA Center for the Study of Evaluation & National Center for Research on Evaluation, Standards, and Student Testing (CRESST)

For over 28 years, the UCLA Center for the Study of Evaluation (CSE) has been at the forefront of efforts to improve the quality of education in America through systematic evaluation practices. CSE has helped pioneer valid and sensitive evaluation and assessment techniques and has promoted vigorously the use of evaluation for more reasoned decision making.

Located within the UCLA Graduate School of Education & Information Studies, CSE is a unique center devoted to educational research, development, training, and dissemination. The result of a national competition, CSE was designated the national center for research in educational evaluation in 1966. This charge was renewed in 1990 when CSE successfully competed for the National Center for Research on Evaluation, Standards, and Student Testing (CRESST), receiving a five-year, \$14 million grant from the Office of Educational Research and Improvement (OERI).

CRESST Initiatives

CRESST projects are conducted in collaboration with colleagues from the University of Colorado at Boulder, University of California, Santa Barbara, National Opinion Research Center at the University of Chicago, Learning Research and Development Center at the University of Pittsburgh, the University of Southern California, and The RAND Corporation.

CRESST Program One, *Building the Infrastructure for Improved Assessment*, seeks to assure that the R&D efforts of CRESST and the many other players currently working on assessment reform will have optimal impact on educational policy and practice. Program Two, *Designs for Learning-Based Assessments: Prototypes and Models*, is creating new prototypes for assessing student performance and new models for analyzing and validating assessment results.

Program Three, *Collaborative Development and Improvement of Assessments in Practice*, also addresses the creation of alternative assessments. However, this developmental work is conducted collaboratively with state and/or local constituencies, subject to their time and resource constraints, and targeted to their goals and objectives.

CSE/CRESST Dissemination

CSE/CRESST produces a variety of dissemination materials. A quarterly newsletter, *CRESST Line*, contains up-to-date news on current CRESST assessment research. *Evaluation Comment*, a semiannual research-based publication, and *CRESST Line* are distributed free to a mailing list of over 11,000 education policy makers, researchers, and practitioners throughout the United States and overseas. For more information about CSE or CRESST, please contact:

Ronald Dietel, Director of Communications
TEL (310) 206-1532, FAX (310) 825-3883
CSE/CRESST
UCLA Graduate School of Education &
Information Studies
405 Hilgard Avenue, 1320 Moore Hall
Los Angeles, CA 90024-1522

The work reported herein was supported under the Educational Research and Development Center Program cooperative agreement number R117G10027 and CFDA catalog number 84.117G as administered by the Office of Educational Research and Improvement, U.S. Department of Education.

The findings and opinions expressed in any CRESST publication or product do not reflect the position or policies of the Office of Educational Research and Improvement or the U.S. Department of Education.

Part I: *Introduction to Performance Assessment*

What Is Your Purpose in Assessing a Child?

Knowing the *purpose of assessment* means understanding the various groups who have diverse interests in assessment results.

For *classroom teachers* like Charlotte Higuchi in *Assessing the Whole Child*, a primary purpose of assessment may be to evaluate where a child is at the beginning of the school year, then establishing goals and activities that will lead to increased student achievement. During the school year, Charlotte also uses assessment to evaluate student progress and how well her children are learning. Yet a third purpose of classroom assessment is that the process itself provides insight to Charlotte on how students learn. As she knows more about student learning, both individual and group, she develops an underlying basis for her instruction.

Thus, assessment for Charlotte at the classroom level is primarily *formative*, useful in identifying strengths and weaknesses of students and instructional programs, then making systematic improvements.

For policy makers, taxpayers, administrators, and school planners, the purpose of assessment may be more *summative* in nature. Is this program doing what it is supposed to do? Is it worth the money we are spending? Should we continue this curriculum or select another one? These same groups may be interested in holding states, schools, school districts, and teachers accountable for student achievement.

For parents and students, the purpose of assessment may be to gauge individual progress, knowing if student performance meets a specific standard. Parents frequently use assessment results to gauge the progress of their school.

Key Point: Regardless of the groups involved, clearly defining the purposes of assessment will lead to better assessment development and use of assessment results. Assessment **purpose, purpose,** and **more purpose** is a theme that underpins all of this guidebook.

What Is Performance Assessment?

Just what is performance assessment? Some have said that performance assessment is anything that requires students to “do” something as opposed to filling in a bubble on an answer sheet. For others, such as Charlotte, performance assessment is an information gathering *process* that leads to improved teaching:

Performance assessments allow a teacher to hone in on what a child’s strengths are and what a child needs to learn. . . . I can tailor the assessments to that child, not the other way around.

In their book, *A Practical Guide to Alternative Assessment*, CRESST authors Joan Herman, Pamela Aschbacher and Lynn Winters provide the following definition of performance assessment:

. . . performance assessments . . . require students to generate rather than choose a response. Performance assessment by any name requires students to actively accomplish complex and significant tasks, while bringing to bear prior knowledge, recent learning, and relevant skills to solve realistic or authentic problems. Exhibitions, investigations, demonstrations, written or oral responses, journals, and portfolios are examples of the assessment alternatives we think of when we use the term “alternative assessment.” (p. 2)

Certainly there are other possible definitions of performance assessment, but this one seems to well embody the concepts frequently discussed, that is, requiring students to perform an authentic task and assimilate prior knowledge and skills with new learning.

What Makes a Good Assessment?

Anyone who has ever tried to create an assessment quickly discovers the challenges inherent to the task. Does this assessment cover *important* content? Have I covered the full *range* of content? What is the *cost* in terms of instructional time? Are the questions *biased* towards a particular culture? Will the results *accurately* represent students’ capabilities?

A good assessment should respond to all of these questions and more. To help teachers and assessment developers answer the question “What makes a good assessment?” CRESST researchers developed a set of criteria for judging the quality of an assessment. On the following pages we discuss *selected* key criteria that we believe are of special interest to teachers.

1. What are the *consequences* of using this assessment?

Use of any assessment often results in teachers spending more time teaching what the assessment covers. As a consequence, important content that would otherwise have been taught may be excluded. For example, CRESST research showed that when Vermont required teachers to use portfolios in math and writing, teachers quickly focused on portfolios in these two subjects and, consequently, spent less time teaching other important subjects such as social studies. Ask yourself, “What are the possible consequences that may result from the use of this assessment, both positive and negative?”

2. Is this assessment *fair*?

Performance assessment is supposed to solve some fairness problems by making assessment and instruction nearly the same type of meaningful activity. But performance assessments cannot guarantee that all children have had an *opportunity to learn* (OTL) the material on which they’re assessed. Equal access to computers, calculators, and a well-trained teaching staff are a few elements that should be considered by teachers, schools, districts, and states when using performance assessments. Also, because the scoring of performance assessments is highly dependent on personal judgment, raters must be monitored for potential bias as should the assessment items themselves. Ask yourself, “How can I make sure that this assessment is fair, and how can I provide my students an equal opportunity to learn the content being assessed?”

3. Does the assessment cover *important, high-quality content*?

Consider the following assessment question:

“Imagine that it is 1858 and you recently heard Abraham Lincoln and Stephen Douglas debating during their campaigns for the Senate seat representing Illinois. After the debates you return home, where your cousin asks you about some of the problems that are facing the nation at this time.

Write an essay in which you explain the most important ideas and issues your cousin should understand. Your essay should be based on two major sources: (1) the general concepts and specific facts you know about American history—especially what you know about the history of the Civil War; and (2) what you have recently learned from listening to the debates.”

(Students read a selection from the Lincoln/Douglas debates.)

Does this question really examine *important, high-quality* content?
Yes.

The Lincoln/Douglas debates were a crucial turning point in U.S. history. Preceding the Civil War by three years, these debates helped

to define key slavery issues between many northern and southern states. As part of the national standards in U.S. history, the excerpts from the Lincoln/Douglas debates used in this assessment clearly cover *important, high-quality* content.

When developing your own assessment tasks, ask yourself, “Is this content important? Is it part of a national or state standard? Is it something my students may need in the future?”

4. Is this assessment *linguistically appropriate*?

The Lincoln/Douglas debate question, as well as many performance tasks, requires students to have excellent command of the English language. Students with other-than-English language backgrounds may have the knowledge, but not the language or writing capability to show their true level of new content understanding. Consider the *linguistic appropriateness* of these types of language-dependent tasks and how you might help students whose native language is other than English to perform well on the assessment.

5. Is this assessment *cognitively complex*?

Some skills, once acquired, become automatic, like the ability to drive a car. Other skills, such as linking the opening of a story to its ending (or *racing* a car) are more complex and will always require deeper thought and more reasoning. CRESST research conducted at the University of Pittsburgh suggests that good assessments should aim to assess “*cognitively complex*” skills. Simply stated, does this assessment make my child think? Memorizing facts is less important than having a repertoire of skills. Ask yourself if this performance task really challenges your students.

6. Is this assessment worth the *cost and time*?

As is obvious from *Assessing the Whole Child*, creating, administering and scoring performance assessments is a time-consuming process. As one CRESST researcher put it, “The key thing we have learned is that this whole [performance assessment] process takes time, time, and more time.” Time means money; consequently, the last of the CRESST criteria refers to the practicality and cost of performance assessments. Ask yourself, “What do I trade off by giving this assessment? Is the cost worth it?”

In answering this question, don’t forget that on the flip side of high costs are the positive benefits that seem to result from performance assessments. For example, principals have said that giving students these new assessments was the best staff development experience for teachers that they had ever seen. Integrating performance

assessments into existing staff development programs may be one way to keep the high costs within reason.

7. **Will student performance on this assessment *transfer* to other settings?**

We want our children to be able to apply the skills they learn from one setting to another, similar setting. Indeed, such *transfer* of learning is, or should be, a principal goal of schooling. But CRESST researchers from the University of California, Santa Barbara, found that doing well on one performance task, a science experiment, did not necessarily mean that students would do well on another, similar experiment. This suggests that students do not necessarily transfer skills from one performance task to another. The problem may be that the tasks themselves were not really comparable or simply the fact that students were not used to performance assessments. Nevertheless, this finding suggests that teachers (and others) should be careful of drawing important conclusions based on just one or two performance tasks.

Key Point: Anyone who wants to use performance assessments should realize that adequate resources must be provided to teachers, schools and districts.

Not all-encompassing, the preceding CRESST criteria serve as a starting basis from which teachers and school administrators might evaluate the quality of their assessment alternatives.

Part II: *Getting Started*

Understanding the relationship between learning, instruction, standards and assessment is a good way to get started with performance assessment. Note in Figure 1 below the various links between instruction, learning and assessment, suggesting that learning is not linear, but takes many directions at once. Standards may be adapted by the teacher from national, state, or district levels, or teachers may develop their own standards as we will soon discuss.

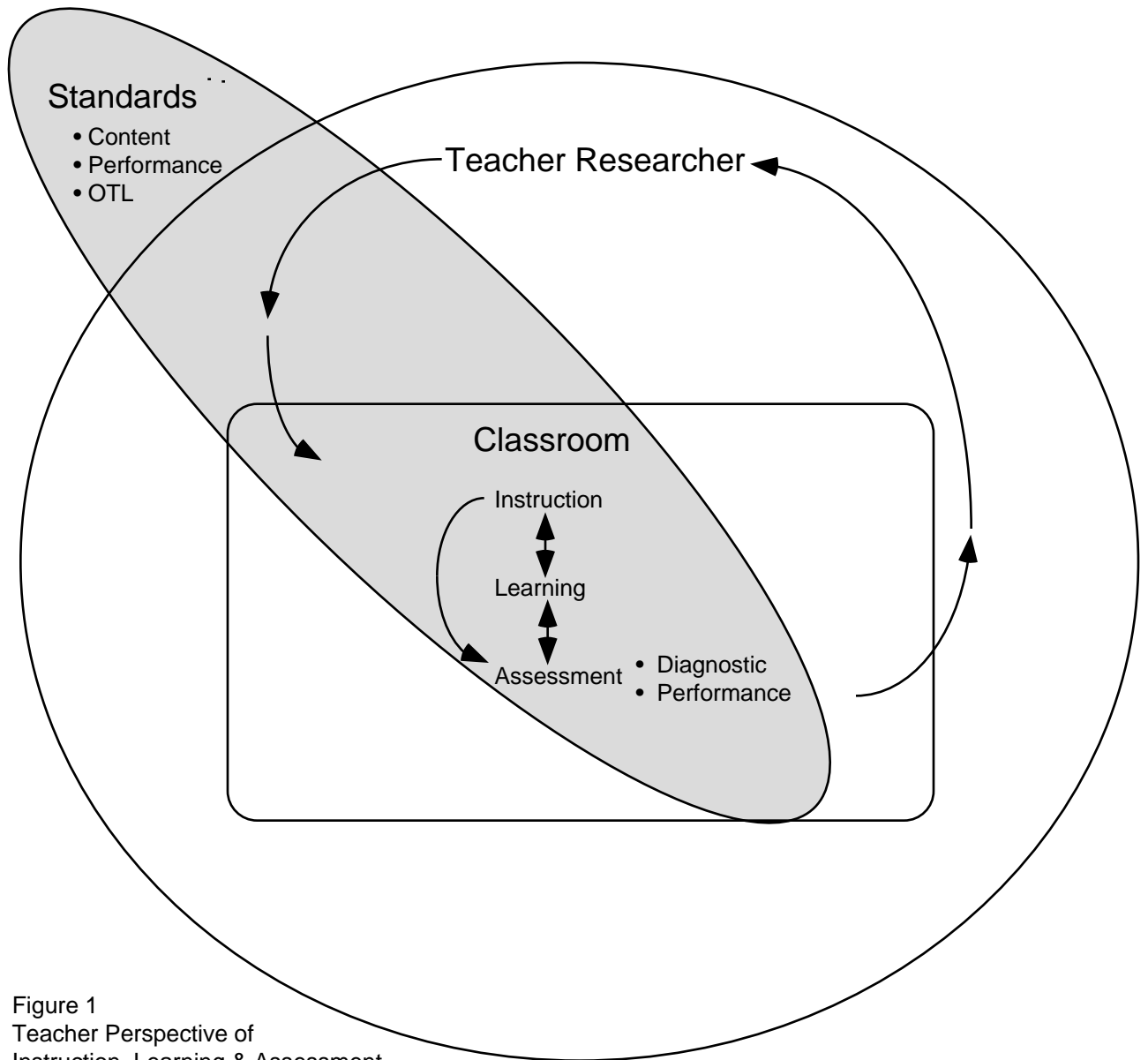


Figure 1
Teacher Perspective of
Instruction, Learning & Assessment

In Figure 1, diagnostic assessments, which may be selected response or performance based, standardized or not, provide information to the teacher on where a student is at a given point in time. Charlotte's use of a cloze assessment, for example, is diagnostic. Other performance assessments may provide the same type of information, but may also be used to determine student performance levels compared to the standards. Well-designed and implemented performance assessments are usually good instructional and learning activities in themselves.

Start With Standards

Goals 2000 and the new Elementary and Secondary Education Act provide financial incentives for states and school districts to develop and implement high standards in thousands of American classrooms. Additionally, the federal government has encouraged the development of national standards in a wide variety of subject areas, from mathematics to vocational education.

But how do teachers adapt these standards? In *Assessing the Whole Child*, Charlotte uses a process where students themselves develop standards for their own performances. By participating in the standard-setting process, students not only support the standards, but develop skill in evaluating their own work as well as that of other students.

There are certainly alternate ways to develop standards. Below and on the following page, we outline some strategies for getting started at the school level.

Start With Standards

1. Develop a specific strategy for the creation of standards in each subject area. Ask yourselves some basic questions: Where are our greatest strengths in subject matter expertise? Where are we weak?

We recommend that you start in your strongest areas first. Once you have created standards in one area, the next ones will become successively easier. Remember that implementation of standards depends on solid understandings of subject matter.

2. Link your efforts to your local goals and resources. Ask:
 - What are our goals for school improvement in each subject area?
 - What topics and concepts need to be covered in what grade levels?

- What are the school resources available, and what grants might be possible?
 - What is a reasonable time frame? Is there a professional development person available to assist us?
3. Build support: Who is available and willing to work in which areas? If you are a teacher, enlist support from your principal, parents, and any stakeholders who have an interest in your efforts. If you are a principal, you will need at least a good percentage of your teachers who are willing to work through the process.
 4. Discuss students' work. A useful beginning point at your first meetings is to review and analyze samples of students' work. Not only will everyone gain an appreciation for what others are doing, but consensus will be built around what kids *should* be doing. Contribute actively to the discussion, and make constructive comments.
 5. Build a library of resources in each subject area and make them available to all teachers at the school. You'll find our own suggestions in the Resources section to this guidebook.

There is no single recipe for successful standards development. But there are some guidelines. Be consistent, be willing to revise. No one gets it right the first time. Assessment is a way of thinking; you are assessing all the time. A teacher researcher needs to be able to step back and look at the whole picture of learning, not just instruction, standards or assessment as separate entities.

Part III: *Criteria and Scoring for Performance Assessment*

What Do You Mean—"Criteria That Work"?

... if we were all to self-evaluate ourselves on our jobs, what criteria would we use? How would we convince our boss that we did a great job? How would we pick our best pieces of work? *Charlotte Higuchi*

Charlotte's statement emphasizes that criteria used for judging performance, either student or adult, lie at the heart of all assessment.

Key Point: Without criteria, a task is just activity. It becomes an assessment when we apply criteria and standards to judge the quality of performance.

Consider a time as a student yourself when you received a grade on a paper accompanied by zero comments from the teacher (or professor!). Naturally, you wondered what criteria the teacher used to give that particular grade. Was it penmanship and spelling, or was it your ability to integrate prior knowledge with current instruction and then draw valid conclusions? With zero comments from the instructor and no knowledge of how you were assessed, you have little idea of your strengths or weaknesses or how to do better. Certainly, you would have learned much more by knowing what the criteria were and how you performed in each of those areas.

Research shows that effective scoring criteria:

- Clarify instructional goals for parents, teachers and, most importantly, students;
- Enhance fairness because students know exactly how they will be assessed and what learning outcomes are most important;
- Help teachers or other raters to be accurate, unbiased, and consistent in scoring;
- Are essential in the development of assessment tasks and clarify important content to be covered.

Effective criteria usually require three key elements say Joan Herman, Pamela Aschbacher and Lynn Winters in *A Practical Guide to Alternative Assessment*. These include:

- A description of the **dimensions** for judging student performance;
- A **scale** of values for rating those dimensions; and, when appropriate,
- The **standards** for judging performance.

The following example demonstrates **dimensions, scales** and **standards** in a basic scoring rubric for a complex performance task in social studies. The **dimensions** are (1) General Impression of Content Quality, (2) Prior Knowledge, (3) Number of Principles or Concepts, (4) Argumentation, (5) Proportion of Text Detail, and (6) Misconceptions.

CRESST Essay Scoring Rubric
from
Performance Assessment Models: Assessing Content Area Explanations
Eva L. Baker, Pamela Aschbacher, David Niemi, & Edynn Sato

1. General Impression of Content Quality (GICQ)
How well does the student know and understand this historical content? (0–5 point global rating: 0 = no response, 5 = highest level of understanding)

2. Prior Knowledge: Facts and Events (PK)
0 - no response
1 - no facts/events mentioned that are not found in the text of the debates
2 - one to two pieces of information that are not found in the text of the debates
3 - three to four pieces of information that are not found in the text of the debates
4 - five to six pieces of information that are not found in the text of the debates
5 - seven or more pieces of information that are not found in the text of the debates

3. Number of Principles or Concepts (PN)
0 - no response
1 - no principles/concepts
2 - one principle/concept
3 - two principles/concepts
4 - three principles/concepts
5 - four or more principles/concepts

4. Argumentation (A)

How well does the student organize historical knowledge to make a convincing argument? (0–5 point global rating: 0 = no response, 5 = highest level of coherent and cohesive argumentation)

5. Proportion of Text Detail (TEXT)

- 0 - no response
- 1 - no information from text
- 2 - material from the text accounts for about 1/4 of the essay
- 3 - material from the text accounts for about 1/2 of the essay
- 4 - material from the text accounts for about 3/4 of the essay
- 5 - the essay uses or is based on material from the text only

6. Misconceptions (MIS)

- 0 - no response
- 1 - one or more serious misconceptions central to the essay
- 2 - at least one serious misconception
- 3 - several minor errors and/or a moderate misconception
- 4 - very minor misconception
- 5 - no misconceptions

In the Performance Assessment Models handbook (see resource list on page 16), each **dimension** is clearly defined. For example, **Argumentation** is defined as:

. . . a measure of how well the student organizes knowledge to present a convincing argument or a coherent explanation.

Scoring of each dimension is accomplished on a **scale** of 0–5, but could alternatively be done through a written description.

Note that values on each scale are clearly defined. For example, in order for a student to receive a “5” under Misconceptions, the performance requirement is *no misconceptions*.

Performance standards are usually left to the user to decide and essentially answer the question “How good is good enough?” You might decide, for example, that for the dimension Misconceptions, a value of “3,” that is, *several minor errors and/or a moderate misconception*, is acceptable for your students. “3” becomes your performance standard for this dimension.

Again, the Performance Assessment Models handbook contains further clarification for each of the dimensions and the related standards.

Part IV: *Involving Parents in Performance Assessment*

In *Assessing the Whole Child*, Charlotte included her students in her parent conferences.

“Students should be included,” says Charlotte, “because the conference is, after all, about them!”

In her meetings with parents and students, Charlotte carefully describes her assessment of the whole child, including student results on standardized and performance assessments, student strengths, needed areas of improvement, and especially what parents can do to help the child at home. Working primarily with disadvantaged Latino and Asian students in East Los Angeles, Charlotte’s research clearly points to the important role that parents play in the education of their children.

During her parent-student conference, Charlotte explains her instructional methods, thereby enlisting and receiving the support of parents who are probably unfamiliar with new types of assessment or instruction. Showing samples of student work is an important indicator to parents of what students are able to do, not just how well they perform on an individual assessment. Additionally, parent conferences enable Charlotte to evaluate the type of support a student is receiving at home, which in turn, enables her to assess the child’s opportunity to learn.

Parent conferences should be adapted to individual situations, but we believe the following might be a minimum point of discussion for each student.

Tips for Parent/Student/Teacher Conferences

- **Involve** the parent and student early in your discussion. Have each of them discuss what they think are the student’s strengths and weaknesses.
- Discuss **multiple** indicators of student performance including combinations of diagnostic, standardized, and performance assessments, formal and informal observations. No single assessment, grade, or performance can tell teachers or parents everything they need to know about a child. When possible, parents and students should be provided comparative information that enables them to know how other students are doing either on a local, statewide or a national basis.

- **Discuss student social skills** including the student's ability to work cooperatively and to assume leadership roles when appropriate. Teachers may wish to create a set of criteria and a checklist that assists them in assessing student social skills.
- **Show and discuss examples** of student work such as might be contained in a portfolio. Work should demonstrate a wide range of problem-solving skills as opposed to one or two single best pieces. Work that shows needed areas of improvement and comparative information with other students will be useful for both the student and parent to work towards improvement. As much as possible, the work should be the student's own rather than group work or work where parents or teachers have had considerable influence.

Part V: *Am I a Teacher Researcher?*

If you are the type of teacher who asks lots of specific questions about how your students learn in the classroom and then takes specific, thoughtful actions to increase their learning, then you are probably already a teacher researcher. Researchers, in general, conduct disciplined inquiry into a set (or sets) of questions for which answers are not already known.

In *Assessing the Whole Child*, Charlotte uses videotape to document her instruction and analytically evaluate student performance as well her own teaching methods. This method is just one way of conducting teacher research. While most teachers may not have the same access to resources as Charlotte, they can still gain a great deal of knowledge from reading current literature related to learning and assessment. An appendix of valuable resources is provided at the end of this guidebook.

Perhaps the single best suggestion we can make for teachers who would like to learn more about how children learn is to meet on a regular basis with other teachers. Teachers should share with one another examples of student work, discussing strengths and weaknesses and what represents good performance. It has been our experience that the productivity of such meetings is likely to increase significantly if an outside, experienced facilitator is available to keep the discussion on track towards clearly defined goals.

For school administrators who would like to create an atmosphere rich in standards, learning, and assessment, we urge you to provide teachers with adequate planning time. CRESST research in Vermont, California, and Colorado has demonstrated that teachers need time, time, and more time, but that such time needs direction if it is to result in significant program improvement. For newer teachers and even those with many years of experience, building confidence in subject matter and learning how children learn is an essential step towards improved instruction and assessment.

Final Key Point: A good next step is to meet with other teachers and begin a standards-setting process like the one described on pages 7–8 of this guidebook. Once again, clearly understand the use of the assessment results, in other words, your **purpose** of assessment!

Part VI: *Resources for Teachers, Schools and Districts*

Many of the following resources are available from CRESST. Call Kim Hurst at (310) 206-1532 or send her an e-mail message at "kim@cse.ucla.edu" for more information. Or write to CRESST at CRESST/UCLA, 10880 Wilshire Boulevard, Suite 700, Los Angeles, CA 90024-4108.

You may also find that your local school district, county or state department of education, or certain nonprofit educational support groups can provide expertise and assistance. Additionally, the American Federation of Teachers and National Education Association have resources available. Contact Eugenia Kemble at the American Federation of Teachers, Educational Issues Department, 555 New Jersey Avenue, NW, Washington, DC 20001, (800) 238-1133; or Rod Riffel at the National Education Association, Educational Policy and Professional Practice, 1201 16th Street, NW, Washington, DC 20036, (202) 822-7384.

<h3>Selected Books on Assessment</h3>
--

A Practical Guide to Alternative Assessment

Herman, J.L., Aschbacher, P.R., & Winters, L. (1992). Washington, DC: Association for Supervision and Curriculum Development.

One of the most widely used books on performance assessment. Contains chapters on rethinking assessment, linking assessment and instruction, determining assessment purpose, selecting assessment tasks, setting criteria, ensuring reliable scoring, and using alternative assessment for decision making.

Modern Educational Measurement: A Practitioner's Perspective

Popham, J.W. (1990). 2nd ed. Englewood Cliffs, NJ: Prentice Hall.

A valuable reference for teachers who would like to learn more about assessment in general, including standardized assessments.

Assessing Student Performance: Exploring the Purpose and Limits of Assessment

Wiggins, G.P. (1993). Jossey-Bass education series, Vol. 1. San Francisco, CA: Jossey-Bass.

In Teachers' Hands: Investigating the Practices of Classroom Assessment

Stiggins, R.J., & Conklin, N.F. (1992). Albany: State University of New York Press.

Reports and Handbooks

CRESST Performance Assessment Models: Assessing Content Area Explanations

Baker, E.L., Aschbacher, P.R., Niemi, D., & Sato, E. (1992). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

A concise model for developing alternative assessments. Contains useful examples of CRESST assessment materials, an effective scoring rubric applicable to multiple topics, and benchmark papers.

Performance-Based Assessment and What Teachers Need

Higuchi, C. (1993). (CSE Tech. Rep. No. 363). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Early attempts to quickly implement performance assessment oftentimes failed to provide teachers with appropriate professional development opportunities. As an early user of performance assessments, Charlotte Higuchi documents in this report many of the needs she had in her own implementation efforts.

Complex, Performance-Based Assessment: Expectations and Validation Criteria

Linn, R.L., Baker, E.L., & Dunbar, S.B. (1991). (CSE Tech. Rep. No. 331). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

One of the first major research articles that established validity criteria for performance-based assessment. Professors Baker and Linn are the CRESST co-directors, and Professor Dunbar teaches at the University of Iowa.

Writing What You Read: A Guidebook for the Assessment of Children's Narratives

Wolf, S., & Gearhart, M. (1994). (CSE Resource Paper No. 10). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Designed to help teachers think about the important role of assessment in guiding students' narrative writing, this guidebook should help teachers draw close connections between curriculum, instruction and assessment. Provides specific examples for the scoring of student writing, including a narrative feedback form and narrative rubric to help teachers evaluate students' understandings and ability to learn.

Other Useful Publications and Articles

Assessment: Authenticity, Context, and Validity

Wiggins, G. (1993, November). *Phi Delta Kappan*, 75(3), 200-214.

Portfolio News

Cooper, W. (Ed.). (1994, Winter). [Entire Issue]. *Portfolio News*, 5(2).

A good resource for portfolios. Contains thoughtful articles on both theory and implementation. Also documents what is happening in some major portfolio programs across the United States.

The ERIC Review: Special Issue on Performance-Based Assessment

Boston, C.L. (Ed.). (1994, Winter). [Special Issue]. *The ERIC Review*, 3(1).

A useful overview of performance assessment and a guide to valuable resources and organizations.

Using Performance Assessment

Brandt, R.S. (Ed.). (1992, May). *Educational Leadership*, 49(8).

An entire issue dedicated to performance assessment. Includes an article by CRESST Associate Director Joan Herman, "What Research Tells Us About Good Assessment."

Databases/Tool Kits

Alternative Assessments in Practice (AAIP) Database

Containing listings from over 250 developers of new assessments, the *Alternative Assessments in Practice* database is of special use to teachers, school district administrators, assessment developers and others interested in new methods for assessing student growth. Released in an easy-to-use Macintosh HyperCard format, the database contains detailed information about each assessment, including subject matter and skills measured, assessment type and purpose, scoring characteristics, and availability of the assessment. The AAIP database, including a 50-page user's manual, is available through the CRESST offices for \$15.00.

Facilitating Systemic Change in Science and Mathematics Education: A Toolkit for Professional Developers (In Development)

This upcoming resource toolkit for teachers will contain information on assessment and effective teaching practices. Contact The Regional Laboratory for Educational Improvement of the Northeast and Islands, 300 Brickstone Square, Suite 950, Andover, MA 01810, (508) 470-0098.

<p style="text-align: center;">On the Information Highway</p>
--

CRESST Gopher Server

Available through the CRESST gopher server are articles from the *CRESST Line* newsletter, listings and descriptions of CRESST technical reports, and the CRESST *Alternative Assessments in Practice* database. Computer users with Internet access and a gopher client program may access the CRESST server by starting a new gopher and pointing at: **gopher.cse.ucla.edu**.

CRESST World Wide Web Server

The CRESST World Wide Web (WWW) server provides access to CRESST assessment research in an exciting visual format. Advanced text formatting, graphics, audio, and video make the CRESST World Wide Web server another valuable product for K-12 educators. Internet users with the appropriate software may connect to the CRESST World Wide Web at: **<http://www.cse.ucla.edu/>**.

From the Department of Education

Useful internet servers are maintained by the Department of Education. Connect to their gopher server at: **gopher.ed.gov**; or their World Wide Web server at: **<http://inet.ed.gov/>**.

Northwest Regional Educational Laboratory (NWREL)

This gopher server contains its own alternative assessment database and a wide variety of other useful education databases. The gopher address is: **gopher.nwrel.org**.

Part VII: *Glossary of Useful Assessment Terms*

Age norms—values representing typical or average performance of people of age groups.

Content standards—specify “what students should know and be able to do” in various subjects and domains, such as mathematics or applied learning. Set out the knowledge, skills and other necessary understandings that schools should teach in order for all students to attain high levels of competency in the subject matter.

Criterion-referenced assessment—a measurement of achievement of specific criteria or skills in terms of absolute levels of mastery. The focus is on performance of an individual as measured against a standard or criterion rather than against performance of others who take the same assessment, as with norm-referenced assessments.

Dimensions, traits, or subscales—the subcategories used in evaluating a performance or portfolio product; e.g., in evaluating students’ writing one might rate student performance on subscales such as organization, quality of content, mechanics and style.

Diagnostic assessment—an intensive, in-depth evaluation process with a relatively detailed and narrow coverage of a specific area. The purpose of this assessment is to determine the specific learning needs of individual students and to be able to meet those needs through regular or remedial classroom instruction.

Grade equivalent—the estimated grade level that corresponds to a given score.

Holistic scoring—Scoring based upon an overall impression (as opposed to traditional assessment scoring which counts up specific errors and subtracts points on the basis of them). In holistic scoring the rater matches his or her overall impression to the point scale to see how the portfolio product or performance should be scored. Raters usually are directed to pay attention to particular aspects of a performance in assigning the overall score.

Item—an individual question or exercise in an assessment or evaluative instrument.

Norm-referenced assessment—an objective assessment that is standardized on a group of individuals whose performance is evaluated in relation to the performance of others; contrasted with criterion-referenced assessment.

Opportunity-to-learn standards—suggest criteria for assessing whether schools are giving students the opportunity to learn material reflected in the content standards. Such criteria may include the availability of instructional materials or whether teachers are prepared to teach according to the content standards.

Percentile—the percent of people in the norming sample whose scores were below a given score.

Percent score—the percent of items that are answered correctly.

Performance assessment—performance assessments require students to generate rather than choose a response. Performance assessment by any name requires students to actively accomplish complex and significant tasks, while bringing to bear prior knowledge, recent learning, and relevant skills to solve realistic or authentic problems. Exhibitions, investigations, demonstrations, written or oral responses, journals, and portfolios are examples of the assessment alternatives we think of when we use the term alternative assessment.

Performance criteria—a predetermined list of observable standards used to rate performance assessments. Effective performance criteria include considerations for validity and reliability.

Performance standards—Answers the question “How good is good enough?” Establishes the degree or quality of students’ performance in the subject matter set out by the content standards. Performance standards are often illustrated by exemplars of student work and may contain gradations of performance, i.e., needs improvement, meets standard, above standard.

Portfolio—A collection of representative student work over a period of time. A portfolio often documents a student’s best work, and may include a variety of other kinds of process information, e.g., drafts of student work, students’ self-assessment of their work, parents’ assessments. Portfolios may be used for evaluation of a student’s abilities and improvement.

Prompt—an assignment or directions asking the student to undertake a task or series of tasks. A prompt presents the context of the situation, the problem or problems to be solved, and criteria or standards by which students will be evaluated.

Rating scales—a written list of performance criteria associated with a particular activity or product which an observer or rater uses to assess the pupil’s performance on each criterion in terms of its quality.

Raw score—the number of items that are answered correctly.

Reliability—the extent to which an assessment is dependable, stable, and consistent when administered to the same individuals on different occasions. Technically, this is a statistical term that defines the extent to which errors of measurement are absent from a measurement instrument.

Rubric—a set of guidelines for giving scores. A typical rubric states all the dimensions being assessed, contains a scale, and helps the rater place the given work properly on the scale.

Standardized assessment—a form of measurement that has been normed against a specific population. Standardization is obtained by administering the assessment to a given population and then calculating means, standard deviations, standardized scores, and percentiles. Equivalent scores are then produced for comparisons of an individual score to the norm group's performance.

Standard scores—a score that is expressed as a deviation from a population mean.

Stanine—one of the steps in a nine-point scale of standard scores.

Task—a goal-directed assessment activity, demanding that students use their background knowledge and skills in a continuous way to solve a complex problem or question.

Validity—the extent to which an assessment measures what it was intended to measure. Validity indicates the degree of accuracy of either predictions or inferences based upon an assessment score.