DOCUMENT RESUME

ED 365 548 SE 054 018

TITLE	Innovative Assessment. Science and Mathematics Bibliographies.
INSTITUTION	Northwest Regional Educational Lab., Portland, OR. Test Center.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE	93
CONTRACT	RP91002001
NOTE	85p.
AVAILABLE FROM	The Test Center, Evaluation and Assessment Program, Northwest Regional Educational Laboratory, 101 S.W. Main Street, Suite 500, Portland, OR 97204.
PUB TYPE	Reference Materials - Bibliographies (131)
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	Academic Achievement; Annotated Bibliographies; *Educational Assessment; Educational Innovation; Elementary Secondary Education; *Informal Assessment; *Mathematics Education; *Science Education; *Student Evaluation
IDENTIFIERS	*Alternative Assessment

ABSTRACT

This annotated bibliography represents Test Center holdings to date in the area of assessment alternatives in mathematics and science. Alternative assessment, for the purpose of this bibliography, means assessment other than standardized, norm-referenced assessment. The list emphasizes examples of assessment, such as performance assessments, portfolios, and technological innovations. The references are presented in two sections. The first section contains 108 documents on assessment alternatives in mathematics. The second section contains 75 documents on assessment alternative in science. (MDH)

Innovative Assessment

Science and Mathematics Bibliographies

Northwest Regional Educational Laboratory

THE TEST CENTER

Sponsored by OERI Office of Educational Research and Improvement U.S. Department of Education

Innovative Assessment

Science and Mathematics Bibliographies

1993

The Test Center
Evaluation and Assessment Program
Northwest Regional Educational Laboratory
101 S.W. Main Street, Suite 500
Portland, Oregon 97204

This publication is based on work sponsored wholly, or in part, by the Office of Educational Research and Improvement (OERI), Department of Education, under Contract Number RP91002001. The content of this publication does not necessarily reflect the views of OERI, the Department, or any other agency of the U.S. Government.

ASSESSMENT ALTERNATIVES IN MATHEMATICS

The following articles represent Test Center holdings to date in the area of assessment alternatives in mathematics. Presence on the list does not necessarily imply endorsement. Articles are included to stimulate thinking and provide ideas. Some of the entries are formal assessments, and are intended mainly for the classroom. For more information, contact Dr. Judy Arter, Unit Manager, or Matthew Whitaker, Test Center Clerk, at (503) 275-9582, Northwest Regional Educational Laboratory, 101 SW Main, Suite 500, Portland, Oregon 97204.

Algina, James, and Sue Legg (Eds.). *Special Issue: The National Assessment of Educational Progress.* Located in: Journal of Educational Measurement, 29, Summer 1992.

This special issue of JEM discusses the National Assessment of Educational Progress (NAEP)--history, specification of content and design of assessments for 1992 and beyond, how students are sampled, and how results are reported. Although some articles are somewhat technical, the general pieces on NAEP's history, and the design of current assessments will be interesting to the general readership.

The current plans for math include:

1. Use of calculators for about 70 percent of the test.

2. Estimation skills tasks using an audio tape.

3. Yes/No questions to determine the extent to which students understand the same information when it is presented in different forms.

4. Constructed response questions in which students are asked to document their solutions by drawing their answers, writing explanations, or providing their computations.

Scoring guides for open-ended questions are tailored to each question. Some examples are provided.

(TC# 150.6JEM292)

Appalachia Educational Laboratory. ***Alternative Assessments in Math and Science: Moving Toward a Moving Target,*** **1992. Available from: Appalachia Educational Laboratory, PO Box 1348, Charleston, WV 25325, (304) 347-0400.**

This document reports on a two-year study by the Virginia Education Association and the Appalachia Educational Laboratory. In the study, 11 pairs of K-12 science and math teachers designed and implemented new methods of evaluating student competence and application of knowledge.

Teachers who participated in the study found that the changes in assessment methods led to changes in their teaching methods, improvements in student learning and better student attitudes. Instruction became more integrated across subjects and shifted from being teacher-driven to being student-driven. Teachers acted more as facilitators of learning rather than dispensers of information.

Included in the report is a list of recommendations for implementing alternative assessments, a list of criteria for effective assessment, and 22 sample activities (with objectives, tasks, and scoring guidelines) for elementary, middle, and high school students, all designed and tested by the teachers in the study.

Most activities have performance criteria that are holistic and specific to each exercise. No technical information or sample student work is included.

(TC# 600.3ALTASM)

Bagley, Theresa, and Catarina Gallenberger. ***Assessing Students' Dispositions: Using Journals to Improve Students' Performance.*** **Located in: <u>The Mathematics Teacher</u>, 85, November 1992, pp. 660-663.**

In this article, the authors discuss the use of journals to elicit behavior that can be examined for high school students' attitude toward math, making mathematical connections, and understanding. They present many questions, tasks, and instructions for getting students to self-reflect, and provide good, practical suggestions for managing the process. However, the authors do not provide criteria for examining student responses (i.e., what to look for in responses that are indicators of attitude, connections or understanding), so the procedure is informal. The procedure will only be useful to the extent that users have the expertise to know what to look for in responses.

(TC# 500.6ASSSTD)

Barton, Paul E. ***National Standards for Education: What They Might Look Like; A Workbook,*** **1992. Available from: Educational Testing Service, Policy Information Center, Mail Stop 04-R, Princeton, NJ 08541, (609) 734-5694.**

This monograph presents examples of standards from eight different projects. The intent is to illustrate and document some existing standards, help policy makers sharpen their thinking about standards, and help people develop common concepts of standards. The eight samples come from NCTM Math Standards, Project 2061 in science, Advanced Placement US History, NAEP Science Objectives, Toronto Benchmarks in math and language arts, NAEP Geography Objectives, National Curriculum in England and Wales in math, and Florida Department of Education on general definitions of terms.

(TC# 500.5NATSTE)

Baxter, Gail P., Richard J. Shavelson, Sally J. Herman, Katharine A. Brown, and James R. Valadez. ***Mathematics Performance Assessment: Technical Quality and Diverse Student Impact.*** **Located in: Journal for Research in Mathematics Education, 1993, 24, 3, pp. 190-216.**

The authors developed 41 hands-on tasks to measure three categories of sixth-grade student competencies: measurement (seven tasks), place value (31 tasks), and probability (three tasks). An example of a measurement task is "describe the object" in which students had to write a description of an object that someone else could use to draw the object. Sixteen of the place value tasks were "card shark" in which students were dealt cards with four numbers (e.g., 6000, 100, 60 and 2). They had to put the cards together to form a specified number, read the number aloud, and name the place value of a particular digit. An example of a probability task was "spin it" in which students were given a spinner with eight sections (four orange, three yellow, and one green). They had to predict which color the pointer would land on most or least often, predict the outcome of 32 spins, and carry out the experiment and graph the results.

Responses were scored either by degree of "correctness" or, in the case of the communication items (e.g., describe an object), holistically for general quality of the response. The tasks and criteria were described only in general terms; further information would have to be obtained from the authors in order to actually reproduce the assessment.

Tasks were pilot tested with 40 sixth graders (Anglo and Hispanic) from two types of instructional settings: hands-on and traditional. Results showed: raters using this type of rating scheme can be trained to be very consistent in their scoring; the assessments are costly and time-consuming; a considerable number of tasks need to be administered to provide a reliable estimate of a student's level of achievement; student performances on the hands-on tasks differed by the type of instructional setting (evidence of validity); and there was differential scoring on the part of Hispanics, leading to some equity concerns.

(TC# 500.6MATPEA)

Braswell, James. *Overview of Changes in the SAT Mathematics Test in 1994. [SAT Mathematics--Student Produced Responses]*, 1991. **Available from: Educational Testing Service, Rosedale Rd., Princeton, NJ 08541, (609) 734-5686.**

This was a paper presented at the annual meeting of the National Council on Measurement in Education, April 5, 1991, Chicago.

Currently, the SAT-Math consists of two parts: regular multiple-choice and quantitative comparison (e.g., solution A is larger than, smaller than, or equal to, solution B, or cannot be determined). A third part called "student-produced responses" will be included on the PSAT in 1993 and the SAT in 1994. In this part, students will solve problems that have integer, fractional, or decimal solutions in the range 0 to 9999. A grid is provided for students to enter their actual answer. Some problems will have more than one right answer or can be any value in a range. For these problems, a correct response is recorded if the student answer is one of the accepted answers. Of the 55-60 items on the test, 10-15 will be in this format.

The materials include a couple of examples of this type of item.

(TC# 500.3SATMAS)

Brown, Larry. *Portfolios in Rural High School Mathematics and Science Classes*, **1992. Available from: Cusick High School, PO Box 270, Cusick, WA 99119, (509) 445-1125.**

This project is still in the developmental process, but is intended to develop the concept that the portfolio is a student's self-selected, self-reflective documentation of growth in understanding and skill over the course of a school year. Students will prepare their portfolios across the curriculum areas of advanced mathematics and physics. Results of the project will be presented together with recommendations for improvement and implications for future work to the Cusick School District, participants of SMART (NWREL), and at the Small Schools Conference at Central Washington University on March 19, 1993.

The author only provided a description of his project. Additional information is available only from the author.

(TC# 660.6PORRUH)

California State Department of Education. *A Question of Thinking: A First Look at Students' Performance on Open-Ended Questions in Mathematics*, **1989. Available from: California State Department of Education, PO Box 944272, Sacramento, CA 94244-2720, (916) 445-1260.**

This report describes the results of 12th grade student assessment using open-ended math problems that were part of the California Assessment Program (CAP). The open-ended problems were scored using rubrics developed for each problem. These rubrics are described, and "anchor" papers for the six scale values for each rubric are provided. Although there is a separate rubric for each problem, they are all intended to reflect the following dimensions of problem solving: understanding of mathematics, use of mathematical knowledge, and ability to communicate about mathematics.

(TC# 500.3AQUESO)

Campbell, Donna. ***Arizona Student Assessment Plan, (ASAP)*, 1990. Available from: Arizona Department of Education, 1535 W. Jefferson, Phoenix, AZ 85007, (602) 542-5393.**

The Arizona Assessment Program has several parts: a short standardized achievement test, non-test indicators, and performance assessments in reading, math and writing. The performance tests are designed to measure the state's Essential Skills. The math portion presents an extended problem-solving situation that requires short answers, extended answers, and explanations of answers. Each extended exercise has its own specific set of scoring procedures that involve assigning a point value if various things are present in the response.

(TC# 060.3ARISTA)

Carpenter, Thomas P., James Hiebert, Elizabeth Fennema, Karen Fuson, Alwyn Olivier, and Diana Wearne. ***A Framework for the Analysis of Teaching and Learning Understanding of Multidigit Numbers.*** **Information on date and availability is unknown.**

This paper presents a way to analyze instruction in math to see whether it is designed to foster understanding, defined as making relevant connections between knowledge. The specific example in the paper relates to multidigit numbers. Dimensions of instruction thought to be critical in promoting understanding include such things as: the scope and sequence of concepts, connections among representations as a basis for establishing meaning for symbols, the nature of problem solving, teacher specification of solution procedures and connections, students' articulation of solution procedures, and coherence between and within lessons.

Most of the paper describes each of these dimensions in detail. Several pages at the end discuss in general terms the kinds of tasks one could give to students to see whether they are making the appropriate connections.

(TC# 500.4FRAANT)

Center for Innovation in Education. ***Math Their Way*, 1990. Available from: Center for Innovation in Education, 19225 Vineyard Ln., Saratoga, CA 95070, (408) 867-3167.**

Math Their Way is an instructional program designed for grades K-2 that emphasizes manipulatives. Chapter 3 deals with assessment; the suggested assessment activities tie into the instructional program. These are suggested "formal assessments" to be used to track student progress two to four times a year. They are really not intended for daily use. There are 18 assessments to evaluate three areas--prenumber concepts and skills, number operations, and place value. All assessments are individual and performance based. No technical information is provided.

(TC# 070.3MATTHW)

Champagne, Audrey B. ***Cognitive Research on Thinking in Academic Science and Mathematics: Implications for Practice and Policy.*** **Located in: <u>Enhancing Thinking Skills in the Sciences and Mathematics</u>, Diane Halpern (Ed.), 1992. Available from: Lawrence Erlbaum Associates, Publisher, 365 Broadway, Hillsdale, NJ 07642, (800) 926-6579.**

Although this article is not strictly about assessment, it discusses some topics of relevance to assessment. Specifically, it has a very nice section on the relationship between the tasks given to students and what they can learn. For example, students can't learn as efficiently to integrate knowledge if they are never given tasks that require them to do this. This also has relevance to designing "authentic" tasks for performance assessments.

(TC# 000.6COGRET)

Charles, Randall. ***Evaluating Progress in Problem Solving*****, 1989. Located in: <u>Communicator</u>, 14, 2, pp. 4-6. Also available from: The California Mathematics Council, 1414 S. Wallis, Santa Maria, CA 93454, (805) 925-0774.**

This article presents a rationale for analyzing student open-ended problem solving in a systematic fashion. One sample analytical scoring rubric is presented. The traits are understanding the problem, planning a solution, and getting the answer. The author also proposes some other questions to ask as one looks at student problem solving: Did the student seem to understand the problem? Were the approaches used to solve the problem feasible for finding a solution? Does the answer make sense in terms of the question to be answered?

(TC# 500.3EVAPRI)

Charles, Randall, Frank Lester, and Phares O'Daffer. ***How to Evaluate Progress in Problem Solving*****, 1987. Available from: National Council of Teachers of Mathematics, 1906 Association Drive, Reston, VA 22091.**

This monograph attempts to assist educators with the challenge of developing new techniques for evaluating the effectiveness of instruction in problem solving by clarifying the goals of problem-solving instruction, and illustrating how various evaluation techniques can be used in practice. Goals include: select and use problem-solving strategies, develop helpful attitudes and beliefs, use related knowledge, monitor and evaluate thinking while solving problems, solve problems in cooperative learning situations, and find correct answers.

Evaluation strategies include: informal observation/questioning and recording results using anecdotal records or a checklist (two are provided); interviews (a sample interview plan is provided); student written or oral self-report of what's happening during a problem-solving experience (a list of stimulus questions is given, as is a checklist of strategies); attitude inventories (two are given); rating scales (three-trait analytic and focused holistic scales are given); and multiple-choice and completion (sample items are given to assess various problem solving abilities; many of these parallel question types mentioned by Marshall, above, to assess procedural and schematic knowledge).

Many sample problems are provided. No student sample performances or technical information is provided.

(TC# 500.6HOWTOE)

Clark, David. *The Mathematics Curriculum and Teaching Program*, 1988. Available from: Curriculum Development Centre, PO Box 34, Woden, ACT 2606, Australia. Also available from: ERIC ED 287 722.

This document was developed to assist classroom teachers to improve their day-to-day assessment of mathematics. Content includes: rationale for assessment alternatives in mathematics, instructions for a two-day in-service program using the materials, instructions on how classroom teachers can use the materials without training, and a series of exercises, formats and ideas for classroom assessment.

Assessment ideas include: help with systematically recording information from informal observations using checklists and "folios" of student work, setting up opportunities for assessment by giving students good tasks to do, assessing problem solving, student self-reflection, and communicating results.

This is written in a very user-friendly manner and contains some good ideas, especially in the areas of designing tasks, problem solving and self-reflection. We found some of the descriptions of activities a little too sketchy.

(TC# 500.3MCTPMA)

Coalition of Essential Schools. [*Various Articles on Exhibitions of Mastery and Setting Standards*], 1982-1992. Available from: Coalition of Essential Schools, Brown University, Box 1969, One Davol Sq., Providence, RI 02912, (401) 863-3384.

Although not strictly about science, this series of articles discusses performance assessment topics and goals for students that are of relevance to math. The articles are: *Rethinking Standards; Performances and Exhibitions: The Demonstration of Mastery; Exhibitions: Facing Outward, Pointing Inward; Steps in Planning Backwards; Anatomy of an Exhibition*; and *The Process of Planning Backwards*.

These articles touch on the following topics: good assessment tasks to give students, the need for good performance criteria, the need to have clear targets for students that are then translated into instruction and assessment, definition and examples of performance assessments, brief descriptions of some cross-disciplinary tasks, the value in planning performance assessments, and the notion of planning backwards (creating a vision for a high school graduate, taking stock of current efforts to fulfill this vision, and then planning backward throughout K-12 to make sure that we are getting students ready from the start)

(TC# 150.6VARARD)

Colison, J. ***Connecticut's Common Core of Learning*****, 1990. Available from: Performance Assessment Project, Connecticut Department of Education, Box 2219, Hartford, CT 06145, (203) 566-4001.**

The Connecticut Department of Education is developing a series of performance assessments in science and math. Each task has three parts: individual work to activate previous knowledge, group work to plan and carry out the task, and individual work to check for application of learning. This document provides:

1. A lengthy description of one of the ninth grade science tasks: "speeders"
2. Short descriptions of 24 performance tasks in science (8 each in chemistry, physics, and earth sciences), and 18 in math.
3. A group discussion self-evaluation form to be used by students

No technical information or general scoring guides are included in this document

(TC# 600.3CONSCI)

Collis, Kevin F. and Thomas A. Romberg. ***Assessment of Mathematical Performance: An Analysis of Open-ended Test Items*****, 1989. Available from: National Center for Research in Mathematical Sciences Education, Wisconsin Center for Education Research, University of Wisconsin, School of Education, 1025 W. Johnson St., Madison, WI 53706, (608) 263-4200.**

This paper discusses the implications of research on cognitive development in math for designing assessments. This discussion leads up to some general considerations for assessment design and a general summary of current assessment trends. Some sample test items are provided to illustrate some of the points. Also some sample performance assessment-type items are shown, but they are not critiqued in light of the previous discussion.

(TC# 500.6ASSMAP)

Collis, Kevin F. and Thomas A. Romberg. *Collis-Romberg Mathematical Problem Solving Profiles*, 1992. **Available from: Australian Council for Educational Research Limited, Radford House, Frederick Street, Hawthorn, Victoria 3122, Australia. Also available from: ASHE, PO Box 31576, Richmond, VA 23294, (804) 741-8991.**

This assessment device for students in grades 5 and 2 has 20 open-ended problems to solve--one problem in each of five areas (algebra, chance, measurement, number, and space) with four questions per problem area. Each question is designed to tap a developmental level of formal reasoning For example, the "A" question determines whether the student can use one obvious piece of information from the item, while the "D" question determines whether the student can use an abstract general principle or hypothesis derived from the information in the problem.

Responses to each question are scored right/wrong. The number of correct responses on each task determine a developmental level. Suggestions are given for instructional strategies for the various developmental levels. Technical information in the manual includes typical performance for various grade levels, teacher judgment on the developmental level indicated by each task, and additional analyses to show validity of the inferences drawn.

(TC# 500.3COLROM)

Commission on Standards for School Mathematics. *Curriculum and/Evaluation Standards for School Mathematics*, 1989. **Available from: National Council of Teachers of Mathematics, 1906 Association Dr., Reston, VA 22091.**

This book contains standards for curriculum and assessment that attempt to create a coherent vision of what it means to be mathematically literate. This book has been quoted extensively and appears to be the current "standard" for what should be in a math curriculum.

The assessment section covers: three statements of philosophy concerning assessment (alignment, multiple sources of information, and appropriate assessment methods and uses); seven sections on assessing various student outcomes (e.g., problem solving, communication, reasoning, concepts, procedures, and dispositions); and four sections on program evaluation (indicators, resources, instruction, and evaluation team). Each of the seven sections on assessing student outcomes briefly describes what the assessment should cover and provides some sample assessment tasks and procedures.

(TC# 500.5CURANE)

Csongor, Julianna E. *Mirror, Mirror On The Wall...Teaching Self-Assessment to Students.* **Located in: The Mathematics Teacher, 85, November 1992, pp. 636-637. Also available from: Saint Maria Gosetti High School, 10th and Moore, Philadelphia, PA 19148.**

The author presents a procedure for getting high school students to self-reflect in math: during the final five minutes of a test, students estimate how sure they are about each answer they gave on the test (100%, 75%, 50%, or 0%). They can earn extra credit on the test if their estimates fall within 3% of their actual score. She reports that students are surprisingly accurate in their estimates and that the procedure works especially well with slow learners.

(TC# 500.3MIRMIW)

EQUALS. *Assessment Alternatives in Mathematics*, **1989. Available from: University of California, Lawrence Hall of Science, Berkeley, CA 94720, (415) 642-1823.**

This document provides an overview of some possible assessment methods in mathematics that cover both process and products. Specific examples are provided for writing in mathematics, mathematical investigations, open-ended questions, performance assessment, observations, interviews, and student self-assessment. Any of the student-generated material could be self-selected for a portfolio of work. The document also includes a discussion of assessment issues and a list of probing questions teachers can use during instruction.

(TC# 500.6ASSALI)

Ferguson, Shelly. ***Zeroing in on Math Abilities*****, 1992. Located in: Learning92, 21, pp. 38-41.**

The paper was written by a fourth grade teacher and describes her use of portfolios in math -- what she has students put in their portfolios, the role of self-reflection, getting parents involved, and grading. She gives a lot of practical help. One interesting idea in the paper has to do with grading. At the end of the grading period she reviews the portfolios for attainment of concepts taught (not amount of work done), and progress toward six goals set by the NCTM standards (e.g., thinks mathematically, communicates mathematically, and uses tools). She marks which goals were illustrated by the various pieces of work in the portfolio and writes a narrative to the student.

Another interesting idea is formal presentations of their portfolios by students to their parents. The article provides a sample comment form for parents and students to complete.

(TC# 500.3ZERMAA)

Fitzpatrick, Anne R., Kadriye Ercikan, and Steven Ferrara. ***An Analysis of the Technical Characteristics of Scoring Rules for Constructed-Response Items*****, 1992. Available from: CTB Macmillan/McGraw-Hill, PO Box 150, Monterey, CA 93942-0150, (800) 538-9547.**

This was a paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, April 1992.

This paper reports on a technical study of the open-response portion of the 1991 administration of the Maryland state tests in reading and math. Items had a variety of scoring formats including different number of possible points and scoring tied to individual tasks. Results showed that the math open-response questions were hard, discriminated well between students having different achievement levels, and worked better when more score points were used. Thus, there is evidence that this set of open-response questions might offer more measurement accuracy than multiple-choice questions.

(TC# 060.6ANATEC)

Fitzpatrick, Robert and Edward J. Morrison. *Performance and Product Evaluation.* Located in: Educational Performance Assessment, Fredrick L. Finch (Ed.), 1991. Available from: The Riverside Publishing Company, 8420 Bryn Mawr Ave., Chicago, IL 60631, (800) 323-9540.

This paper has interesting discussions of the following topics:

1. What "authenticity" in tasks means. The authors' position is that there are many degrees and kinds of artificialities in tests. "Performance and product evaluation are those in which some criterion situation is simulated to a much greater degree than is represented by the usual paper-and-pencil test.... [However,] there is no absolute distinction between performance tests and other classes of tests--the performance test is one that is *relatively* realistic."

2. Criteria for deciding how much "reality" to include in tasks.

3. Descriptions of various types of tasks that can be used in performance assessments: in-basket, games, role-plays, projects, etc.

4. Steps for developing performance assessments: analysis of the important dimensions of the skills to be covered, identification of tasks that cover as many of the important skills as possible, developing instructions and materials, and developing the scoring procedure.

Most specific examples are taken from military and business applications.

(TC# 150.6PERPRE)

Fraser, Barry J., John A. Malone, and Jillian M. Neale. *Assessing and Improving the Psychosocial Environment of Mathematics Classrooms.* Located in: Journal for Research in Mathematics Education, 20, 2, 1989, pp. 191-201.

This article describes the development of a short form of the *My Class Inventory* to be used in sixth grade math classes to measure the psychosocial characteristics of the classroom learning environment, i.e., social interactions.

(TC# 500.3ASSIMP)

Glaser, Robert. *Expert Knowledge and Processes of Thinking*. **Located in: Enhancing Thinking Skills in the Sciences and Mathematics, Diane Halpern (Ed.), 1992. Available from: Lawrence Erlbaum Associates, Publisher, 365 Broadway, Hillsdale, NJ 07642, (800) 926-6579.**

In this article the author describes research on expert performance. Although not directly about assessment, expert performance can be used to help understand and define the targets we have for students, which is the first step toward designing assessment. For example, expert performance can be used to develop criteria for evaluating performance tasks.

The author points out that although expertise is very subject-specific, generalizations can be made about its nature across subjects: experts perceive large, meaningful patterns, have skillful self-regulatory processes, etc.

A critical point made by the author is that, "Practice, as it comes about in the usual course of training, is not necessarily very efficient. On the basis of our knowledge of the specific aspects of competence and expertise, we are able to find ways to compress or shortcut experience...." This is one goal for performance assessment, we help students understand current conceptions of the relevant dimensions of a task so that they don't have to rediscover this themselves.

(TC# 050.6EXPKNP)

Grady, Emily. *Grady Profile Portfolio Assessment Product Demo*, **1991. Available from: Aurbach & Associates, Inc., 8233 Tulane Ave., St. Louis, MO 63132, (314) 726-5933.**

This document contains demo materials for a software package that allows the user to collect, store and retrieve a variety of student products and information using a Mac Hypercard system. The document includes a rationale statement for portfolios, a description of the software product, and a demo disk that allows the user to see how the system works with one case example. The user still needs to plan what work will be collected and how to assess progress (although there does appear to be some sort of checklist built into the system).

(Note: the disk and written materials are shelved separately. In the shelf numbers below, "d" is the demo disk, and "t" is the written materials.)

(TC# 000.3GRAPRPd *and* TC# 000.3GRAPRPt)

Grobe, R. P., K. Cline, and J. Rybolt. ***[Mount Diablo] Curriculum Based Assessment For Math: A Summary of 1990 Field-Test Results*, 1990. Available from: Mt. Diablo Unified School District, 1936 Carlotta Dr., Concord, CA 94519.**

The 1990 project in Mt. Diablo Unified School District entailed scoring open-ended math problems holistically on a scale of 0-4. The scale for grades 3, 5 and 8 defines an exemplary response as: systematic or elegant, organized recording system, completed and accurate, and clear and thorough explanation. One problem for each grade, along with sample student responses, is included for each grade level. A rationale for using open-ended problems is also provided. Some information on teacher reactions is included. No other technical information is included.

(TC# 500.3MIDIAC)

Hall, Greg. ***Alberta Grade 9 Performance-Based Assessment--Math*****, 1992. Available from: Greg Hall, Student Evaluation Branch, Alberta Education, Box 43, 11160 Jasper Ave., Edmonton, AB T5K 0L2, Canada.**

The 1992 ninth grade math performance assessment entailed six stations with hands-on activities. Students circulate through the stations; testing time for each group of six students is 90 minutes. Some of the six tasks were open-response and some were open-ended; all were assessed for problem solving. The six tasks involved applications of rearranging squares to form different perimeters for the same area, measurement and mapping, surface area, collecting and graphing information, estimation, and combinations/permutations.

Responses were scored using an analytical trait system having two dimensions: problem solving and communication. Each trait was scored on a scale of 0 (totally misunderstood or blank) to 3 (readily understood the task, developed a good strategy, carried out the strategy and generalized the conclusion). A few *possible* student responses are included to illustrate scoring, but no *actual* student responses. No technical information is included.

(TC# 500.3ALBGRN)

Halpern, Diane (Ed.). ***Enhancing Thinking Skills in the Sciences and in Mathematics*****, 1992. Available from: Lawrence Erlbaum Associates, Publishers, 365 Broadway, Hillsdale, NJ 07642, (800) 926-6579.**

This book is not strictly about assessment. Rather, it discusses the related topics of "What should we teach students to do?" and "How do we do it?" The seven authors "criticize the conventional approach to teaching science and math, which emphasizes the transmission of factual information and rote procedures applied to inappropriate problems, allows little opportunity for students to engage in scientific or mathematical thinking, and produces inert knowledge and thinking skills limited to a narrow range of academic problems." (p. 118). In general, they recommend that teachers focus on the knowledge structures that students should know, use real tasks, and set up instruction that requires active intellectual engagement.

The authors give various suggestions on how to bring this about: instructional methods, videodiscs, group work, and a host more. The final chapter analyzes the various positions and raises theoretical issues.

(TC# 500.6ENHTHS)

Harvey, John G. *Mathematics Testing With Calculators: Ransoming the Hostages.* Located in: Mathematics Assessment and Evaluation: Imperatives for Mathematics Educators, Thomas A. Romberg (Ed.), 1992. Available from: State University of New York Press, State University Plaza, Albany, NY 12246.

This paper looks at the use of calculators in mathematics testing. The premise is that if we want students to investigate, explore and discover, assessment must not just measure mimicry math. Tests designed to really require calculators are more likely to be able to do this. Additionally, it is important to incorporate calculators into the curriculum because in the technological world of the future, calculators will be essential. If we want teachers to use calculators in instruction, we need to incorporate them into testing.

The author analyzes three types of test with respect to calculator use, describes things to consider when designing calculator tests, and describes current activity in developing "calculator-active" tests.

(TC# 500.6MATTEC)

Hawaii Department of Education. ***Using Portfolios: A Handbook for the Chapter 1 Teacher*****, 1991. Available from: Hawaii Department of Education, Chapter 1 Office, 3430 Leahi Ave., Bldg. D, Honolulu, HI 96815, (808) 735-9024.**

This handbook was developed to help teachers explore the possibilities of using portfolios for documenting progress of Chapter 1 students. The handbook includes rationale, philosophy, suggestions for contents, and the tie to Chapter 1 regulations. There are separate sections for reading, writing and math. Each section contains a sample portfolio, sample student outcomes, possible portfolio entries, and other resources.

(TC# 010.6USIPOH)

Illinois State Board of Education. ***Defining and Setting Standards for the Illinois Goal Assessment Program, (IGAP)*****, 1991. Available from: Illinois State Board of Education, 100 N. 1st St., Springfield, IL 62777.**

This paper describes Illinois' procedure for setting standards on the IGAP in grades 3, 6, 8, and 11. The steps include:

1. Creating descriptions of what students look like at three levels of competence: does not meet the state goal for learning, meets the state goal for learning, and exceeds the state goal for learning

2. Judgments by educators of the percent of students at each level that are likely to get each item correct

3. Adjustment of judgments by looking at the actual percentage of students getting the items correct

The paper includes a description of the process and descriptions of students at grades 3, 6, 8, and 11 at each level of competence in math.

(TC# 000.6DEFSES)

Kansas State Board of Education. ***Kansas Mathematics Standards and 1991 Kansas Statewide Pilot Assessment Results*****, 1991. Available from: Kansas State Board of Education, Kansas State Education Building, 120 SE 10th Ave., Topeka, KS 66612.**

This is an overview of the 1991 Kansas pilot math assessment and a description of results. Students from grades 3, 7, and 10 were tested. The pilot included both multiple-choice and open-performance problems. The performance assessment portion entailed giving 1/6 of the students tested one task each. A total of 31 tasks were used altogether in the three grades. Nine problems are included in the report.

Responses were scored using both a holistic scale (0-6) for overall correctness of response, and a four-trait analytic model focusing on problem-solving processes (understanding the question, planning, implementing the strategies selected, and verifying the results). Each trait is rated on a six-point scale (A-F). Scoring guides are included, but detailed instructions and sample student work are not.

Some information on student performance is included, but no other technical information on the test itself.

(TC# 500.3KASMAS)

Kentucky Department of Education. ***Kentucky Instructional Results Information System (KIRIS) Open-Response Released Items*** **, 1991-92. Available from: Advanced Systems in Measurement & Evaluation, Inc., PO Box 1217, 171 Watson Rd., Dover, NH 03820, (603) 749-9102. Also available from: Kentucky Department of Education, Capitol Plaza Tower, 500 Mero St., Frankfurt, KY 40601, (502) 564-4394.**

This document contains only the released sets of exercises and related scoring guides from Kentucky's 1991-92 grade 4, 8, and 12 open-response tests in reading, math, science, and social studies. It does not contain any support materials such as: rationale, history, technical information, etc.

There are three to five tasks/exercises at each grade level in each subject. Most are open-response (only one right answer), but some are open-ended (more than one right answer). Examples in math are: write a word problem that requires certain computations, determine how many cubes are needed for a given figure, follow instructions, explain an answer, arrange a room, and explain a graph. Examples in science are: experimental design for spot remover, graph and interpret results of a study on siblings, and predict the weather from a weather map.

Scoring for each exercise is holistic/primary trait. Each exercise has its own set of scoring criteria.

Kentucky has given educators permission to copy this document for their own use.

(TC# 060.3KENINR)

Knight, Pam. ***How I Use Portfolios in Mathematics*****, 1992. Located in: Educational Leadership, 49, pp. 71-72. Also available from: Twin Peaks Middle School, Poway Unified School District, 14012 Valley Springs Road, Poway, CA 92064.**

The author describes her first year experimentation with portfolios in her middle school algebra classes. She had her students keep all their work for a period of time and then sort through it to pick entries that would best show their effort and learning in algebra and the activities that had been the most meaningful. There is some help with what she did to get started and discussion of the positive effects on students. There is some mention of performance criteria, but no elaboration. One student self-reflection is included, but no technical information.

(TC# 530.3HOWIUS)

Koretz, Daniel, Daniel McCaffrey, Stephen Klein, Robert Bell, and Brian Stecher. ***The Reliability of Scores from the 1992 Vermont Portfolio Assessment Program--Interim Report*****, December 1992. Available from: RAND Institute on Education and Training, National Center for Research on Evaluation, Standards, and Student Testing, UCLA Graduate School of Education, 10880 Wilshire Blvd., Los Angeles, CA 90024, (310) 206-1532.**

Beginning in 1990, RAND has been carrying out a multi-faceted evaluation of Vermont's portfolio assessment program. This paper reports on reliability findings of the study conducted during school year 1991-92. Basically, RAND found that interrater agreement on portfolio scores was very low for both writing and math. The authors speculate that this resulted from aspects of scoring systems, aspects of the operation of the program, and the nature and extent of training raters.

This report provides good advise and caution for others setting up portfolio systems for large-scale assessment.

(TC# 150.6RELSCV)

Koretz, Daniel, Brian Stecher, and Edward Deibert. *The Vermont Portfolio Assessment Program: Interim Report on Implementation and Impact, 1991-92 School Year.* **Available from: RAND Institute on Education and Training, National Center for Research on Evaluation, Standards, and Studen Testing, UCLA Graduate School of Education, 10880 Wilshire Blvd., Los Angeles, CA 90024, (310) 206-1532.**

Beginning in 1990, RAND has been carrying out a multi-faceted evaluation of Vermont's portfolio assessment program. This paper reports on questionnaires and interviews conducted during school years 1990-91 and 1991-92. Results indicated that:

1. There was a significant impact on instruction, but teachers felt somewhat confused about what they were supposed to do.

2. The portfolios took a lot of classroom space and tended to be viewed by teachers as an add-on rather than as "the" instruction.

3. Teachers felt they knew more about students as the result of doing portfolios.

4. Students had some difficulty doing portfolio problems.

5. Reported effect on low achieving students was mixed.

(TC# 150.6VERPOP)

Kulm, Gerald. (Ed.) *Assessing Higher Order Thinking in Mathematics*, **1990. Available from: American Association for the Advancement of Science, 1333 H Street NW, Washington, DC 20005, (301) 645-5643.**

This book contains a series of articles that address various topics in mathematics assessment. The articles address three broad topics:

1. The rationale for assessing mathematics problem solving and the need to have assessment devices that reflect this emphasis.

2. Issues that come up when trying to assess higher-order thinking skills in mathematics.

3. General discussions of what to assess and how to assess it.

There are a few examples of actual assessment techniques. The most relevant articles are included on this bibliography as separate entries.

(TC# 500.6ASSHIO)

Lane, Suzanne. ***QUASAR Cognitive Assessment Instrument, (QCAI)*****, 1993. Available from: QUASAR (Quantitative Understanding: Amplifying Student Achievement and Reasoning), Learning Research & Development Center, University of Pittsburgh, 3939 O'Hara St., Pittsburgh, PA 15260. (412) 624-7791.**

The QCAI (QUASAR Cognitive Assessment Instrument) is designed to measure long-term growth of students in the area of math thinking and reasoning skills. Information for this review was taken from the following publications: *Principles for Developing Performance Assessments: An Example of Their Implementation* (Lane & Carol Parke, AERA, 1992); *Empirical Evidence for the Reliability and Validity of Performance Assessments* (Lane, Clement Stone, Robert Ankenmann & Mai Liu, AERA, 1992); *The Conceptual Framework for Development of a Mathematics Performance Assessment Instrument* (Lane, AERA, 1991); *Validity Evidence for Cognitive Complexity of Performance Assessments: An Analysis of Selected QUASAR Tasks* (Maria Magone, Jinfa Cai, Edward Silver, and Nign Wang, AERA, 1992); and *Conceptual and Operational Aspects of Rating Student Responses to Performance Assessments* (Patricia Kenney and Huixing Tang, AERA, 1992).

Thirty-three tasks were designed for sixth and seventh graders. No single student receives more than nine tasks in any 45-minute sitting. The tasks were designed to provide a good sample of math thinking and reasoning skills by having a variety of representations, approaches and problem strategies. Specifically, students were asked to provide a justification for a selected answer or strategy, explain or show how an answer was found, translate a problem into another representation (picture or equation), pose a mathematical question, interpret provided data, and extend a pattern and describe underlying regularities. The tasks were carefully field-tested for bias and confusing or difficult instructions. General descriptions for all the tasks, and details on a few individual tasks are provided in these materials.

Scoring is done via a generalized holistic 4-point rubric which directs raters to consider mathematical knowledge, strategic knowledge and communication. (Each of these dimensions is laid out very clearly and could be used as the basis of an analytical trait scoring scale.) The generalized rubric is then applied to each problem by specifying features of responses that would fall at different scale points. The generalized scoring guide is included in these materials, but not the task-specific adaptations.

(TC# 500.3QUACOA)

Larter, Sylvia. ***Benchmarks: The Development of a New Approach to Student Evaluation*****, 1991. Available from: Toronto Board of Education, 155 College Street, Toronto, ON M5T 1P6, CANADA, (416) 598-4931.**

Benchmarks are student performances on tasks tied to Provincial educational goals. Each Benchmark activity lists the goals to be addressed, the task, and the scoring system. To develop the Benchmarks, two observers were used for each student--one to interact with the student and one to record observations. Tasks vary considerably. Some require very discrete responses (e.g., knowledge of multiplication facts using whatever means the student needs to complete the task), while some are more open-ended. There are 129 Benchmarks developed in language and mathematics for grades 3, 6, and 8.

For many of the tasks, a general, holistic, seven-point scale ("no response" to "exceptional performance [rare]") was used as the basis to develop five-point holistic scoring scales specific to each task. For other tasks, scoring appears to be right/wrong. Holistic scoring seems to emphasize problem solving, method of production, process skills, and accuracy, although students can also be rated on perseverance, confidence, willingness, and prior knowledge, depending on the Benchmark.

The percentage of students at each score point (e.g., 1-5) is given for comparison purposes, as are other statistics (such as norms) when appropriate. Anchor performances (e.g., what a "3" performance looks like) are available either on video or in hard copy.

This report describes the philosophy behind Benchmarks, how they were developed, and a few of the specific Benchmarks. Some technical information is described (factor analysis, rater agreement), but no student performances are provided.

(TC# 100.6BENCHM)

Lash, Andrea. ***Arithmetic Word Problems: Activities to Engage Students in Problem Analysis*****, 1985. Available from: Far West Laboratory, 730 Harrison St., San Francisco, CA 94107, (415) 565-3000.**

This is a book of arithmetic word problems selected by the author to promote problem solving. Some are multiple-choice and some are open-response. The author categorizes problems as being "word problems," "process problems," "applied problems," and "puzzle problems.". The author also presents a model for the steps in problem solving and a discussion of the implications for instruction. Problems are grouped according to the step in the problem-solving process they relate to.

Most of the problems have only one right answer and none seem to utilize manipulatives. However, problems are presented for addition, subtraction, multiplication, division, multi-step problems, and problems containing unnecessary information.

(TC# 500.2ARIWOP)

Lash, Andrea. ***An Assessment of Mathematical Problem-Solving Skills*****, 1985. Available from: Far West Laboratory, 730 Harrison St., San Francisco, CA 94107, (415) 565-3000.**

This monograph describes a study which examined seventh graders' skill in one aspect of mathematical problem solving--problem analysis. Problem analysis includes identifying information necessary to solve a problem, separating relevant from irrelevant information.

identifying intermediate steps, and representing the information in a problem with a table or diagram.

The monograph describes possible assessment procedures for problem analysis (rating of open-ended solutions, purposeful multiple-choice), why they selected the latter procedure, and the types of problems that elicit problem analysis skills. The complete instrument is included.

(TC# 510.3ANASSO)

Leach, Eilene L. ***An Alternative Form of Evaluation that Complies with NCTM's Standards.*** **Located in: The Mathematics Teacher, 85, November 1992, pp. 628-632. Also available from Centaurus High School, 10300 S. Boulder Rd., Lafayette, CO 80026.**

This teacher uses scored discussions to assess and promote problem solving, communicating mathematically, and group process skills in her high school math classes. She has three to six students face each other in front of the rest of the class and spend about five minutes trying to solve a problem. Individuals can earn positive points for such things as "determining a possible strategy to use," "recognizing misused properties or arithmetic errors," or "moving the discussion along." They can earn negative points by doing such things as: "not paying attention or distracting others," and "monopolizing."

The article has a thorough discussion of how the teacher sets up the classroom, introduces the procedure to students, scores the discussion, and handles logistics. She also discusses the positive effects this procedure has had on students, and the additional insight she has obtained about her students.

All her scoring is teacher-centered, but it wouldn't necessarily have to be. No technical information is included.

(TC# 500.3ALTFOE)

Lehman, Michael. ***Assessing Assessment: Investigating a Mathematics Performance Assessment*****, 1992. Available from: The National Center for Research on Teacher Learning, 116 Erickson Hall, Michigan State University, East Lansing, MI 48824-1034.**

This monograph, by a high school math teacher, describes his attempt to develop a better method of assessing algebra problem solving, concepts, and skills than traditional paper and pencil tests. The assessment technique involves giving students problems to solve as a group, and then having them explain their results in front of a panel of judges. Three examples of problems are provided, as is a brief description of the scoring criteria (making sense of the problem, and problem-solving strategies), accuracy of results, interpreting results, ability to communicate results, and an explanation of what they did. However, these criteria are not elaborated on, and, although samples of student explanations are provided, these are used to describe the understandings the teacher reached about his students, not to anchor the performance criteria.

The author also provides a brief summary of the strategies he uses to help students develop greater depth in their understanding of algebraic principals and their interrelationships--small group cooperative learning, r[illegible]ring justifications of approaches, etc.

(TC# 530.3ASSASI)

Lehman, Michael. ***Performance Assessment--Math*, 1992. Available from: Michael Lehman, Holt Senior High School, 1784 Aurelius Rd., Holt, MI 48842, (517) 694-2162.**

This paper is related to the one above, and provides additional information. Students are given six problems (some having only one right answer and some having more than one right answer) to solve as a team (four students per team). The team then spends an hour with a panel of three judges Judges can ask any student to explain the team's solution and problem-solving strategy on any of the six problems. (Therefore, all students must have knowledge of all six problems.) Then the judges assign the team a new problem to work on while they watch.

Student responses are scored on: making sense of the problem, solution strategies, accuracy of results, ability to communicate results, ability to answer questions posed by the judges, three judgments of group process skills, and an overall judgment of student understanding

A complete set of 10 tasks (six pre-assigned, and four on-the-spot) are included for Algebra II The scoring guide and a few sample pre-calculus projects are also included No technical information or sample student performances are included.

(TC# 500.3PERASM)

Lesh, Richard. ***Computer-Based Assessment of Higher Order Understandings and Processes in Elementary Mathematics.* Located in: <u>Assessing Higher Order Thinking in Mathematics</u>, Gerald Kulm (Ed.), 1990. Available from: American Association for the Advancement of Science, 1333 H Street NW, Washington, DC 20005, (301) 645-5643.**

This article is as much about how meaningful learning occurs and the nature of the structure of knowledge in mathematics, as it is about use of computers in math instruction and assessment. The basic premise is that computer-based tests should not simply be pencil-and-paper tests delivered on-line. They should be part of an integrated instruction and assessment system that supports both learning facts and developing the meaningful internal structuring of these facts to form a coherent knowledge system.

The article discusses three things:

1. principles underlying a modeling perspective of learning and assessment (ideas such as learning and problem-solving situations are interpreted by the learner by mapping them to internal models, and several "correct" alternative models may be available to interpret a given situation)

2. five objectives that should be emphasized in K-12 math (such as going beyond isolated bits of knowledge to construct well-organized systems of knowledge, and think about thinking)

3 specific types of assessment items that can be used to measure these deeper and broader understandings (such as conceptual networks and interactive word problems)

Many sample problems are provided.

(TC# 500.6COMBAA)

Lester, Frank K, Jr. ***An Assessment Model for Mathematical Problem Solving.*** **Located in: Teaching Thinking and Problem Solving, 10, September/October, 1988, pp. 4-7. Also available from: Lawrence Erlbaum Associates, Inc., Journal Subscription Department, 365 Broadway, Hillsdale, NJ 07642, (800) 962-6579**

This article presents a model for assessing both the problem solving performance of students and assessing the task demands of the problem to be solved. The dimensions of problem solving (which could be used as a scoring rubric) are: understanding/formulating the question in a problem, understanding the conditions and variables in the problem, selecting the data needed to solve the problem, formulating subgoals and selecting appropriate solution strategies to pursue, implementing the solution strategy and attaining subgoals, providing an answer in terms of the data in the problem, and evaluating the reasonableness of an answer. The article describes these in some detail.

The problem features that can affect a student's success in solving a problem are: the type of problem, the strategies needed to solve it, the mathematical content/types of numbers used, and the sources from which data need to be obtained to solve the problem.

(TC# 500.3ANASSM)

Lester, Frank K. Jr., and Diana Lambdin-Kroll. ***Assessing Student Growth in Mathematical Problem Solving.*** **Located in: Assessing Higher Order Thinking in Mathematics, Gerald Kulm (Ed.), 1990. Available from: American Association for the Advancement of Science, 1333 H Street NW, Washington, DC 20005, (301) 645-5643.**

The authors present a model of factors that influence problem-solving performance, and discuss several problem-solving assessment techniques

A good assessment program in math should collect information about the following: affect (attitudes, preferences, and beliefs), and cognitive/processes ability to get the right answer (both whether they get the right answer, and the strategies used). The program should also systematically define and cover the features of tasks (problem type, math content, required strategies, etc.) since these affect performance and should be reflected in instruction

In order to gather information on these three categories of factors, the authors briefly review: observations, interviews, student self-reports, and holistic and analytic scoring of performances. They recommend against multiple-choice questions.

This paper is a general theoretical discussion; no actual tasks, problems or scoring guidelines are provided.

(TC# 500.6ASSSTG)

Long, Donna J. ***Mathematics Proficiency Guide*****, 1991. Available from: Indiana Department of Education, Room 229, State House, Indianapolis, IN 46204, (317) 232-9155.**

Although not strictly about assessment, this document has a nice description of mathematics proficiencies at various grade levels tied to specific instructional tasks. Proficiencies include: problem solving strategies, reasoning, communication, developing cognitive structures, applying math across the curriculum, and various knowledges (e.g., decimal places, measurement, and geometry).

(TC# 500.5MATPRG)

Marshall, Sandra P. ***Assessing Knowledge Structures in Mathematics: A Cognitive Science Perspective.*** **Located in: <u>Cognitive Assessment of Language and Mathematics Outcomes</u>, Sue Legg & James Algina (Eds.), 1990. Available from: Ablex Publishing Company, 355 Chestnut St., Norwood, NJ 07648.**

This article discusses the implications of recent advances in cognitive science for mathematics assessment. The goal in using this research to develop assessment techniques is to determine the extent to which students have acquired specific cognitive skills rather than merely whether they can correctly solve particular problems.

Cognitive theory holds that people solve problems by using three knowledge structures--declarative (facts), procedural (algorithms and production rules), and schema (frames that relate facts and production rules). To solve a problem, a person must first find the right schema, must then correctly implement a set of production rules, and must have stored correctly the facts and knowledge required to carry out the necessary algorithms specified by the production rules. Errors can occur in any of these three areas.

Researchers are currently engaged in specifying these knowledge structures in such detail that they can develop computer simulations that can, first, solve problems, and second, reproduce student errors by leaving out or altering various parts of the necessary structures. In this way, errors in student responses can be tracked back to the erroneous structure used. The author specifically mentions work in the area of simple arithmetic operations, geometry, and word problems

Additionally, the author discusses two other ways of assessing these things in students--reaction time (to assess how automatic a function is); and multiple-choice problems (e.g., "which of the following problems can be solved in the same way as the one stated above?" to get at schema knowledge). Some time is spent with multiple-choice problems to explore various types of problems and the technical issues that arise with them.

It should be pointed out that all these procedures are experimental; none have progressed to the point where there is a final product that can be ordered and installed.

(TC# 500.6ASSKNS)

Marshall, Sandra P. ***The Assessment of Schema Knowledge for Arithmetic Story Problems: A Cognitive Science Perspective*, 1990. Located in: <u>Assessing Higher Order Thinking in Mathematics</u>, Gerald Kulm (Ed.). Available from: American Association for the Advancement of Science, 1333 H Street NW, Washington, DC 20005, (301) 645-5643.**

The Story Problem Solver (SPS) was created to support instruction based on a theory of memory architecture called schemata. Under such theories, human memory consists of networks of related pieces of information. Each network is a schema--a collection of well-connected facts, features, algorithms, skills, and/or strategies.

Adult students are explicitly taught five problem-solving schemas and how to recognize which schema is represented by a story problem. SPS is a computerized assessment method in which several different item types are used: students pick out the schema or general solution strategy that fits a given story problem, decide which information in the story problem fits into the various frames of the schema, identify the steps needed to solve a problem, and decide whether the necessary information is given in the problem.

Some of the schema shells and item types are given as examples. No technical information is included.

(TC# 500.3ASSOFS)

Maryland Department of Education. ***Maryland School Performance Assessment Program*, 1991. Available from: Gail Lynn Goldberg, Maryland Department of Education, Maryland School Performance Assessment Program, 200 W. Baltimore St., Baltimore, MD 21201, (410) 333-2000.**

Maryland has released six performance tasks that illustrate the 1992 assessment. This review is based on three of them, one task at each of grades 3, 5 and 8. The tasks are integrated across subject areas and use some combination of information and skills in science, math, writing, reading, and social studies. The three tasks we have relate to the weather (Grade 3), snowy regions of the country (Grade 5) and collisions (Grade 8). Each task has both individual and group work and proceeds through a series of tasks that require reading, designing and conducting experiments, observing and recording information, and writing up results.

Student responses are scored using two basic approaches: generalized holistic or analytical trait scoring guides for the "big" outcomes such as communication skills, problem solving, scientific process, and reasoning; and specific holistic ratings of conceptual knowledge and applications. For example, the task on collisions is scored both for knowledge of the concepts of mass and rate/distance, and for general science process skills (collecting and organizing data, and observation) and communication skills. Thus, some scoring guides are generalized across tasks, and some list specific features from individual tasks to watch for.

The materials we have allude to anchor performances and training materials, but these are not included in our samples. Neither information about student performance, nor technical information about the tests is included.

(TC# 500.3MDSCMA)

Maryland State Department of Education. ***Scoring MSPAP (Maryland School Performance Assessment Program): A Teacher's Guide*****, 1993. Available from: Gail Lynn Goldberg, Maryland Department of Education, Maryland School Performance Assessment Program, 200 W. Baltimore St., Baltimore, MD 21201, (410) 333-2000.**

This document presents information about the 1993 MSPAP: philosophy, general approach, sample tasks, and performance criteria. There are sample tasks, performance criteria and student responses for the following areas: expository, persuasive and expressive writing, reading comprehension, math, science, and social studies.

Scoring can be done three different ways depending on the task: generalized scoring rubrics that can be used across tasks (e.g., persuasive writing); generalized scoring rules that are not as detailed as rubrics (e.g., language usage); and scoring keys that are task-specific (e.g., many math tasks are scored for the degree of "correctness" of the response)

No technical information is included.

(TC# 000.3SCOMST)

Maryland State Department of Education. ***Teacher to Teacher Talk: Student Performance on MSPAP (Maryland School Performance Assessment Program)*****, 1992. Available from: Gail Lynn Goldberg, Maryland Department of Education, Maryland School Performance Assessment Program, 200 W. Baltimore St., Baltimore, MD 21201, (410) 333-2000.**

This publication presents teacher reactions to their experience of scoring performance assessment tasks on the 1992 Maryland School Performance Assessment Program (MSPAP). The MSPAP covered reading, writing, math, social studies and science in grades 3, 5, and 8

Comments are organized by grade and subject. Most comments have to do with two topics: what teachers learned about students as the result of participating in the scoring, and how the performance tasks should be revised.

(TC# 000.6TEATET)

Marzano, Robert J., Debra J. Pickering, Jo Sue Whisler, et al. ***Authentic Assessment*****, undated. Available from: Mid-Continent Regional Laboratory (McREL), 2550 S. Parker Rd., Suite 500, Aurora, CO 80014, (303) 337-0990.**

This document appears to be a series of handouts used in training. Although not specifically about math, the document does discuss some "big" outcomes related to math such as complex thinking, information processing, communication, etc.

Materials include definitions of assessment terms, a procedure for developing performance assessment tasks, and samples of tasks and scoring guides. The general approach is mix and match--tasks are meant to elicit several target behaviors on the part of students which are then scored with generic performance criteria. For example, a problem-solving task requires students to draw a picture of their neighborhoods without using any circles or squares. Performances are scored for knowledge (geometry), complex thinking (ability to identify obstacles in the way of achieving desired outcomes), and effective communication (ability to express ideas clearly).

Sample tasks are in the areas of science, math and social studies. There are general mix and match scoring guides for: Knowledgeable Person, Complex Thinker, Information Processor, Effector Communicator/Producer, Self-Directed Learner, and Collaborative Worker. Scoring guides are generally not very descriptive. For example, one of the three traits included in the scoring guide for Skilled Information Processor is "effectively interprets and synthesizes information." To get a "4" (the highest score possible) the student "consistently interprets information gathered for tasks in accurate and highly insightful ways and provides synthesis of that information that are highly creative and unique." This is basically just a restatement of the trait title.

The authors have begun to develop a useful approach to performance assessment (mix and match tasks and performance criteria), but the criteria need to be filled out a little more.

(TC# 150.6AUTASS)

Massachusetts Educational Assessment Program. ***On Their Own: Student Response to Open-Ended Tests in Mathematics*****, [Massachusetts Educational Assessment Program --** ***Math Open-Ended and Performance Tasks.*****], 1991. Available from: Dr. Allan Hartman, Commonwealth of Massachusetts, Department of Education, 1385 Hancock St., Quincy, MA 02169, (617) 770-7334.**

The document we received contained assessment materials for grades 4, 8, and 12 from three years (1988-1990) in four subject areas (reading, social studies, science and math). This entry describes the math portions of the assessments. The 1988 and 1990 materials described open-ended test items in which students had to solve a problem and then explain their answer. In 1988 eight problems were administered to each of the three grades (some problems were repeated between grades). In 1990, ten problems were administered. These problems emphasized the major areas of patterns/relationships, geometry/measurement, and numerical/statistical concepts. All problems were done individually in written format. Problems were distributed in such a way that different students responded to different questions. Responses were scored both for correctness of solution and for quality of the explanation. No specific criteria for judging quality of explanation were given. Many examples of student responses illustrating various conclusions are included.

In 1989, a sample of 2,000 students was assigned one of seven performance tasks (four in math required manipulatives) to do in diads. Each pair was individually watched by an evaluator. Each evaluator could observe between six and ten pairs each day. It took 65 evaluators five days to observe the 2,000 performances. Evaluators were to both check off those things that students did correctly (e.g., measured temperature correctly), and record observations of students' conversations and strategies as completely as possible. A sample checklist of skills includes: measuring, proportional reasoning, equivalency, numeration, attitude, and planning/execution.

Some information on results for all the assessments is provided: percentages of students getting correct answers, using various strategies, using efficient methods, giving good explanations, etc., depending on the task. Many examples of student responses illustrating these various points are provided. No technical information about the assessments themselves is provided.

(TC# 500.3MASOPM)

McTighe, Jay. ***Maryland Assessment Consortium: A Collaborative Approach to Performance Assessment*, 1991. Available from: Maryland Assessment Consortium, c/o Frederick County Public Schools, 115 E. Church St., Frederick, MD 21701, (301) 694-1337.**

This entry contains handouts from a presentation by the author in 1991. The following topics are covered:

1. A description of the consortium--what it is and what it does.

2. An overview of the process used for developing performance tasks, and review criteria for performance tasks.

3. Examples of three performance assessment tasks developed by the consortium: one math problem-solving task for grade six and two fifth grade reading tasks. All tasks are

scored using a four-point holistic scoring guide. Scoring appears to be generalized rather than tied to individual tasks. The reading tasks, for example, are scored using the same, generalized scoring guide.

(TC# 500.3MARASC)

McTighe, Jay. ***Teaching and Testing in Maryland Today: Education for the 21st Century*****, 1992. Available from: Maryland Assessment Consortium, c/o Frederick County Public Schools, 115 E. Church St., Frederick, MD 21701, (301) 694-1337.**

This 13-minute video is designed to introduce parents and community members to performance assessment.

(TC# 150.6TEATEMv)

Mead, Nancy. ***IAEP (International Assessment of Educational Progress) Performance Assessment (Science and Math)*****, 1992. Available from: Educational Testing Service, Rosedale Rd., Princeton, NJ 08541, (609) 734-1526.**

This document supplements the report by Brian Semple (also described in this bibliography) (TC# 500.6PERASS). The document contains the administrators manual, scoring guide, equipment cards, and released items from the Second International Assessment of Educational Progress in science and mathematics.

(TC# 500.3IAEPPA)

Medrich, Elliott A., and Jeanne E. Griffith. ***International Mathematics and Science Assessments: What Have We Learned?*****, 1992. Available from: National Technical Information Service, US Department of Commerce, 5285 Port Royal Rd., Springfield, VA 22161, (703) 487-4650.**

This report provides a description of the international assessments of math and science (First International Mathematics and Science Studies, 1960's; Second International Mathematics and Science Studies, 1980's; and First International Assessment of Educational Progress, 1988), some of their findings, and issues surrounding the collection and analysis of these data. It also offers suggestions about ways in which new data collection procedures could improve the quality of the surveys and the utility of future reports.

(TC# 000.6INTMAS)

Meinhard, Richard. ***A Developmental Baseline Profile of 12 Key Elementary Science Concepts/Processes*, 1990. Available from: Institute for Developmental Sciences, Oregon Cadre for Assistance to Teachers of Science (OCATS), 3957 E. Burnside, Portland, OR 97214, or by calling (214) 234-4600.**

The OCATS (Oregon Cadre for Assistance to Teachers of Science) project is designed to encourage concept/process-based science education in order to promote long-range student growth in science. One part of this project has been to gather information on how twelve science and math concepts develop in students from K to 5. The concepts are: organization of objects (simple classification, multiple classification, seriation, whole number operations), geometrical and spatial relationships of objects (perimeter, area, multiplicative projective relationships); physical properties of objects (quantity, weight, volume); experimental reasoning (controlling variables); causal explanation (proportional reasoning).

One performance task was given to the students for each concept area. Performance was rated for developmental stage: sensory-motor, pre-operational, operational, and form. Each stage has two substages for a final scale having eight points.

Descriptive information is available for 40 K-5 students. Neither the performance tasks nor the scoring techniques are described in detail in this paper. No technical information, except distribution of performance, is included.

(TC# 600.6DEVBAP)

Meltzer, L. J. ***Surveys of Problem-Solving & Educational Skills*, 1987. Available from: Educator's Publishing Service, Inc., 75 Moulton St., Cambridge, MA 02138.**

Although this is an individual test published primarily for diagnosing learning disabilities for students aged 9-14, it has some interesting ideas that could be more generally applied. There are two parts to the test--a more-or-less standard individualized aptitude test, and a series of achievement subtests. The math subtest involves a fairly standard test of computation. The interesting part comes in the scoring. Each problem is scored on choice of correct operations, ability to complete the word problem, efficiency of mental computation, self-monitoring, self-correction, attention to operational signs, and attention to detail (one point for evidence of each trait).

After the entire subtest is administered, the teacher is guided through analysis of the student's strategies in completing the task--efficiency of approaching tasks. flexibility in applying strategies, style of approaching tasks, attention to the task, and responsiveness during assessment. (Each area is assigned a maximum of three points for the presence or absence of three specific features of performance. For example, under "efficiency" the students get a point if he or she does not need frequent repeating of instructions, a second point if the student implements the directions rapidly, and a third point if the student perseveres to complete the task.) Examples of scoring are included.

A fair amount of technical information is included. This covers typical performance, factor analysis, inter-rater reliability, relationship to other measures of performance, and comparison of clinical groups.

(TC# 010.3SUROFP)

Mullen, Kenneth B. ***Free-Response Mathematics Test*, 1992. Available from: American College Testing Program, PO Box 168, Iowa City, IA 52240, (319) 337-1051.**

This was a paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, April 1992.

This paper reports on a study by ACT that compares multiple-choice, open-response, and gridded response item formats on reliability, difficulty and discrimination. In gridded response items, students fill in "bubbles" that correspond to the answer rather than choosing the answer from a given list. "Testlets" were designed to cover the same content and have the same test length for each format. Results indicated that all formats had about the same reliability; there was good rater agreement on the open-ended problems; and grid and open-ended problems discriminated better between students with different achievement levels. The correlation between performances on the various types of items ranged from 0.5 to 0.7.

A few sample problems are provided. All open-response questions used scoring criteria that emphasize degree of correctness of the response and were tied to the task (i.e., there was a different scoring guide for each problem).

(TC# 500.3FREREM)

Mumme, Judy. ***Portfolio Assessment in Mathematics*, 1990. Available from: California Mathematics Project, University of California--Santa Barbara, 522 University Rd., Santa Barbara, CA 93106, (805) 961-3190.**

This booklet describes what mathematical portfolios are, what might go into such portfolios, how items should be selected, the role of student self-reflection, and what might be looked for in a portfolio. Many student samples are provided. Criteria for evaluating portfolios include: evidence of mathematical thinking, quality of activities and investigation, and variety of approaches and investigations. No technical information is included.

(TC# 500.6PORASI)

National Science Foundation. ***Educating Americans for the 21st Century: A Plan of Action for Improving Mathematics, Science and Technology Education*, 1983. Available from: National Science Board Commission on Precollege Education in Mathematics, Science and Technology, Forms & Publications Unit, 1800 G St. NW, Room 232, Washington, DC 20550, (202) 357-3619.**

This is not strictly a document regarding assessment, but rather a statement of what students need to know and be able to do in science and math. As such, it also provides an outline for what assessments should measure.

(TC# 000.5EDUAMF)

Nicholls, John G., Paul Cobb, Erna Yackel, et al. ***Students' Theories About Mathematics and Their Mathematical Knowledge: Multiple Dimensions of Assessment*. Located in: Assessing Higher Order Thinking in Mathematics, Gerald Kulm (Ed.), 1990. Available from: American Association for the Advancement of Science, 1333 H Street NW, Washington, DC 20005, (301) 645-5643.**

This paper reports on a series of studies on student attitudes toward mathematics and their relationship to mathematical knowledge and understanding. Dimensions of attitudes toward math were:

1. how motivated students are to do math
2. student beliefs about what causes success in math
3. student views of the benefits of learning math.

All items are included.

(TC# 500.3STUTHA)

Oregon Department of Education. ***Oregon Dimensions of Problem Solving*, 1992. Available from: Michael Dalton, Oregon Department of Education, 700 Pringle Parkway, SE, Salem, OR 97310, (503) 378-8004.**

The Oregon Department of Education began giving open-ended math problems to a sample of students in grades 3, 5, 8, and 11 in 1992. The five short, written problems used in each grade in 1992 are included in this document, as are student instructions. Responses are scored on four dimensions, or traits: (1) conceptual understanding of the problem--the ability to interpret the problem and select appropriate information to apply a strategy for solution; (2) procedural knowledge--the ability to demonstrate appropriate use of math; (3) skills to solve the problem; and (4) communication--the ability to use math symbols well and ability to explain the problem solution.

Each trait is scored on a scale of 1-5. The scoring guides are included in this document along with one sample student problem. No anchor papers or technical information is included.

(TC# 500.3ORDIPS)

Padilla, Michael. ***Group Assessment of Logical Thinking, 1982.*** **Available from: University of Georgia, 212 Aderhold Hall, Athens, GA 30602, (706) 542-3000.**

The two documents we received describe enhanced multiple-choice tests to assess the level of student development from concrete to formal logical thinkers based on Piaget. The test has 21 items for students with a reading level of grade six and above. Six logical operations are assessed: conservation, proportional reasoning, controlling variables, combinatorial reasoning, probabilistic reasoning, and correlational reasoning. Content is taken from the sciences and daily life. Each item is presented pictorially. The student chooses both a statement he or she believes is true about the situation pictured, and the reason for this choice. All items are multiple-choice except for the combinatorial reasoning items for which students list all possible combinations.

There is technical information to support the conclusion that the test can distinguish groups at concrete, transitional, and formal stages of development. The authors recommend using the information obtained to design instruction at the proper developmental level for students. No concrete examples of how to do this are provided.

(TC# 600.3GROASL)

Pandey, Tej. ***Power Items and the Alignment of Curriculum and Assessment.*** **Located in: <u>Assessing Higher Order Thinking in Mathematics</u>, Gerald Kulm (Ed.), 1990. Available from: American Association for the Advancement of Science, 1333 H Street NW, Washington, DC 20005, (301) 645-5643.**

The author presents a philosophy and approach for thinking about the development of a test of mathematics problem solving, and provides some examples of multiple-choice and short-answer "power" questions developed by the California Assessment Program.

The author maintains that typical content by process matrices used to specify the content of tests tend to result in tests that measure minuscule pieces of information that are fragmented and non-integrated. The author prefers to have assessment tasks that are broader in focus and cut across several process/content areas, so that in order to get the right answer, students must use skills like organizing information, representing problems, and using strategies.

Multiple-choice or short-answer power questions:

1. Assess essential mathematical understandings and inter-connectedness of mathematical ideas, rather than isolated facts and knowledge

2. Are not directly teachable, even though teaching for them will result in good instruction.

3. Result in teacher agreement that such questions represent worthwhile teaching goals

(TC# 500.6POWITA)

Pandey, Tej. ***A Sampler of Mathematics Assessment*, 1991. Available from: California Department of Education, Bureau of Publications, Sales Unit, PO Box 944272 Sacramento, CA 94244, (916) 445-1260.**

This sampler describes the types of assessment that the California Assessment Program (CAP) is proposing to support curricular reforms. Illustrated and discussed are open-ended problems, enhanced multiple-choice questions, investigations, and portfolios. These four types of activities are intended to measure mathematical understandings that students develop over a period of several years.

This monograph includes a definition of "mathematical power"--the ultimate goal of mathematics instruction, guidance in the characteristics of assessment tasks that will encourage and measure power, a few sample student responses to problems, and help with implementation of alternative assessment.

All performance-based techniques will use a six-point holistic scale. This scale is briefly described. The scale will be tailored for individual tasks.

(TC# 500.3SAMMAA)

Paulson, Leon. ***Portfolio Guidelines in Primary Math*, 1992. Available from: Multnomah County Educational Service District, PO Box 301039, Portland, OR 97220, (503) 255-1842.**

This monograph provides some assistance with getting started with portfolios in the primary grades. The author believes that the most important purpose for mathematics portfolios is to prompt students to take control of their own learning. Therefore, the student should be in control of the portfolio. (The author, however, also points out that there might be other audiences and purposes for the portfolios that might have to be addressed.)

The author provides some ideas for tasks that students could do to generate material for the portfolio, provides some very practical suggestions for getting started, gives ideas for activities to encourage student self-reflection, and shows some draft holistic criteria for evaluating portfolios.

An example of the user-friendly way this monograph provides practical help is: "Remember, the portfolio is telling a story. Each item in a portfolio is there for a reason. It should not require a mind reader to figure out why it is there. A portfolio entry includes a piece of work plus information that makes its significance clear--the reason it was selected, the learning goals illustrated, student self-reflections, and (always!) the date."

(TC# 500.6PORGUP)

Paulson, Leon, and Pearl Paulson. ***An Afternoon to Remember: A Portfolio Open House for Emotionally Disabled Students*, 1992. Available from: Multnomah County Educational Service District, PO Box 301039, Portland, OR 97220, (503) 255-1842.**

Reynolds School District adapted Crow Island's "portfolio night" for use with severely emotionally disabled students. This paper describes how the afternoon was set up, what happened, student debriefing sessions, and changes in format based on student comments.

(TC# 000.6AFTREP)

Pfeiffer, Sherron. ***NIM Game Project*, 1992. Available from: Southeast EQUALS, 14 Thornapple Dr., Hendersonville, NC 28739, (704) 692-4078.**

The assessment described in this document is a math project task appropriate for upper elementary and middle school students. Two project tasks are included, one individual and one group. The projects require students to create a game that requires application of math skills. These extended projects are used after students have had many opportunities to work with different kinds of NIM games. The extended nature of the project emphasizes persistence and the importance of quality products. Projects become part of a portfolio that shows growth over time.

The projects are scored using criteria specific to these tasks. The criteria revolve around the quality of the game and its usefulness in teaching the math skills specified. The project instructions and scoring guide are included. No sample student work nor technical information is included. This exercise is part of a book of teaching strategies produced by and available from the author: *Successful Teaching Strategies.*

The author has given educators permission to copy this document for their own use.

(TC# 500.3NIMGAP)

Pritchard, Diane. ***Student Portfolios--Are They Worth the Trouble?*, 1992. Available from: Sisters Middle School, PO Box 555, Sisters, OR 97759, (503) 549-8521.**

This paper was written by a middle school math and English teacher. It provides practical help with how to set up a portfolio system in math by describing her purpose for having a portfolio, the types of activities included, and activities to get students to self-reflect (including an idea for tests).

(TC# 500.3STUPOT)

Psychological Corporation. ***GOALS: A Performance-Based Measure of Achievement*, 1992. Available from: Psychological Corporation, Order Service Center, PO Box 839954, San Antonio, TX 78283, (800) 228-0752.**

GOALS is a series of open-response questions that can be used alone or in conjunction with the MAT-7 or SAT-8, or any achievement test. Three forms are available for 11 levels of the test covering grades 1-12 in the subject areas of science, math, social studies, language and reading. Each test (except language) has ten items. The manual states that the math questions assess student problem solving, communication, reasoning, connections to other subjects, estimation, numeration, geometry, patterns, statistics, probability and algebra. Tasks are multiple, short problems. The manual draws the distinction between the approach taken in *GOALS* (efficiency in large-scale assessment), and the related publication "Integrated Assessment System" which has fewer tasks pursued in more depth.

Responses are scored on a scale of 0-3, where 0 is "response is incorrect" and 3 is "accurate and complete with supporting information." The scoring guide is generalized and is used for all problems. Scoring can be done locally or by the publisher. There is good assistance with scoring philosophy and procedures. There are two sample student performances for each score point for each question.

The holistic scales are combined in various ways to provide indicators of overall conceptual understanding and various specific aspects of problem solving and using procedures. These are, however, not scored directly. Rather, it is analogous to multiple-choice tests in which the correct items are combined in various ways to give subtest scores.

Both norm-referenced (percentiles) and criterion-referenced (how students perform on specific concepts) score reports are available. A full line of report types (individual, summary, etc.) are available.

The materials we obtained did not furnish any technical information about the test itself.

(TC# 510.3GOALS)

Psychological Corporation. ***Integrated Assessment System: Mathematical Performance Assessment*, 1991. Available from: Psychological Corporation, Order Service Center, PO Box 839954, San Antonio, TX 78283, (800) 228-0752.**

This is a series of 14 tasks designed to be used with students in grades 2-8. Two task booklets were designed for each grade level, but can also be used in spring testing of the grade below or fall testing of the grade above. Each task booklet presents a problem situation that is expanded on and applied to a series of questions. For example, various task booklets focus on symmetry, breaking a tie in an election, planning an orchard to maximize yield, and bar codes. Questions involve such things as figuring out an answer and explaining how the solution was reached, and generating a principle and applying it to a new situation.

Solutions are scored either holistically (0-6) or analytically (four, 4-point scales). The performance criteria represent generalized features of problem solving and so can be used to score performance on any task. The holistic scale is used to provide an overall picture of performance; raters look for quality of work, evidence of understanding of concepts, logical reasoning, and correct computations. The analytical traits are reasoning, conceptual knowledge, communication, and procedures. Scoring can be done either locally or by the publisher.

The set of materials we obtained includes a brief description of the scoring rubrics and one example of a scored student test. Technical information was not included.

(TC# 500.3INTASM)

Romberg, Thomas A. ***Assessing Mathematics Competence and Achievement*, 1989. Available from: National Center for Research in Mathematical Sciences Education, Wisconsin Center for Educational Research, University of Wisconsin, School of Education, 1025 W. Johnson St., Madison, WI 53706, (608) 263-4200.**

This paper describes the author's view of what it means to be literate mathematically. It then describes the instructional and assessment implications of this goal. The author believes that we need to assess not only mathematical knowledge but also the structure of the knowledge.

(TC# 500.5ASSMAC)

Romberg, Thomas A. ***The Domain Knowledge Strategy for Mathematical Assessment*, 1987. Available from: National Center for Research in Mathematical Sciences Education, Wisconsin Center for Educational Research, School of Education, 1025 W. Johnson St., Madison, WI 53706, (608) 263-4200.**

This document provides a brief overview of the "Domain Knowledge" strategy used by the National Center for Research in Mathematical Sciences Education to assess math knowledge of students. This approach is contrasted to the typically used "Content by Behavior Matrix" approach in which content topics are crossed with behavior (usually some form of Bloom's taxonomy). The author maintains that this approach is outdated; the behavior dimension fails to reflect contemporary notions of how information is processed and the content dimension is an inadequate way to describe what is meant by "knowing mathematics."

The "Domain Knowledge" approach involves making a "map" or network of a concept domain. This reflects a more integrated and coherent picture about knowledge. These maps can be used to generate tasks, assessment criteria, and formats that get at both "correctness" of responses and the strategies used to arrive at the answer.

(TC# 500.6DOMKNS)

Romberg, Thomas A. *Evaluation: A Coat of Many Colors*, 1988. **Available from: National Center for Research in Mathematical Sciences Education, Wisconsin Center for Educational Research, University of Wisconsin, School of Education, 1025 W. Johnson St., Madison, WI 53706, (608) 263-4200. Also located in: Mathematics Assessment and Evaluation: Imperatives for Mathematics Educators, Thomas A. Romberg (Ed.), 1992.**

This paper describes the impact of assessment information on decision making and describes the ways in which assessment must change if it is to have a positive impact on such decisions.

(TC# 500.6EVACOM)

Romberg, Thomas A. *Mathematics Assessment and Evaluation: Imperatives for Mathematics Educators*, **1992. Available from: State University of New York Press, State University Plaza, Albany, NY 12246.**

This book covers several interesting topics with respect to assessment in math. Specifically:

1. How tests communicate what is valued.

2. How current tests will not promote the recommendations in the NCTM standards.

3. Various considerations when developing tests. calculators, how to adequately model knowledgeable students, etc.

4. Setting up assessment that is intended to influence instruction.

Although authoritative, this book is written in a very academic style, which makes it less accessible to general readers. Articles that are most relevant to this bibliography are entered separately.

(TC# 500.6MATASE)

Romberg, Thomas A., and Linda D. Wilson. *Alignment of Tests with the Standards.* **Located in: Arithmetic Teacher, September 1992, pp. 18-22.**

The authors make the argument that teachers teach to tests. Therefore, if we want the NCTM standards to be implemented we need to have tests that reflect the standards. The authors feel that many current norm-referenced tests do not match the standards. Finally, they present tasks from several innovative assessments that they feel do reflect the standards.

(TC# 500.6ALITEW)

Romberg, Thomas A., Linda Wilson, and 'Mamphono Khaketla. *An Examination of Six Standard Mathematics Tests For Grade Eight*, 1989. Available from: National Center for Research in Mathematical Sciences Education, Wisconsin Center for Educational Research, School of Education, 1025 W. Johnson St., Madison, WI 53706, (608) 263-4200.

This study is a follow-up to the survey of teachers described above. The authors analyzed the six tests most commonly cited by the eighth grade teachers in that study as being used with their students. The authors conclude that the six standardized tests are not appropriate instruments for assessing the content, process, and levels of thinking called for in the NCTM standards.

(TC# 500.6EXASIS)

Romberg, Thomas A., Linda D. Wilson, 'Mamphono Khaketla, and Silvia Chavarria. *Curriculum and Test Alignment*. Located in: Mathematics Assessment and Evaluation: Imperatives for Mathematics Educators, Thomas A. Romberg (Ed.), 1992. Available from: State University of New York Press, State University Plaza, Albany, NY 12246.

This article reports on two studies on the alignment of current standardized tests and alternative assessments to the NCTM standards. Results showed that current standardized tests are weak in five of six content and process areas, and place too much emphasis on procedures and not enough on concepts. The authors present several examples of test questions that they feel do match the standards.

(TC# 500.6CURTEA)

Romberg, Thomas A., E. Anne Zarinnia, and Steven R. Williams. *The Influence of Mandated Testing on Mathematics Instruction: Grade 8 Teachers' Perceptions*, 1989. Available from: National Center for Research in Mathematical Sciences Education, Wisconsin Center for Educational Research, School of Education, 1025 W. Johnson St., Madison, WI 53706, (608) 263-4200.

This monograph reports on the first of a sequence of studies on mandated testing in mathematics. This study was a large-scale questionnaire survey to find out from Grade 8 teachers how influential mandated testing was on their teaching of mathematics. The results of the study showed that nearly 70 percent of the teachers reported that their students take a mandated test. Secondly, because teachers know the form and character of the tests their students take, most teachers make changes in their teaching to reflect this knowledge. Third, the kinds of changes teachers make are in contrast to the recommendations made by the NCTM standards. Specific examples are given.

Although this paper does not describe an alternative assessment device, it does provide reasons for seeking alternative ways of assessing math.

(TC# 500.6INFMAT)

Schoenfeld, Alan H. ***Teaching Mathematical Thinking and Problem Solving.*** **Located in: <u>Toward the Thinking Curriculum: Current Cognitive Research</u>, Loren B. Resnick & Leopold E. Klopfer (Eds.), 1989. Available from: Association for Supervision and Curriculum Development, 1250 N. Pitt St., Alexandria, VA 22314-1403, (703) 549-9110.**

Although this article is more about defining what mathematical problem solving is than about assessment, it presents an interesting visual way to represent how students spend their time when solving a problem. It also compares a plot of time use for a good problem solver to a plot for an inefficient problem solver.

Essentially, the plotting procedure involves tracking the sequence in which people use different steps in the problem-solving process (reading the problem, analyzing the problem, exploring a solution strategy, planning, implementing a strategy, and verifying the results) and the amount of time spent on each. Good problem solvers spend a lot of time analyzing and planning, with many self-checks on "how it is going." Poor problem solvers tend to fixate on a possible line of attack and pursue it relentlessly even when it is clearly not going well. Additionally, there are very few stops to self-check on how it is going.

(TC# 500.5STOWTET)

Semple, Brian McLean. ***Performance Assessment: An International Experiment,*** **1991. Available from: Educational Testing Service, The Scottish Office, Education Department, Rosedale Rd., Princeton, NJ 08541, (609) 734-5686.**

Eight math and eight science tasks were given to a sample of thirteen-year-olds in five volunteer countries (Canada, England, Scotland, USSR, and Taiwan). This sample was drawn from the larger group involved in the main assessment. The purpose of the assessment was to provide an information base to participating countries to use as they saw fit, and to examine the use of performance assessments in the context of international studies.

The 16 hands-on tasks are arranged in two 8-station circuits. Students spend about five minutes at each station performing a short task. Most tasks are "atomistic" in nature; they measure one small skill. For example, the 8 math tasks concentrate on measuring length, angles, and area, laying out a template on a piece of paper to maximize the number of shapes obtained, producing given figures from triangular cut-outs, etc. Some tasks require students to provide an explanation of what they did. All 16 tasks are included in this document, although some instructions are abbreviated and some diagrams are reduced in size the complete tasks, administration and scoring guides are available from ETS.

Most scoring is right/wrong; student explanations are summarized by descriptive categories. There is also observation of the products of students' work.

Student summary statistics on each task are included. There is a brief summary of teacher reactions, student reactions, the relationship between student performance on various tasks, and the relationship between performance on the multiple-choice and performance portions of the test. A few sample student performances are included.

(For related information, see Nancy Mead, also listed in this bibliography.)

(TC# 600.3PERASS)

Silver, Edward A., and Jeremy Kilpatrick. ***Testing Mathematical Problem Solving.*** **Located in: The Teaching and Assessing of Mathematical Problem Solving, Randall Charles and Edward Silver (Eds.), 1988. Available from: National Council of Teachers of Mathematics, Inc., 1906 Association Dr., Reston, VA 22091.**

This paper discusses two topics: how assessment can inform instructional decision making and how it communicates what we value. The authors propose that the National Assessment of Educational Progress and many other math tests do not provide the type of information needed for the improvement of mathematics instruction. The information useful for improvement of instruction would be types of errors kids make, how automatic mathematical processes are, and the cognitive structures and abilities associated with expertise in the domain being tested.

(TC# 500.6TESMAP)

Stalker, Veronica. ***Urbandale Alternative Assessment Project*****, 1991. Available from: Urbandale Community Schools, 7101 Airline Avenue, Urbandale, IA 50322, (515) 253-2300.**

Urbandale High School is "working to implement authentic forms of assessment throughout all of the disciplines." In all subject areas, teachers are asked to develop at least one "authentic" unit in which students are given an engaging task and which are assessed using a pre-defined rubric.

This package contains Urbandale's policy statement setting up this effort, and includes five samples of these units: projects on the environment, earthquakes, writing in math, and American history.

In a personal communication, the teacher developing the American history units makes the following points:

1. She has seen students empowered by clear performance targets presented ahead of time.

2. Assessment is daily and on-going.

3. Having an "authentic final" did not work if the rest of the class is lecture based. Students need practice with open-ended units and performance criteria.

4. The biggest challenge is not coming up with the tasks for the "authentic units" but is coming up with good performance criteria, and clearly communicating these to students.

5. In the past, she has developed a different set of performance criteria for each task report. However, now she sees that there are common threads through them, and she feels she can come up with a "master rubric" that can apply across many reports. To this master rubric, criteria specific to a given task or report can be added. The master rubric will include such things as accuracy of historical facts and how interesting the report is to read.

(TC# 0C0.3URBALA)

Stenmark, Jean Kerr. ***Mathematics Assessment: Myths, Models, Good Questions, and Practical Suggestions*****, 1991. Available from: National Council of Teachers of Mathematics, 1906 Association Drive, Reston, VA 22091.**

This monograph was designed for teachers in the elementary grades. It is a collection of examples of assessment techniques that focus on student thinking. Topics include the rationale for new ways of assessing mathematics, the necessity of integrating assessment and instruction, designing performance assessments (most emphasis is on designing the task, although sample holistic and analytical trait scoring systems are shown), what to look for during classroom observations and interactions (including questions to ask to get at various types of thinking), portfolios (including types of items to include and the types of information they can demonstrate about students, and criteria for evaluation), student self-assessment, and hints to make assessment work in the classroom. No technical information is provided.

(TC# 500.3MATASM)

Surber, John R. ***Mapping as a Testing and Diagnostic Device*****, 1984. Located in: <u>Spatial Learning Strategies--Techniques, Applications, and Related Issues</u>, C. D. Holley & D. F. Dansereau (Eds.). Available from: Academic Press, 1250 6th Ave., San Diego, CA 92101.**

The book is a general discussion of the advantages of, and procedures for, integrating the production of cognitive networks into instruction. The premise is that knowledge of facts, rules, algorithms, etc. is only part of what students need to know. They also need to know how these facts fit together to form a body of knowledge. Without knowledge of the interrelationships, students are not likely to remember the facts or be able to use them correctly when they are remembered.

The Surber paper discusses a particular type of cognitive networking scheme--mapping--and its use in assessment of knowledge structures. The basic procedure consists of taking a completed map for the topic to be tested, and deleting portions in various ways. Students then complete the map given various types of cues.

(TC# 000.6MAPASA)

Surber, John R., Philip L. Smith, Frederika Harper. ***MAP Tests*****, 1981 - undated. Available from John R. Surber, University of Wisconsin-Milwaukee, Department of Educational Psychology, Milwaukee, WI 53201, (414) 229-1122.**

Our review is based on four reports from the author: *Testing for Misunderstanding* (John R. Surber and Philip L. Smith, Educational Psychologist, 1981, 16, 3, pp. 165-174; *Technical Report No. 1, Structural Maps of Text as a Learning Assessment Technique: Progress Report for Phase I;* Surber, Smith, and Frederika Harper, undated, University of Wisconsin-Milwaukee; *Technical Report No. 6, The Relationship Between Map Tests and Multiple Choice Tests,* Surber, Smith and Harper, 1982, University of Wisconsin-Milwaukee; and *Mapping as a Testing and Diagnostic Device,* Surber: Spatial Learning Strategies, 1984, Academic Press, Inc., pp. 213-233 (also available as TC# 000.6MAPASA).

These reports and papers describe the development of map tests as an assessment technique to identify conceptual misunderstandings that occur when students learn from text. The purpose is to diagnose student understanding in order to plan instruction. In this testing technique, the test developer graphically represents concepts and their interrelationships in a map. Then, information from the map is systematically removed. Students complete the map shells. Four different levels of deletion associated with different types of content clues are described. Maps are scored by comparing the student-completed version to the original. Scoring involves looking both at the content included or omitted from the map and the proper relationship between this content. Report #6 describes scoring in more detail.

The authors did a series of studies on this technique, reported on in "Mapping as a Testing and Diagnostic Device." They found good interrater reliability and good consistency between developers of "master maps." They report on comparisons to multiple-choice tests.

Text maps and tests can be constructed in any content area at any grade level. The specific examples in these materials come from chemistry (matter), study skills, and sociology (the development of early warfare).

A manual, designed to teach students how to construct concept maps, is included in Report #1. The authors have given educators permission to copy these documents for their own use.

(TC# 150.6MAPTES)

Szetela, Walter and Cynthia Nicol. *Evaluating Problem Solving in Mathematics.* **Located in: Educational Leadership, May 1992, pp. 42-45.**

This short article presents a statement of the need to assess problem solving, describes steps in the problem-solving process, shows some sample scoring guides, and discusses some question types that prompt problem solving. Scoring guides are somewhat sketchy and no samples of student work are included.

(TC# 500.6EVAPRS)

Vermont Department of Education. ***Vermont Mathematics Portfolio Project: Grade Eight Benchmarks*, 1991. Available from: Vermont Department of Education, Vermont Mathematics Portfolio Project, 120 State Street, Montpelier, VT 05602, (802) 828-3135.**

This document provides lots of samples of grade eight student work that illustrate different scores for each of the seven analytical traits used in the Vermont Mathematics Portfolio Project. Samples were taken from the 1991 portfolio pilot.

(TC# 500.3GRAEIB)

Vermont Department of Education. ***Vermont Mathematics Portfolio Project: Grade Four Benchmarks*, 1991. Available from: Vermont Department of Education, Vermont Mathematics Portfolio Project, 120 State Street, Montpelier, VT 05602, (802) 828-3135.**

This documents provides lots of samples of grade four student work that illustrate different scores for each of the seven analytical traits used in the Vermont Mathematics Portfolio Project. Samples were taken from the 1991 portfolio pilot.

(TC# 500.3GRAFOB)

Vermont Department of Education. ***Looking Beyond "The Answer"--The Report of Vermont's Mathematics Portfolio Assessment Program*, 1991. Available from: Vermont Department of Education, Vermont Mathematics Portfolio Project, 120 State Street, Montpelier, VT 05602, (802) 828-3135.**

This report describes the results of the pilot year of the Vermont's grade 4 and 8 mathematics portfolio system used for large-scale assessment. The report contains information on the rationale for the portfolio approach, a description of what students were to include, a description of the criteria used to evaluate the portfolios (with sample student performances to illustrate the scoring scale), the scoring and training process, results, and what was learned about large-scale assessment using portfolios.

(For related documents, see entries under "Koretz.")

(TC# 500.3REPOFV)

Vermont Department of Education. ***Vermont Mathematics Portfolio Project: Resource Book*****, 1991. Available from: Vermont Department of Education, Vermont Mathematics Portfolio Project, 120 State Street, Montpelier, VT 05602, (802) 828-3135.**

This document includes sample performance tasks taken from portfolio entries submitted by teachers as part of Vermont's 1991 math portfolio pilot project, a resource bibliography, and a list of suggested readings. The purpose is to provide colleagues with tasks that have worked well with students to promote problem solving. *This is meant as a companion document to the Teacher's Guide (TC# 500.3TEAGU1).*

(TC# 500.3RESBOO)

Vermont Department of Education. ***Vermont Mathematics Portfolio Project: Teacher's Guide*****, 1991. Available from: Vermont Department of Education, Vermont Mathematics Portfolio Project, 120 State Street, Montpelier, VT 05602, (802) 828-3135.**

This document presents Vermont's current view of what should go into a mathematics portfolio, provides detailed information about the scoring criteria for portfolio entries and the portfolio as a whole, discusses how to develop tasks that will invite student problem solving, and provides help with how to manage the portfolios. *This is a companion piece to the Resource Book (TC# 500.6RESBOO).*

(TC# 500.3TEAGU1)

Webb, Noreen. ***Alternative Strategies for Measuring Higher Order Thinking Skills in Mathematics: The Role of Symbol Systems*****, 1991. Available from: CRESST, University of California -- Los Angeles, 145 Moore Hall, Los Angeles, CA 90024. (213) 825-4711.**

This document presents an overview of a study that is currently taking place at CRESST in which students are asked to represent problems in various equivalent ways (graphs, tables, equations, word problems, and diagrams). The premise is that if a student really understands a problem, he or she should be able to solve the problem presented in any format, and translate from one format to another. Examples are provided of problems represented in different ways.

(TC# 500.6ALTSTF)

Webb, Norman, and Thomas A. Romberg. ***Implications of the NCTM Standards for Mathematics Assessment.*** **Located in <u>Mathematics Assessment and Evaluation: Imperatives for Mathematics Educators</u>, Thomas A. Romberg (Ed.), 1992. Available from: State University of New York Press, State University Plaza, Albany, NY 12246.**

This paper provides a good summary of the NCTM standards, both goals for students and standards for assessment. It uses four of the standards for assessment to develop criteria for assessments:

1. The assessment instrument should provide information that will contribute to decisions for the improvement of instruction

2. The assessment instruments should be aligned with the instructional goals, the goals for the overall program, and a holistic conceptualization of mathematical knowledge

3. The assessment instruments should provide information on what a students knows

4. The results from one assessment instrument should be such that when combined with results from other forms of assessment, a global description is obtained of what mathematics a person or group knows

The authors then illustrate their points with several assessment tasks that they feel would elicit the correct behavior from students. (These generally have only one correct answer and appear to be scored for degree of correctness.)

(TC# 500.6IMPNCM)

Wells, Barbara G. ***Journal Writing in the Mathematics Classroom.*** **Located in: <u>Communicator</u>, 15, 1, 1990, pp. 30-31. Also available from: California Mathematics Council, Ruth Hadley, 1414 South Wallis, Santa Maria, CA 93454, (805) 925-0774.**

This brief article describes one method that a teacher uses to elicit thinking on the part of high school math students. The teacher puts a short phrase on the board at the beginning of each class period and students write what they know about that phrase as the teacher takes attendance. Sample "prompts" and student responses are included. Although no criteria for evaluating responses are included, this article is added here because it represents an attempt to do writing in math, and because some of the prompts are designed to elicit metacognition, e.g., "What three problems on the final should have been eliminated and why?" or "What mathematical fact, concept, skill or insight that you learned in this class this year are you most likely to remember and why?"

(TC# 500.6JOUWRI)

Whetton, Chris. ***An Evaluation of the 1992 National Curriculum Assessment at Key Stage 1 in the Core Subjects*****, 1992. Available from: National Foundations for Educational Research (NFER), The Mere, Upton Park, Slough, Berks S11 2DQ, England, United Kingdom.**

This set of four documents reports on the results of the 1992 assessment. They contain: results of surveys of educators, use of the assessments with special education students, overall summary results, and recommendations for the 1993 assessment.

(TC# 060.6EVANAC)

Whetton, Chris. ***Key Stage 1, 1992, Teacher's Pack*, 1992. Available from: HMSO Publications Centre, PO Box 276, London, SW8 5DT, England, United Kingdom.**

This document contains all administration materials for the 1992 assessment. The assessments consist of a combination of hands-on and paper and pencil activities for primary students. English, science and mathematics are covered. In science and math, some activities are scored for the correctness of the answer and some are scored for correctness of approach or explanation. For example, one math task requires students to sort and tabulate the frequency of objects in a cupboard pictured in the student booklet. (Students get a "correct" mark if they miss no more than one item.) One science task requires students to select and describe five objects. (The response is "correct" if the student describes at least three objects in terms of at least two physical characteristics.)

All tasks are administered by the classroom teacher in large and small group settings. (The 1992 assessment took 24 hours, including English.) *A summary and technical report on the 1992 assessment is cataloged separately (Whetton: TC# 060.6EVANAC)*

(TC# 070.3KEY192)

Whetton, Chris. ***Key Stage 1, 1993, Teacher's Pack*, 1993. Available from: HMSO Publications Centre, PO Box 276, London SW8 5DT, England, United Kingdom.**

This document contains all administration materials for the 1993 assessment. The assessments consist of a combination of hands-on and paper and pencil activities for primary students. English, science and mathematics are covered. In science and math, some activities are scored for the correctness of the answer and some are scored for correctness of approach or explanation. For example, one math task consisted of adding and subtracting using a small number of objects. (The student must get three out of four correct to be scored as "pass.") One science task has students draw pictures or verbally explain what forces are acting on a raft as it floats on the water. (Responses are scored correct if the student conveys the knowledge that there are forces acting down and up on the raft.) Scoring is always tied directly to the task, and tasks usually are designed to cover discrete skills or pieces of knowledge.

All tasks are administered by the classroom teacher in large and small group settings. Results of the 1993 administration are not yet available, so it is unknown how long the most current version takes. (The 1993 assessment was greatly streamlined from the 1992 assessment which took 24 hours, including English.)

(TC# 070.3KEY193)

Whetton, Chris, Graham Ruddock, Steve Hopkins, et al. ***Standard Assessment Tasks for Key Stage 1*****, 1991. Available from: HMSO Publications Centre, PO Box 276, London SW8 5DT, England, United Kingdom.**

In spring 1991, all seven-year-olds in England and Wales (N=600,000) where tested using a set of performance assessments tied to a new National Curriculum. Areas tested included reading, writing, spelling, handwriting, math, and science. The assessment consisted of a series of tasks given to students. For each task, students were assessed on several "statements of attainment (SoA) [goals in the curriculum]." In math, thirty-eight SoA's were covered in 19 tasks. SoA's included those that are fairly traditional (e.g., "use addition and subtraction facts up to 10") but also included some self-reflection and problem solving (e.g., "talk about own work and ask questions," "make predictions based on experience," "explore and use the patterns in addition and subtraction facts to 10").

This package contains all the materials used by teachers for the age 7 *Standard Assessment Tasks*--administration handbooks, detailed description of tasks and scoring procedures, information recording booklets, and student worksheets. *For related information see other entries from Whetton.*

(TC# 070.3STAAST -- In house use only)

Wilson, Mark. ***Measuring Levels of Mathematical Understanding*****. Located in: <u>Mathematics Assessment and Evaluation: Imperatives for Mathematics Educators</u>, Thomas A. Romberg (Ed.), 1992. Available from: State University of New York Press, State University Plaza, Albany, NY 12246.**

The premise of this article is that if we want students to be reasoners and thinkers, we need to move from tests that fragment knowledge into "atomistic" pieces, each of which are assessed independently of the others, to assessment procedures that reveal student understanding of the concepts in a domain and their interrelationships. Many current tests are based on lists of skills, each of which is tested separately. "The primary focus of a mathematics testing methodology based on an active, constructive view of learning is on revealing how individual students view and think about key concepts in a subject. Rather than comparing students' responses with a 'correct' answer to a question so that each response can be scored right or wrong, the emphasis is on understanding the variety of responses that students make to a question and inferring from those responses students' levels of conceptual understanding."

The author presents a few examples. One is the SOLO taxonomy which looks at degree of formal reasoning. *(See the Collis-Romberg TC# 500.3COLROM on this bibliography.)*

This is a very technical and theoretical article and points up the need to be well grounded in current theory before beginning to develop math assessments.

(TC# 500.6MEALEM)

Zarinnia, E. Anne, and Thomas A. Romberg. ***A Framework for the California Assessment Program to Report Students' Achievement in Mathematics.*** **Located in: Mathematics Assessment and Evaluation: Imperatives for Mathematics Educators, Thomas A. Romberg (Ed.), 1992. Available from: State University of New York Press, State University Plaza, Albany, NY 12246.**

This paper takes the position that assessment affects instruction, and therefore, regardless of the other purposes for the assessment, the instructional implications of our assessments must be taken into account. "If one acknowledges student learning as the central mission of schooling, it further suggests that not only the tasks, but also the system and structures for gathering accountability information and reporting the data, should be designed with instructional needs in mind "

Other points made by this paper are:

1. We need to change the view of math held by many teachers and the general public, that math is a set of rules and formalisms invented by experts that everyone else is to memorize. The authors maintain that both the test itself and the way results are reported will influence these perceptions

2. Mathematical power means that citizens can use math to solve day-to-day problems. This means we need to seek evidence of students using, reflecting on, and inventing mathematics in the context of value and policy judgments. These experiences should be built into our instruction and assessments

Implications for turning power over to students are also discussed

(TC# 500.6FRACAA)

ASSESSMENT ALTERNATIVES IN SCIENCE

The following entries represent current Test Center holdings in the area of alternative assessment ideas for science. "Alternative," for this purpose, means "other than standardized, norm-referenced." The list emphasizes performance assessments, portfolios, technological innovations, etc. Some of the entries may be intended for informal, classroom use. For more information, contact Judy Arter, Senior Research Associate, or Matthew Whitaker, Test Center Clerk, at (503) 275-9582, Northwest Regional Educational Laboratory, 101 SW Main, Suite 500, Portland, Oregon 97204.

Abraham, Michael R., Eileen Bross Grzybowski, John W. Renner, and Edmund A. Marek. *Understandings and Misunderstandings of Eighth Graders of Five Chemistry Concepts Found in Textbooks.* **Located in: Journal of Research in Science Teaching, 29, 1992, pp. 105-120.**

The study reported in this paper looked at how well grade eight students understand five concepts in chemistry: chemical change, dissolution, conservation of atoms, periodicity, and phase-change. There are five problems, one associated with each concept. Each problem describes (and/or shows) a problem situation and asks one to three questions. Some questions require short answers and some require explanations of answers.

Each response is scored on a six-point scale from "no response" to "specific misunderstanding" to "sound understanding" of the concept. The paper gave some examples of misunderstandings shown by the students.

The authors found that very few students really understood the concepts. They speculate that this may either be due to the nature of instruction (mostly textbook driven and little hands-on) or because students are not developmentally ready for the formal logic found in these concepts.

The paper reports some information on student status and the relationship between scores on this test and another measure of formal logical thinking.

A related study using the same five tasks is Michael R. Abraham, Vickie M. Williamson, and Susam Westbrook, *A Cross-Age Study of the Understanding of Five Chemistry Concepts.* Available from: The Department of Chemistry and Biochemistry, University of Oklahoma, 620 Parrington Rd., N[illegible]man, OK 73019 (TC#600.3CROAGS)

(TC#650.3UNDMIE)

Appalachia Educational Laboratory. *Alternative Assessments in Math and Science: Moving Toward a Moving Target,* 1992. Available from: Appalachia Educational Laboratory, PO Box 1348, Charleston, WV 25325, (304) 347-0400.

This document reports on a two-year study by the Virginia Education Association and the Appalachia Educational Laboratory. In the study, 11 pairs of K-12 science and math teachers designed and implemented new methods of evaluating student competence and application of knowledge.

Teachers who participated in the study found that the changes in assessment methods led to changes in their teaching methods, improvements in student learning and better student attitudes. Instruction became more integrated across subjects and shifted from being teacher-driven to being student-driven. Teachers acted more as facilitators of learning rather than dispensers of information.

Included in the report is a list of recommendations for implementing alternative assessments, a list of criteria for effective assessment, and 22 sample activities (with objectives, tasks, and scoring guidelines) for elementary, middle, and high school students, all designed and tested by the teachers in the study.

Most activities have performance criteria that are holistic and specific to each exercise. No technical information or sample student work is included.

(TC#600.3ALTASM)

Barnes, Lehman W., and Marianne B. Barnes. *Assessment, Practically Speaking.* Located in: <u>Science and Children</u>, March 1991, pp. 14- 15.

The authors describe the rationale for performance assessment in science. Traditional tests (vocabulary, labeling, matching, multiple-choice, short-answer, puzzle, questions, essay) accurately assess student mastery of the verbal aspects of science. But, they do not allow students to demonstrate what they know.

(TC#600.6ASSPRS)

Baron, Joan B. *Performance Assessment: Blurring the Edges Among Assessment, Curriculum, and Instruction*, **1990. Located in: Champagne, Lovitts and Calinger (Eds.), Assessment in the Service of Instruction, pp. 127-148. Available from: American Association for the Advancement of Science, 1333 H St. NW, Washington, DC 20005, [AAAS Books: (301) 645-5643]. Also in: G. Kulm & S. Malcom (Eds.), Science Assessment in the Service of Reform, AAAS, 1991, pp. 247-266.**

After a brief discussion of the rationale for doing performance assessments in science, this article describes current work in Connecticut. The tasks for these assessments have three parts that involve a blend of individual work at the beginning and end, and group work in the middle:

1. At the beginning, each student provides information about his or her prior knowledge and understandings of the scientific concepts and processes relevant to the task. The student also provides a preliminary solution to the task. This serves to encourage preliminary thinking, brings diversity to the thinking of the group, makes more obvious what each student brings to the task, has instructional value, and provides a baseline for students to refer to later.

2. Then, students work as a team to produce a group product. Throughout this process individual students report their views/summaries/insights of the work of the group.

3. After the group work, a transfer task is completed individually.

The paper then spends some time discussing how to structure the tasks used in such assessments, and the learning theory and collaborative learning research that underpin the approach.

The paper concludes with a discussion of current issues in performance assessment in science including:

- They take a lot of time.
- The concepts assessed are harder to teach and harder for students to grasp.
- Teachers are concerned about covering the material that is required in course guides.
- It requires a great deal of expertise on the part of the teacher.

This article does not discuss the criteria by which student performances would be evaluated. However, the author discussed preliminary plans at a session at the Annual Meeting of the American Educational Research Association in April 1991. Scoring plans include: the content of solutions, the processes used to arrive at solutions, interpersonal and communication skills used by the group, and the manifestation of science-related attitudes.

At this session, the author also discussed the fact that they have developed 50 tasks, and are currently doing research and development on them. This includes:

1. Getting expert opinion on the degree to which the tasks invite the skills and behaviors that are to be observed.

2. Having experts and novices come up with criteria by which to score performances, and modifying these by looking at actual student work.

3. Exploring what happens in the classroom after this type of assessment is implemented.

4. Looking at potential bias in these tasks. Do all students have an equal opportunity to show what they can do?

(TC#600.6PERASB)

Bennett, Dorothy. *Assessment & Technology Videotape*. Available from: The Center for Technology in Education, Bank Street College of Education, 610 W. 112th St., New York, NY 10025, (212) 875-4550.

The Center for Technology in Education (CTE) has been conducting research on how best to use technology in assessment. It supports the use of video to capture aspects of students' performance that cannot be assessed with paper and pencil.

This document consists of a video and handbook that focus on the assessment of thinking skills, communication skills and interpersonal skills in the context of a group project that requires applying physics to the design of motorized devices which produce at least two simultaneous motions in different directions to accomplish an action or set of actions.

The first part of the video describes an alternative assessment system that uses students' personal journals, group logs, projects, and presentations. Personal journals document students' personal experiences with technology outside the classroom and their observations about how things work. Group logs document group problem-solving and dynamics. The group projects and presentations are the major part of the assessment. Presentations are videotaped and scored by a panel of experts and other students.

The second part of the video contains four examples of students' presentations (car wash, tank, garbage truck, oscillating fan) which can be used to practice scoring using the criteria set forth in the handbook. The basic criteria for assessing students' presentations are:

1. Thinking skills:

   - Understanding
   - Critical thinking/meta-processing
   - Extensions of knowledge and inquiry/creativity

2. Communication/presentation skills

   - Clarity and coherence of presentation
   - Presentation aesthetics

3. Work management/interpersonal skills

   - Teamwork (for group work only)
   - Thoroughness and effort
   - Reflectiveness

Brief descriptions of the above criteria are contained in the handbook. The procedure is a prototype. Feedback by those attempting to use the criteria is requested.

(TC#600.3ASSTEVh and 600.3ASSTEVv)

Brown, Larry. ***Portfolios in Rural High School Mathematics and Science Classes*****, 1992. Available from: Cusick High School, Cusick, WA 99119, (509) 445-1125.**

This project is still in the developmental process, but is intended to develop the concept that the portfolio is a student's self-selected, self-reflective documentation of growth in understanding and skill over the course of a school year. Students will prepare their portfolios across the curriculum areas of advanced mathematics and physics. Results of the project will be presented together with recommendations for improvement and implications for future work to the Cusick School District, participants of SMART (NWREL), and at the Small Schools Conference at Central Washington University on March 19, 1993.

The author only provided a description of his project. Additional information is available only from the author.

(TC#660.6PORRUH)

California Assessment Program (CAP). ***New Directions in CAP Science Assessment*****, 1990. Available from: California Department of Education, PO Box 944272, Sacramento, CA 94244, (916) 445-1260.**

The new California science curriculum identifies six themes (energy, evolution, patterns of change, scale/structure, stability and systems/interactions) that cut across three content areas (earth sciences, physical sciences, and life sciences.) CAP is developing multiple-choice, open-ended, and performance items to match this curriculum.

CAP administered open-ended science questions to 8,000 sixth graders in spring, 1989. The questions required students to create hypotheses, design investigations, and write about social and ethical issues in science. Each task took 10 to 15 minutes.

CAP also field tested five performance assessment tasks to about 50,000 sixth graders in spring, 1990. Tasks were administered at five stations and took about 10 minutes each. The tasks were:

1. Building a circuit, and predicting, testing, and recording the conductivity of various materials.

2. Creating a classification system for a collection of leaves, and explaining the adjustments necessary when a "mystery leaf" is introduced into the group.

3. Performing a number of tests on a collection of rocks, and recording and classifying the results.

4. Estimating and measuring water volumes.

5. Performing chemical tests on samples of lake water.

This document includes instructions for administering one of the performance tasks (electricity), seven letters written by students commenting on the assessment, and two open-ended questions with sample student responses.

Grade 12 students were supposed to have similar pilot-testing during fall, 1990. CAP has plans to use performance and open-ended tasks in their 1991 assessment for grade 6 and 1992 assessment for grade 12.

(TC#600.3NEWDII)

Champagne, Audrey, B. Lovitts, and B. Calinger. ***Assessment in the Service of Instruction*****, 1990. Available from: American Association for the Advancement of Science, 1333 H St. NW, Washington, DC 20005 [AAAS Books: (301) 645-5643].**

This book is a compilation of eleven papers that address the issue of making assessment a tool for meaningful reform of school science. The book contains papers that cover: an overview of good assessment, national and state assessment initiatives, traditional assessments, innovative assessments (performance, group, portfolio, and dynamic), and experiences in England and Wales.

The introductory article by two of the editors *(Assessment and Instruction: Two Sides of the Same Coin)* covers the following topics:

1. Reasons for assessing, including instruction, conveying expectations, monitoring achievement, accountability and program improvement.

2. What should be assessed, and the inability of multiple-choice tests to assess the most important aspects of scientific competence: generating and testing hypotheses, designing and conducting experiments, solving multi-step problems, recording observations, structuring arguments, and communicating results; or scientific attitudes: comfort with ambiguity and acceptance of the tentative nature of science.

3. A definition of "authentic" assessment: "An assessment is authentic only if it asks students to demonstrate knowledge and skills characteristic of a practicing scientist or of the scientifically literate citizen." Simply matching the curriculum is not enough, because the curriculum may be lacking.

Other articles from this book that are particularly relevant to this bibliography are described separately.

(TC#600.6ASSINT)

Chi, M.T., P.J. Feltovich, and R. Glaser. ***Categorization and Representation of Physics Problems by Experts and Novices*****, 1981. Located in: <u>Cognitive Science</u>, 5, pp. 121-152.**

The authors report on a series of studies to determine the differences between expert and novice problem solvers in physics. Although this paper is not about assessment per se, the observations in the paper might help users to define what good physics problem solving looks like, which in turn can serve as the basis for forming performance criteria to be used with performance assessments.

Expert problem solvers begin with a brief analysis of the problem statement to categorize the problem (i.e., determine which schema to activate). Once activated, the schema itself specifies further tests for its appropriateness. When the expert has decided that a particular principle is, indeed, appropriate, the knowledge contained in the schema provides the general form that specific equations to be used for solution will take. This is contrasted to novice problem solvers which use superficial characteristics to categorize problems and lack procedural connections.

Several samples of expert and novice thinking are provided.

(TC#660.6CATREP)

Coalition of Essential Schools. Various articles on exhibitions of mastery and setting standards, 1982-1992. Available from: Coalition of Essential Schools, Brown University, Box 1969, One Davol Sq., Providence, RI 02912, (401) 863-3384.

Although not strictly about science, this series of articles discusses performance assessment topics and goals for students that are of relevance to science. The articles are: Rethinking Standards; Performances and Exhibitions: The Demonstration of Mastery; Exhibitions: Facing Outward, Pointing Inward; Steps in Planning Backwards; Anatomy of an Exhibition; and The Process of Planning Backwards.

These articles touch on the following topics: good assessment tasks to give students, the need for good performance criteria, the need to have clear targets for students that are then translated into instruction and assessment, definition and examples of performance assessments, brief descriptions of some cross-disciplinary tasks, the value in planning performance assessments, and the notion of planning backwards (creating a vision for a high school graduate, taking stock of current efforts to fulfill this vision, and then planning backward throughout K-12 to make sure that we are getting students ready from the start).

(TC#150.6VARARD)

Colison, J. *Connecticut's Common Core of Learning*, 1990. Available from: Performance Assessment Project, Connecticut Department of Education, Box 2219, Hartford, CT 06145, (203) 566-4001.

The Connecticut Department of Education is developing a series of performance assessments in science and math. Each task has three parts: individual work to activate previous knowledge; group work to plan and carry out the task; and individual work to check for application of learning. This document provides:

1. A lengthy description of one of the ninth grade science tasks: "speeders."

2. Short descriptions of 24 performance tasks in science (8 each in chemistry, physics, and earth sciences), and 18 in math.

3. A group discussion self-evaluation form to be used by students.

No technical information or general scoring guides are included in this document. Additional information should be forthcoming soon.

(TC#600.3CONSCI)

Collins, Angelo. *Portfolios for Assessing Student Learning in Science: A New Name for a Familiar Idea?*, 1990. Located in: Champagne, Lovitts and Calinger (Eds.), Assessment In The Service of Instruction, pp. 157-166. Available from: American Association for the Advancement of Science, 1333 H St. NW, Washington, DC 20005 [AAAS Books: (301) 645-5643]. Also in: G. Kulm and S. Malcom (Eds.), Science Assessment in the Service of Reform, AAAS, 1991, pp. 291-300.

This paper presents the rationale for using portfolios in science, defines and provides the characteristics of such portfolios, and discusses what should go in them. There is no one "right" way to do a portfolio. They will differ due to three factors--purpose, context and design.

Purpose includes what is to be shown with the portfolio--mastery of content? understanding of and use of the processes by which this knowledge is constructed? student attitudes toward science? student comfort with ambiguity and the tentative nature of science? Purpose also includes how the portfolio will be used--student self-reflection? accountability? instruction?

Context includes such things as the age of the students and student interests and needs.

Design covers such considerations as what will count as evidence, how much evidence is needed, how the evidence will be organized, who will decide what evidence to include, and evaluation criteria.

This article focused mostly on considerations when designing a portfolio system in science, but a few, brief examples are given.

(TC#600.6PORFOA)

Collins, Angelo. *Portfolios: Questions for Design.* Located in: Science Scope, 15, March 1992, pp. 25-27.

This repeats a lot of the information on this topic presented by the author on other entries on this bibliography. This is a nice, short summary. The author appears to use the term "purpose" (as in "determine the purpose for the portfolio") to mean "target" (what do we want to show about the student).

(TC#600.3PORQUD)

Collins, Angelo. *Portfolios for Science Education: Issues in Purpose, Structure, and Authenticity.* Located in Science Education, 76, 1992, pp. 451-463.

The author teaches preservice science teachers. This paper discusses design considerations for portfolios in science and applies these considerations to portfolios for student science teachers, practicing science teachers, and elementary students. The design considerations he suggests are:

1. Determine what the portfolio should be evidence of. What will the portfolio be used to show?
2. Determine what types of displays should go in the portfolio to provide evidence of #1. He suggests and describes several types: artifacts (actual work produced), reproductions of events (e.g., photos, videotapes), attestations (documents about

the work of the person prepared by someone else), and productions (documents prepared especially for the portfolio such as self-reflections).

3. View the portfolio as a "collection of evidence" that is used to build the case for what is to be shown. Those developing the portfolio should determine the story to be told (based on all the evidence available) and then lay this out in the portfolio so that it is clear that the story told is the correct one.

(TC#600.6PORSCE)

Collins, Allan, Jan Hawkins, and John R. Frederiksen. ***Three Different Views of Students: The Role of Technology in Assessing Student Performance, Technical Report No. 12,*** **April 1991. Available from: Center for Technology in Education, Bank Street College of Education, 610 W. 112th St., New York, NY 10025, (212) 875-4550. Also available from ERIC: ED 337 150.**

This paper begins by discussing why assessment in science needs to change: if tests continue to emphasize facts and limited applications of facts the curriculum will be narrowed to these goals. The paper then gives several good examples of how high stakes uses of tests have negative, unintended side effects on curriculum and instruction.

The authors use the term "systemically valid" to refer to assessments that are designed to foster (create) the learning they also assess. The authors discuss four criteria for "systemically valid" tests (the test directly measures the attribute of interest, all relevant attributes are assessed, there is high reliability, and those being assessed understand the criteria), criteria for quality tasks, examples of alternative assessment ideas, cost, cheating, and privacy.

(TC#600.6THRDIV)

Dana, Thomas M., Anthony W. Lorsbach, Karl Hook, and Carol Briscoe. ***Students Showing What They Know: A Look at Alternative Assessments,*** **1991. Located in: G. Kulm and S. Malcolm (Eds.), Science Assessment in the Service of Reform, pp. 331-337. Available from: American Association for the Advancement of Science, 1333 H St. NW, Washington, DC 20005 [AAAS Books: (301) 645-5643].**

The authors present short descriptions of assessment activities they have developed and used with students at the Florida State University School for grades 6-12 in physical science, biology, and chemistry. The assessments are based on the theory that students construct knowledge for themselves as they participate in educational activities. The authors briefly mention the following techniques: concept mapping, journals, scrap books, and oral interviews. The examples include mostly descriptions of tasks; there is mention, but not elaboration, of the criteria for judging responses. The techniques emphasize student self-evaluation.

(TC#600.3STUSHW)

Doran, R. ***Performance Assessment in Science at the 12th Grade Level,*** **1991. Available from: Graduate School of Education, University of New York at Buffalo, Buffalo, NY 14260, (716) 636-2000.**

This document is an outline of a series of activities to assess student laboratory skills. The outline includes the list of tasks that students must perform for each lab, a list of the labs, the traits that are scored for each performance, and a set of references to previous efforts in assessing laboratory skills. Although just an outline, the document is of interest because of the scoring guidelines. Each performance is scored on two dimensions having five traits each:

1. Experiment Planning and Design--Statement of hypothesis, procedures for investigation, diagram of equipment, safety procedures, and collection of observations/measurements.

2. Laboratory Performance and Analysis of Results--Organization of observations/data, accuracy of observations/data, calculation of means, presentation of data on graph, and statement of conclusions.

This document does not include detailed information on tasks, performance criteria, sample student performances. No technical information is included. A request late in 1992 yielded no response. The author previously reported that complete copies of the instruments would be available in 1992.

(TC#600.3PERASI)

Halpern, Diane (Ed.). *Enhancing Thinking Skills in the Sciences and in Mathematics*, 1992. **Available from: Lawrence Erlbaum Associates, Publishers, 365 Broadway, Hillsdale, NJ 07642, (800) 926-6579.**

This book is not strictly about assessment. Rather, it discusses the related topics of "What should we teach students to do?" and "How do we do it?" The seven authors "criticize the conventional approach to teaching science and math, which emphasizes the transmission of factual information and rote procedures applied to inappropriate problems, allows little opportunity for students to engage in scientific or mathematical thinking, and produces inert knowledge and thinking skills limited to a narrow range of academic problems." (p. 118). In general, they recommend that teachers focus on the knowledge structures that students should know, use real tasks, and set up instruction that requires active intellectual engagement.

The authors give various suggestions on how to bring this about: instructional methods, videodiscs, group work, and a host more. The final chapter analyzes the various positions and raises theoretical issues.

(TC#500.6ENHTHS)

Hardy, Roy. *Options for Scoring Performance Assessment Tasks*. **A presentation to the National Council on Measurement in Education, San Francisco, California, April 23, 1992. Available from: Educational Testing Service, 1979 Lakeside Parkway, Suite 400, Tucker, GA 30084**

Four assessment tasks were developed to explore the feasibility of performance assessment as part of a statewide assessment program. Tasks were: shades of color (grades 1-2), discovering shadows (grades 3-4), identifying minerals (grades 5-6), and designing a carton (grades 7-8). The tasks are described in the paper, but all relevant materials are not included. Each task was designed to take one hour. Most tasks are completed individually, but one (cartons) is a group task.

Response modes were varied (multiple-choice, figural, short narratives, products), in part to see which are feasible, and in part to see how different kinds of scores relate to each other. Most scoring was right/wrong or holistic on degree of "correctness" of answer.

Cartons was scored holistically on problem solving. The scoring procedures are described but not presented in detail. The paper also describes the process used to develop scoring rubrics, to train scorers at the state level, and to analyze the data. No sample student responses are included in this document, but were used in training.

The tasks were completed by 1128 students in 66 classes in 10 school districts. Teachers completed a survey (questions are included in the paper). Results showed that it took from 1/2 to three minutes to score the performances, interrater agreement ranged from .76 to the high .90's, relationships between scoring procedures varied, and teachers liked the procedures. In all, the author concluded that it is feasible to use performance tasks in statewide assessment.

(TC#600.3OPTSCP)

Harlen, Wynne. ***Performance Testing and Science Education in England and Wales.*** **Located in: Gerald Kulm and Shirley M. Malcom (Eds.), Science Assessment in the Service of Reform, 1991. Available from: American Association for the Advancement of Science, 1333 H St. NW, Washington, DC 20005 [AAAS Books: (301) 645-5643].**

This is a good summary of the approach to science education and assessment currently under way in England and Wales. (For related information, see the entries under Whetton.) It discusses the history of the project, provides three hands-on test questions as examples, and describes the issues and problems which have arisen thus far--for example, comparability of tasks, amount of reading required by students, and trying to accomplish too many purposes with a single assessment.

From the examples provided, it appears that the performance tasks are a series of open-response questions which address a single science process skill, e.g., interpreting information, planning an investigation, or observing. Students provide short-answers which are evaluated according to degrees of completeness or right/wrong. Criteria differ by task.

(TC#600.6PERTES)

Hibbard, Mike. ***What's Happening?*****, 1991. Available from: Region 15 School District, PO Box 395, Middlebury, CT 06762, (203) 758-8250.**

This document is a series of performance tasks in which assessment is integrated with instruction. The tasks include chemical reaction, consumer action research, plant growth, physiological responses of the human body, survival in the winter, science fiction movie development, and food webs. Each task includes assessment rating forms and checklists, some of which are designed for student self-assessment. For example, the *survival in winter* exercise includes a rating scale that assesses 12 features of the project on a scale of 1-5, and a rating scale for an oral presentation. Other tasks include performance criteria for group work and self-rating on perseverance. The P.C. are a mixed bag. Some directly refer to specific features of the task (e.g., "detailed descriptions were given of each plants' growth"). Others are general features that could be applied to many tasks (e.g., "shows persistence"). However, there is no standard of criteria across tasks: there is a different number of criteria and a different mix of specific and general criteria depending on task.

The assessments were developed for classroom use and do not include detailed definitions of traits to be rated, nor sample anchor performances. No technical information is included.

(TC#600.3WHAHAP)

Johnson, David W. and Roger T. Johnson. ***Group Assessment as an Aid to Science Instruction*****, 1990. Located in: Champagne, Lovitts and Calinger (Eds.), Assessment in the Service of Instruction, pp. 267-282. Available from: American Association for the Advancement of Science, 1333 H St. NW, Washington, DC 20005 [AAAS Books: (301) 645-5643]. Also located in: G. Kulm & S. Malcom (Eds.), Science Assessment in the Service of Reform, AAAS, 1991, pp. 103-126.**

The authors favor cooperative learning in science because of research that shows positive effects on student learning and attitudes. Their suggestions for group assessment build on this same philosophy--group assessment involves having students complete a lesson, project, or test in small groups while a teacher measures their level of performance. If done well, this format allows assessment of outcomes that are difficult to assess in other ways, such as reasoning processes, problem solving, metacognitive thinking, and group interactions. The authors also maintain that it increases the learning it is designed to measure, promotes positive attitudes toward science, parallels instruction, and reinforces the value of cooperation. The article describes how to structure performance tasks in a cooperative framework. .

The authors then describe, in general, different ways to record the information from the task--observational records, interviews, individual and group tests, etc. This is a general overview of the possibilities, however, and provides no specific rubrics, forms, questions, etc.

(TC#600.6GROASA)

Kamen, Michael. ***Use of Creative Drama to Evaluate Elementary School Students' Understanding of Science Concepts*****, 1991. Located in: G. Kulm and S. Malcom (Eds.), Science Assessment in the Service of Reform, pp. 338-341. Available from: American Association for the Advancement of Science, 1333 H St. NW, Washington, DC 20005 [AAAS Books: (301) 645-5643].**

This article emphasizes kinesthetic learning--reinforcing and assessing knowledge of scientific concepts through acting them out. For example, students can demonstrate their knowledge of waves by forming a line and creating waves with different wave length and amplitude. Other examples are given for air pressure, solar energy, and land snails. The assessment appears to occur by seeing the extent to which students can illustrate the concept properly. No other performance criteria are discussed. Tasks were designed for students in grades K-6.

(TC#600.6USECRD)

Kanis, I.B. ***Ninth Grade Lab Skills*****. Located in: The Science Teacher, January 1991, pp. 29-33.**

This paper provides a summary description of the six performance tasks given to ninth grade students as part of the 1985-86 Second International Science Study to assess laboratory skills. A brief description, a picture of the lab layout, and a list of scoring dimensions is provided for each task. It appears that scoring is essentially right/wrong and tied to each task. Students were scored on ability to manipulate material, collect information, and interpret results. A brief discussion of some results of the assessment are provided. There is enough information here to try out the tasks, but not enough to use the performance criteria. No sample student performances are included. The paper also discusses problems with many current lab activities (too cookbook) and how to redesign lab exercises to promote higher-order thinking skills.

(TC#600.3NINGRL)

Karnes, F.A. and S. Bean. *Process Skills Rating Scales*, 1990. Available from: United Educational Services, Inc., PO Box 1099, Buffalo, NY 14224, (716) 668-7691.

The *Process Skills Rating Scales* were designed for use in grades 1-12 to obtain ratings of students' facility in using process skills that relate to their ability to think, reason, search for knowledge independently, and communicate with others. Students may use the scales to rate themselves, or teachers and parents may use them to rate the student.

There are 12 rating scales altogether. Although several might be relevant to science, one is specifically called *Scientific Research Skills*, which contains 94 ratings. Examples are: "Can develop inferences from observation," and "Can focus on essentials."

No technical information is provided. There are no detailed definitions of the various skills to be rated and no sample student performances to help anchor the rating scales.

(TC#050.3PROSKR)

Kentucky Department of Education. *Kentucky Instructional Results Information System*, 1991-92. Available from: Advanced Systems in Measurement & Evaluation, Inc., PO Box 1217, 171 Watson Rd., Dover, NH 03820, (603) 749-9102. Also available from: Kentucky Department of Education, Capitol Plaza Tower, 500 Mero St., Frankfurt, KY 40601, (502) 564-4394.

This document contains only the released sets of exercises and related scoring guides from Kentucky's 1991-92 grade 4, 8, and 12 open-response tests in reading, math, science, and social studies. It does not contain any support materials such as: rationale, history, technical information, etc.

There are three to five tasks/exercises at each grade level in each subject. Most are open-response (only one right answer), but some are open-ended (more than one right answer). Examples in math are: write a word problem that requires certain computations, determine how many cubes are needed for a given figure, follow instructions, explain an answer, arrange a room, explain a graph. Examples in science are: experimental design for spot remover, graph and interpret results of a study on siblings, and predict the weather from a weather map. Scoring for each exercise is holistic/primary trait. Each exercise has it's own set of scoring criteria. It appears that the scoring emphasizes the correctness of the response and not the process by which the response was obtained.

(TC#060.3KENINR)

Koballa, T.R. ***Goals of Science Education*****, 1989. Located in: D. Holdzkom and P. Lutz (Eds.), Research Within Reach: Science Education, pp. 25-40. Available from: National Science Teachers Association, Special Publications Department, 1742 Connecticut Ave. NW, Washington, DC 20009, (202) 328-5800.**

Assessment should be designed to cover important student processes and outcomes. This article is included because it discusses what our goals for students should be. Specifically, the author maintains that most science curricula are oriented toward those students that want to pursue science academically and professionally. We should also, however, be looking at science education as a means of promoting other important goals for students such as: longing to know and understand, respect for logic, and helping students to acquire capacities to cope with change.

(TC#600.5GOASCE)

Kulm, Gerald, Shirley M. Malcom. ***Science Assessment in the Service of Reform*****, 1991. Available from: American Association for the Advancement of Science, 1333 H St. NW, Washington, DC 20005 [AAAS Books: (301) 645-5643].**

This book contains articles from various authors who discuss: current issues surrounding science assessment, the rationale for considering alternatives, curriculum issues and trends, and alternative assessment initiatives in various states and countries. There are good summaries of what is occurring with the *National Assessment of Educational Progress*, with test publishers, in England and Wales, and with various states. An appendix presents brief descriptions of alternative assessments under development by various organizations. The individual articles that appeared to be of most interest for the purpose of this bibliography are entered separately.

(TC#600.6SCIASI)

Laboratory Leadership Group. ***Laboratory Assessment Builds Success*****, 1990. Available from: Institute for Chemical Education, University of Wisconsin-Madison, Department of Chemistry, 1101 University Ave., Madison, WI 53706, (608) 262-3033.**

This document contains 18 chemistry labs for high school students. They were developed over a period of three years to make important concepts of chemistry very understandable to high school students. It also contains a rationale for laboratory-based learning and describes three kinds of outcomes important for labs--psychomotor, affective, and cognitive. The labs appear to be well developed. Student responses are assessed by correctness of final answers and ability to explain concepts and apply them to new situations. Although each exercise is classified according to the knowledge, critical thinking skills, and other skills it requires, only the knowledge type of skill is assessed in the student work. For the other outcomes, the only assistance given is a list of 14 things to look for as students do the labs (e.g., following instructions, selecting appropriate apparatus, interpreting results, communicating plans and results, etc.) without any help on how to do this.

No technical information is available.

(TC#650.3LABASB)

Lawrence Hall of Science. *Full Option Science System--Water Module*, 1992. **Available from: Encyclopedia Britannica Educational Corporation, 310 S. Michigan Ave., Chicago, IL 60604, (800) 534-9862. Also available from: Lawrence Hall of Science, University of California, Berkeley, CA 94720, (510) 642-8941.**

The Full Option Science System is a series of hands-on instructional modules with associated assessments. The modul reported here is on water. There are three parts to the assessment, all of which are describ d in detail in the document. The first part is a series of hands-on tasks set up in stations. Examples are: "Put three drops of mystery liquids on wax paper and observe what happens." and "What do your observations tell you about the mystery liquids?" The answer key indicates that scoring proceeds by looking at the correctness of the response. Two different testing configurations are outlined (8 students and 24 students). Each group takes about 30 minutes.

The second part of the assessment is an open-response paper and pencil test that takes about 15 minutes. Again, it appears that responses are scored for degree of correctness. The third part of the assessment is an application of concepts in paper and pencil format that takes about 20 minutes. Again, it appears to be scored by degree of correctness.

All administration and scoring information is provided, but no technical information on the tests nor information about typical performance is given.

(TC#660.3FOSSWM)

Liftig, Inez Fugate, Bob Liftig, and Karen Eaker. ***Making Assessment Work: What Teachers Should Know Before They Try It.*** **Located In: Science Scope, 15, March 1992, pp. 4-8.**

The authors contend that students have trouble taking alternative assessments because they have no practice doing do. For example, they don't know the higher-order thinking skills vocabulary that is often used in performance tasks, so they don't know what to do. They also don't know what it takes to do well. The authors recommend learning vocabulary, practicing oral and written communication, and being careful not to leave anything out because you figure that the teacher already knows you know it. A list of vocabulary is included.

(TC#600.6MAKASW)

Lock, Roger. *Gender and Practical Skill Performance in Science*, 1992. **Located in: Journal of Research in Science Teaching, 29, pp. 227-241.**

This paper is not included here because of the results of the study of student gender differences in high school students. Rather, it is included because of its brief descriptions of the performance tasks used, procedures, and method of scoring student performances. The four tasks were: measuring the rate of movement of blow fly larvae in dry and damp atmospheres, finding out how the size of the container with which a burning candle is covered affects the length of time for which the candle burns, determining the mass supported by a drinking straw, and identifying an unknown solution. Only one of these (straws) is described in enough detail to replicate. There are separate performance criteria for each task. Student performance is assessed live by listening to what the student says while he or she does the task, by watching what the student does, and by looking at what the student writes down. The criteria for the unknown solution task are given.

Because of the nature of the research reported, some technical information is included on the tasks. An attempt to obtain more information from the author was unsuccessful.

(TC#600.6GENPRS)

Lunetta, Vincent N., and Pinchas Tamir. ***Matching Lab Activities.*** **Located in: <u>The Science Teacher</u>, 46, May 1979, pp. 23-25.**

The authors list 24 skills and behaviors related to the scientific process, recommend using these skills to analyze the tasks given to students to make sure that students are being required to apply/use all the skills of importance, and report on a study in which they analyzed several tasks using the list. They discovered that most lab activities do not require students to use many of the skills on the list.

(TC#600.6MATLAA)

Lunetta, Vincent N., A. Hofstein, and G. Giddings. ***Evaluating Science Laboratory Skills.*** **Located in: <u>The Science Teacher</u>, January 1981, pp. 22-25.**

The authors present the following information:

1. A listing of the most common pedagogical objectives for laboratory work. These fall into three categories--cognitive, practical, and affective.

2. A summary of the advantages and disadvantages of four methods of assessing lab skills--written reports (e.g., lab notebooks), test items, structured performance assessment, and observation during regular classroom activities. The authors recommend the last.

3. Criteria to rate student performance during classroom observation. Criteria are given for five dimensions of performance--planning an experiment, conducting an experiment, data collection, interpreting results, and work habits. These were adapted from work in England in 1979, and are not completely reproduced in this paper.

(TC#600.6EVASCL)

Macdonald Educational. ***Learning Through Science*****, 1989. Available from: Macdonald Educational, Wolsey House, Wolsey Road, Hemel Hempstead HP2 4SS, England, UK. Also available from: Teachers' Laboratory, Inc., PO Box 6480, Brattleboro, VT 05301, (802) 254-3457.**

This is one of a series of publications developed to promote instructional reform in science in the British Isles. The reform movement emphasizes active learning and concept development. (An overview of this curriculum reform movement is included in 600.6SCIASI--W. Harlen, *Performance Testing and Science Education in England and Wales.*)

In addition to sections covering such topics as "why do science" and how to organize instruction, one chapter covers record keeping. This chapter proposes keeping track of student development toward mastery of broad scientific concepts and habits of thought rather than keeping track of activities completed. The chapter provides a brief description of a rating procedure (presented in more detail in another publication) for 24 attributes such as: curiosity, perseverance, observing, problem solving, exploring, classifying, area, and time. A sample five-point rating scale for one of the attributes, curiosity, is given.

An appendix to the book also provides developmental continua for: attitudes, exploring observations, logical thinking, devising experiments, acquiring knowledge, communicating, appreciating relationships, and critical interpretation of findings. These could be adapted for use in keeping track of student progress in a developmental fashion.

(TC#600.6LEATHS)

Macdonald Educational. ***With Objectives in Mind*, 1984. Available from: Macdonald Educational, Wolsey House, Wolsey Road, Hemel Hempstead HP2 4SS, England, UK. Also available from: Teachers' Laboratory, Inc., PO Box 6480, Brattleboro, VT 05301, (802) 254-3457.**

This is one of a series of publications developed to promote instructional reform in science in the British Isles. This instructional reform emphasizes active learning and concept development. (An overview of this curriculum reform movement is included in 600.6SCIASI--W. Harlen, *Performance Testing and Science Education in England and Wales.*)

This document covers topics such as the contribution of science to early education, objectives for children learning science, and how to use the various instructional units that have also been produced as part of this series. There is a good discussion of how student understanding in science develops, which includes many samples of student behavior as illustrations of the various stages. This discussion could be adapted to constructing developmental continua for tracking student progress to be used for performance assessment.

(TC#600.6WITOBM)

Marshall, G. *Evaluation of Student Progress*, 1989. Located in: D. Holdzkom and P. Lutz (Eds.), Research Within Reach: Science Education, pp. 59-78. Available from: National Science Teachers Association, Special Publications Department, 1742 Connecticut Ave. NW, Washington, DC 20009, (202) 328-5800.

This paper presents a general overview of assessment development targeted at classroom teachers. The author emphasizes the need to clearly define outcomes for students and then match the outcome to the proper assessment technique--multiple-choice, essay, projects, practical tests and lab reports. Examples of each item type (using science content) are provided.

(TC#600.6EVASTP)

Martinez, M. *Figural Response in Science and Technology Testing*, 1991. Located in: G. Kulm and S. Malcom (Eds.), Science Assessment in the Service of Reform, pp. 384-390. Available from: American Association for the Advancement of Science, 1333 H St. NW, Washington, DC 20005 [AAAS Books: (301) 645-5643].

This paper briefly describes two field tests of "figural response" items. Figural response items require open ended responses by students in which students draw graphs, label diagrams, etc. They can be computer scored because the computer looks for the placement of key features in certain places on the answer sheet. For example, did the graph extend up to the point expected?

The first experiment involved field testing of 25 items to determine their feasibility for the *National Assessment of Educational Progress*. The second experiment involved computer-delivered items in which features could be moved around on the screen.

Several examples of items are provided. For more information see 600.6COMMUC.

(TC#600.6FIGRES)

Massachusetts Department of Education. *Massachusetts Educational Assessment Program. Open-ended and performance tasks in science*, 1989-91. Available from: The Commonwealth of Massachusetts, Department of Education, 1385 Hancock St., Quincy, MA 02169, (61") 770-7334.

The materials we received contain assessment materials for grades 4, 8 and 12 from three years (1988-1990) in four subject areas: science, math, social studies and reading. This entry describes the science portion of the materials.

The 1988 and 1990 materials describe open-ended test items in which students were given a written problem in which they had to apply concepts of experimental design or use concepts in life or physical sciences to explain a phenomenon. In 1988, three problems were given to fourth graders, six problems to eighth graders, and seven problems to twelfth graders. In 1990, three problems were given to fourth graders, and four were given to eighth and twelfth graders. Some of these were repeated across grade levels. All problems are included. Responses were analyzed for ability to note important aspects of designing an experiment or the amount of understanding of concepts they displayed. No specific performance criteria or scoring procedures are provided. However, there is extensive discussion of what students did, illustrated by sample responses. Because some of the information was also presented in multiple-choice format, the state was able to conclude that "although students appear to know and recognize the rules and principles of scientific inquiry when presented as stated options, unstructured situations that demand an application of these principles seem to baffle them."

In 1989, a sample of 2,000 students was assigned one of seven performance tasks (the three in science required lab equipment and/or manipulatives) to do in pairs. Each pair was individually watched by an evaluator. Each evaluator could observe between six and ten pairs each day. It took 65 evaluators five days to observe the 2,000 performances. Evaluators were to both check off those things that students did correctly (e.g., measured temperature correctly), and record observations of students' conversations and strategies. Again, detailed scoring procedures are not provided. There is, again, much discussion of observations illustrated by samples of student responses.

Some information about results for all the assessments is provided: percentages of students getting correct answers, using various strategies, using efficient methods, giving good explanations, etc., depending on the task. No technical information about the tests themselves are provided.

(TC#600.3MASOPS)

Medrich, Elliott A., and Jeanne E. Griffith. ***International Mathematics and Science Assessments: What Have We Learned?*****, 1992. Available from: National Technical Information Service, Springfield, VA 22161, (202) 219-1395.**

This report provides a description of the international assessments of math and science (First International Mathematics and Science Studies, 1960's; Second International Mathematics and Science Studies, 1980's; and First International Assessment of Educational Progress, 1988), some of their findings, and issues surrounding the collection and analysis of these data. It also offers suggestions about ways in which new data collection standards could improve the quality of the surveys and the utility of future reports.

Meinhard, Richard. ***A Developmental Baseline Profile of 12 Key Elementary Science Concepts/Processes*****, 1990. Available from: The Institute for Developmental Sciences, 3957 E. Burnside, Portland, OR 97214, (503) 234-4600.**

The OCATS (Oregon Cadre for Assistance to Teachers of Science) project is designed to encourage concept/process based science education in order to promote long range student growth in science. One part of this project was to gather information on how twelve science concepts develop in students from kindergarten through grade five. The concepts were:

1. Logical-mathematical organization of objects--simple classification, multiple classification, seriation, and whole number operations.

2. Geometrical and spatial relationships of objects--perimeter, area, and multiplicative projective relationships.

3. Physical properties of objects--quantity, weight, and volume.

4. Experimental reasoning--controlling variables.

5. Causal explanation--proportional reasoning.

One performance task was given to the students for each concept area. Performance was rated using a holistic developmental scale with four stages: sensory-motor (student engages in the activity without representational thought of the activity), preoperational (intuitive, no real understanding), operational (conceptual understanding under some circumstances), and formal (concept used as a variable in a more complex system of explanatory reasoning). Each stage has two substages for a final scale having eight points.

After discussing the results for the sample of 40 K-5 students in this study, the authors point out that the advantages of assessing students in this fashion are in knowing:

1. The readiness of students to handle instruction of certain types.

2. How to teach concepts to students in ways they can understand.

3. What needs to be done to move the student to higher developmental levels.

Neither the performance tasks nor the scoring techniques are described in detail in this paper. No technical information, except distribution of performance, is included.

(TC#600.6DEVBAP)

Mergendoller, J.R., V.A. Marchman, A.L. Mitman, and M.J. Packer. ***Task Demands and Accountability in Middle-Grade Science Classes*****, 1987. Located in: Elementary School Journal, 88, pp. 251-265.**

The authors maintain that the types of thinking students engage in and the quality of learning that occurs are largely influenced by the nature of the tasks students complete. After analyzing a large number of instructional and assessment tasks given to eighth graders, the authors conclude that, in general, the tasks given students present minimal cognitive demands. The article also provides suggestions about analyzing and modifying curriculum tasks.

Although not strictly about assessment, the article is included here to reinforce the notion that, as in instruction, the task given to students in a performance assessment can affect how well one can draw conclusions about student ability to think--if students are not given performance tasks that require thinking, it would be difficult to analyze responses for thinking ability.

(TC#600.6TASDEA)

Moran, Jeffrey B., and William Boulter. ***Step by Step.*** **Located in: Science Scope, 15, March 1992, pp. 46-47, 59.**

This article describes the assessment approach used in the Second International Science Study. Students were presented with an exercise and then lead through 13 questions, each of which builds upon the previous questions. After the student responds to each step of the exercise, he or she is shown the ideal response. As a result, a student's successful performance does not depend upon every task being performed correctly. The authors partially demonstrate the procedure using two examples: water temperature and length v. pitch.

No technical information is provided.

(TC#600.6STESTS)

National Assessment of Educational Progress (NAEP--1987). ***Learning by Doing: A Manual for Teaching and Assessing Higher-Order Thinking in Science and Mathematics. Report No. 17-HOS-80.*** **Available from: Educational Testing Service, CN 6710, Princeton, NJ 08541, (800) 223-0267.**

The National Assessment of Educational Progress was established in 1969 to monitor student achievement status and trends. Samples of students aged 9, 13 and 17 are tested periodically, with science assessments having occurred in 1970, 1973, 1982, 1986, and 1990.

Learning by Doing is an overview of a pilot test of "higher-order thinking skills" that was added to the 1986 assessment. This pilot consisted of 30 tasks/items in the areas of sorting/classifying, observing/formulating hypotheses, interpreting data, and designing/conducting an experiment. The tasks included open-ended paper and pencil items, use of equipment at stations, and complete experiments. *Learning by Doing* briefly describes 11 of the exercises presented to students. (The full report is available from NAEP at the above address.)

Lisa Hudson in *Assessment in the Service of Instruction* (TC#600.6ASSINT) discusses some issues with respect to this pilot test and the 1990 science assessment. These include whether the time and cost of giving the performance items really provides that much extra information; how the ability to read, listen, and write might affect scores; and whether this type of task would differentially encourage inquiry-based instruction. (These are questions that relate to all performance assessments and not just the NAEP pilot.)

(TC#050.6LEABYD)

National Center for Improving Science Education. *Getting Started in Science: A Blueprint for Elementary School Science Education*, 1989. Available from: National Center for Improving Science Education, 2000 L St. NW, Suite 602, Washington, DC 20036, (202) 467-0652. Also available from ERIC: ED 314 238.

This report covers such topics as the rationale for science instruction, how children learn science, teacher development and support, and assessment. The chapter on assessment promotes the idea of assessment in the service of instruction--measuring the full range of knowledge and skills required for science, alignment with instruction, and a range of assessment approaches.

The authors outline the characteristics of a good assessment system, including characteristics of tests, measuring affective as well as cognitive dimensions, and assessing instruction and curriculum.

(TC#600.6GETSTS)

National Science Foundation. *Educating Americans for the 21st Century: A Plan of Action for Improving Mathematics, Science and Technology Education*, 1983. Available from: National Science Board Commission on Precollege Education in Mathematics, Science and Technology, Forms & Publications Unit, 1800 G St. NW, Room 232, Washington, D.C. 20050, (202) 357-3619.

This is not strictly a document regarding assessment, but rather a statement of what students need to know and be able to do in science and math. As such, it also provides an outline for what assessments should measure.

(TC#000.5EDUAMF)

National Science Teachers Association. *Scope, Sequence and Coordination of Secondary School Science, Volume 1: The Content Core, A Guide for Curriculum Designers*, 1992. Available from: The National Science Teachers Association, Special Publications Dept., 1742 Connecticut Ave. NW, Washington, DC 20009, (202) 328-5800.

This book is one of two (the other is Science for All Americans, TC#600.5SCIFOA) that appear to currently be the standard for defining the emphasis of secondary science (grades 6-12). The document emphasizes the need to do more than have students memorize facts, the philosophy that students need to be involved in the practical applications of science, the approach that the various subject areas need to be coordinated, the philosophy that all students need to be scientifically literate, and the belief that students learn best when they construct their own meaning. However, the scope, itself, concentrates mainly on the knowledge part of the curriculum.

(TC#600.5SCOSEC)

***New York State Elementary Science Program Evaluation Test (ESPET)*, 1989. General information available from: Bureau of Science Education, Office of General and Occupational Education, Division of Arts and Sciences Instruction, The State Education Department, The University of the State of New York, Albany, NY 12234, (518) 474-7746.**

This paper provides only a general description of the Elementary Science Program Evaluation Test (ESPET) and some commonly-asked questions and answers. Actual test items will not be available for a few years (when they are no longer being used by the state). The information provided indicates that only grade 4 students are tested.

ESPET consists of two required components and five optional components. These components include:

Required:

- Objective Test--45 multiple-choice items
- Manipulatives Skills Test

Optional:

- Student Science Attitudes
- Student Survey
- Teacher Survey
- Administrator Survey
- Parent/Guardian Survey

The Manipulative Skills Test was developed to evaluate a number of inquiry and communication skills. It consists of five stations with a total of 15 exercises and requires about an hour to administer to a class. A brief description of the 5 tasks is included, but no detail on either tasks or scoring is included. All testing materials are available in several languages.

The results from the New York State ESPET are used to provide data to help local educators make decisions to improve their elementary science program and to help the state identify those programs in need of state technical assistance. The results are reported at the program level--no individual student achievement results are given.

These tests are scored at the local level and results reported to the state. The test is given to about 200,000 grade 4 students each year.

(TC#600.3NEWYOE)

O'Rafferty, Maureen Helen. *A Descriptive Analysis of Grade 9 Pupils in the United States on Practical Science Tasks*, 1991. Available from: University Microfilms International Dissertation Services, 300 N. Zeeb Rd., Ann Arbor, MI 48106, (800) 521-0600.

This dissertation was a re-analysis of some of the information from the Second International Science Assessment (1986), but also includes a good description of the performance portion of the SISS and three of the six performance tasks. (The SISS also contained a multiple-choice portion and several surveys.)

The three tasks included in this document (Form B) were: determining the density of a sinker, chromotography observation and description, and identifying starch and sugar. The other three tasks (Form A) in the SISS, not included in the document, are: using a circuit tester, identifying solutions by ph, and identifying a solution containing starch. Each task has a series of questions for the student to answer using the equipment provided. Form A had 11 total questions and Form B had 10. These questions asked students to observe, calculate, plan and carry out a simple experiment, explain, and determine results. Each subquestion was classified as being one of three types of process skills: performing, reasoning, or investigating. The six tasks were set up at 12 alternating stations A, B, A, B, ...). Students had 10 minutes at each station, plus five minutes in between. So, 12 students could be tested each 45 minutes.

One to two points were given for each answer. The basis for assigning points was not clear, but appears to be based on a judgement of the correctness of the response.

The dissertation includes a number of student responses to the tasks, overall performance of the U.S. population, and several reinterpretations of the results. For example, student performance on questions classified as measuring the same skill were widely different. The author speculates that this is either because the definitions of the skills are imprecise, or because such unitary skills don't exist.

The author also examined student responses for patterns of errors, and discussed the implications of this for instruction.

(TC#600.3DESANP)

Ostlund, Karen. *Sizing Up Social Skills.* Located in: Science Scope, 15, March 1992, pp. 31-33.

The author presents a taxonomy of social skills important for the science classroom, provides a few ideas for how to teach them, and a couple of ideas on student and teacher monitoring techniques.

(TC#223.6SIZUPS)

Padilla, Michael J., Vantipa Roadrangka, and Russell H. Yeany. *Group Assessment of Logical Thinking*, 1982. Available from: Michael J. Padilla, University of Georgia, 212 Aderhold Hall, Athens, GA 30602, (706) 542-3000.

This assessment is an enhanced multiple-choice test based on Piaget logical operations. The test consists of 21 items and is for students with a reading level of grade 6 and above. It purports to measure six logical operations: conservation, proportional reasoning, controlling variables, combinatorial reasoning, probabilistic reasoning, and correlational reasoning. Each item is presented pictorially. The student chooses both a statement he/she believes is true about the pictures and the reason for this choice. All items are multiple-choice except for the combinatorial reasoning items for which students list all possible combinations.

There is technical information to support the conclusion that the test can distinguish groups at concrete, transitional, and formal stages of development. Although the authors recommend that this information be used to plan instruction at the proper developmental level, no concrete examples of how to do this are provided.

(TC#050.3GROASL)

Pine, Jerry, Gail Baxter, and Richard J. Shavelson. *Assessments for Hands-On Elementary Science Curricula*, 1991. Available from: Physics Department, California Institute of Technology, Pasadena, CA 91125, (818) 356-6811.

The authors present the case that science curriculum should enable students to learn how to pursue an experimental inquiry, and should give them the ability to construct new knowledge from their observations. Assessment should match this. But, they question whether it is always necessary to have hands-on assessment tasks. The authors designed a study that compared observer rating of fifth- and sixth-grade student performance on hands-on tasks with five other surrogates: ratings of student lab notebooks that covered the same hands-on tasks, a computer simulation of the tasks, free-response paper and pencil questions, multiple-choice items, and CTBS scores. The surrogates (with the exception of the CTBS) were designed to parallel the hands-on tasks as closely as possible.

This paper reports on the relationship of observer ratings, notebook ratings, simulations, and CTBS scores. Results showed:

1. It was possible to get consistent ratings of student performance on hands-on tasks with trained observers.

2. Ratings of lab notebooks were a promising surrogate for observations, but they have to be designed carefully.

3. Computer simulations, open-ended questions, and multiple-choice questions were not good surrogates.

4. CTBS scores were moderately related to hands-on performance, but appeared to mainly reflect general verbal and numerical skills.

5. In order to assess inquiry instruction rather than general natural ability, hands-on tasks need to be carefully designed.

The paper briefly describes all the tasks used in the study, but does not present them in enough detail to replicate. A companion paper, *New Technologies for Large-Scale Science Assessments: Instruments of Educational Reform* (TC#600.3NEWTEF), describes all the tasks in more detail.

(TC#600.3ASSFOH)

Psychological Corporation. *Integrated Assessment System--Science Performance Assessment*, 1992. Available from: Psychological Corporation, Order Service Center, PO Box 839954, San Antonio, TX 78283-3954, (800) 228-0752.

This is a series of seven tasks designed to be used with students in grades 2-8 (one task per grade level). The tasks involve designing and conducting an experiment based on a problem situation presented in the test. Students are provided various materials with which to work. Students may work individually or in teams, but all submitted products must be individually generated. Students generate a hypothesis they wish to test, write down (or show using pictures) the procedures used in the experiment, record data, and draw conclusions. At the end, students are asked to reflect on what they did and answer questions such as: "What problem did you try to solve?" "Tell why you think things worked the way they did," and "What have you seen or done that reminds you of what you have learned in the experiment?" The final question in the booklet asks students how they view science. This question is not scored but can be used to gain insight into students' performances.

Only the written product in the answer booklet is actually scored. (However, the publisher recommends that teachers watch the students as they conduct the experiment to obtain information about process. A checklist of things to watch for is provided.) Responses can be scored either holistically or analytically using criteria generalized so that they can be used with any task. The holistic scale (0-6) focuses on an overall judgment of the performance based on quality of work, conceptual understanding, logical reasoning, and ability to communicate what was done.

The four analytical traits are experimenting (ability to state a clear problem, and design and carry out a good experiment), collecting data (precise and relevant observations), drawing conclusions (good conclusions supported by data), and communicating (use of appropriate scientific terms, and an understandable presentation of what was done.). Traits are scored on a scale of 1-4.

There is a scoring guide that describes the procedure. However, in the materials we obtained, there are no student performances provided to illustrate the scoring. No technical information about the assessment is included.

(TC#600.3INTASS)

Psychological Corporation. *GOALS: A Performance-Based Measure of Achievement--Science*, 1992. Available from: Psychological Corporation, Order Service Center, PO Box 839954, San Antonio, TX 78283-3954, (800) 228-0752.

GOALS is a series of open-response questions (only one right answer) that can be used alone or in conjunction with the MAT-7 and SAT-8. Three forms are available for 11 levels of the test covering grades 1-12 for each of science, math, social studies, language and reading. Each test (except language) has ten items. On the science test, tasks cover content from the biological, physical, and earth/space sciences. Each task seems to address the ability to use a discrete science process skill (e.g., draw a conclusion, record data) or use a piece of scientific information. The tasks require students to answer a question and then (usually) provide an explanation.

Responses are scored on a four-point holistic scale (0-3) which emphasizes the degree of correctness of plausibility of the response and the clarity of the explanation. A generalized scoring guide is applied to specific questions by illustrating what a 3, 2, 1 and 0 response look like.

Both norm-referenced and criterion-referenced (how students look on specific concepts) score reports are available. Scoring can be done either by the publisher or locally. A full line of report types (individual, summary, etc.) are available.

The materials we obtained did not furnish any technical information about the test itself.

(TC#610.3GOALSS)

Raizen, Senta and J. Kaser. *Assessing Science Learning in Elementary School: Why, What, and How?* Located in: Phi Delta Kappan, May 1989, pp. 718-722.

This paper describes some of the limitations of current standardized, multiple-choice tests to assess science, discusses how this combines with inadequate teacher preparation and textbooks to create inferior science instruction, and provides a list of questions to ask about any test being considered for use. The list of questions includes such things as "Are problems with more than one correct solution included?" and "Are there assessment exercises that encourage students to estimate their answers and to check their results?"

(TC#600.6ASSSCL)

Raizen, Senta A., Joan B. Baron, Audrey B. Champagne, E. Haertel, Ina V.S. Mullis, and Jeannie Oakes. *Assessment in Elementary School Science Education*, 1989. Available from: The National Center for Improving Science Education, 2000 L St. NW, Suite 602, Washington, DC 20036, (202) 467-0652. Also available from: ERIC ED 314 236.

The authors discuss the following topics: why assessment is important, issues in assessment, what to assess, how to assess, using assessment in instruction, and assessment of program features. The emphasis is on using assessment to enhance instruction not to undermine it. A lengthy appendix describes "fundamental organizing concepts in science that all students, by the time they finish sixth grade, should incorporate in the way they think about and engage their world." These include: orderliness, cause and effect, systems, scale, models, change, structure and function, variations, and diversity. There is a definition of each area and examples of K-6 instructional activities.

This appears to be a longer and more detailed version of 600.6GETSTS--see *National Center for Improving Science Education.*

(TC#600.6ASSELS)

Riggs, Iris M. and Larry G. Enochs. *Toward the Development of an Elementary Teacher's Science Teaching Efficacy Belief Instrument*, 1989. Paper presented at the 62nd Annual Meeting of the National Association for Research in Science Teaching, San Francisco, CA. Available from: ERIC ED 308 068.

This publication reports on a study in which the Personal Science Teaching Efficacy Belief Scale and the Science Teaching Outcome Expectancy Scale were administered to measure teacher feelings of self-efficacy and outcome expectancy. The authors present evidence that the combined instrument is valid for studying elementary teacher's beliefs toward science teaching and learning. The instrument is included.

(TC#600.4TOWDEE)

Roth, Wolff-Michael. *Dynamic Evaluation*. Located in: Science Scope, 15, March 1992, pp. 37-40.

The author describes a method by which students plan and report experiments: the Vee Map. The Vee Map requires students to list vocabulary related to the topic they are reporting, develop a concept map of these terms, describe the experimental design, describe the data collected, and present their conclusions. One extended example in earth science is given. Performance criteria for assessing the Vee Map is sketchy. No technical information is included.

(TC#630.6DYNEVA)

Rutherford, F. James and Andrew Ahlgren. ***Science for All Americans--Science Literacy*, 1990. Available from: Oxford University Press, Inc., 200 Madison Ave., New York, NY 10016, (800) 334-4249.**

This book is one of two (the other is Scope, Sequence, and Coordination, TC#600.5SCOSEC) that appear to currently be the standards for defining what the content and emphasis of science instruction should be. The premise is that, although not everyone will be a scientist, future success of humanity requires that everyone have a certain level of scientific literacy--knowledge, habits of mind, and the desire to be a critical thinker. The chapters cover the following kinds of goals we should have for students: the scientific endeavor as a human enterprise, basic knowledge about the world, major scientific themes, and habits of mind.

(TC#600.5SCIFOA)

Semple, Brian McLean. ***Performance Assessment: An International Experiment*, 1992. Available from: ETS, Scottish Office, Education Department, Rosedale Rd., Princeton, NJ 08541, (609) 734-5686.**

This report describes the Second International Assessment of Educational Progress on math and science conducted in 1991. Eight math and eight science tasks were given to a sample of thirteen-year-olds in five volunteer countries (Canada, England, Scotland, USSR, and Taiwan). This sample was drawn from the larger sample involved in the main assessment.

The 16 hands-on tasks are arranged in two 8-station circuits. Students spend about five minutes at each station performing a short task. Most tasks are "atomistic" in nature; they measure one small skill. For example, the 8 math tasks concentrate on measuring length, angles, and area, laying out a template on a piece of paper to maximize the number of shapes obtained, producing given figures from triangular cut-outs, etc. Some tasks require students to provide an explanation of what they did. All 16 tasks are included in this document, although some instructions are abbreviated and some diagrams are reduced in size.

Most scoring appears to be right/wrong. (However, it is not entirely clear how the explanations are scored. It consists of some kind of judgement of reasonableness of the explanation.) There must also have been some observation of how the students approached the tasks, because a detailed analysis of such strategies for one problem is given.

Student summary statistics on each task are included. There is a brief summary of teacher reactions, student reactions, the relationship between student performance on various tasks, and the relationship between performance on the multiple-choice and performance portions of the test.

(TC#600.3PERASS)

Shavelson, Richard J., Neil B. Carey, and Noreen M. Webb. *Indicators of Science Achievement: Options for a Powerful Policy Instrument.* Located in: Phi Delta Kappan, May 1990, pp. 692-697.

The authors review reasons for moving from multiple-choice tests of science achievement to more performance-based measures, and then discuss three examples: looking at how well students can move between different representation of a problem, mental models, and performance assessments/surrogates.

(TC#600.6INDSCA)

Shavelson, Richard J., Gail P. Baxter, Jerry Pine, and J. Yure. *New Technologies for Large-Scale Science Assessments: Instruments of Educational Reform*, 1991. Available from: University of California, 552 University Rd., Santa Barbara, CA 93106, (805) 893-8000.

This document is a series of papers that report in more detail on the studies of hands-on versus surrogate assessment tasks also described in *Assessments for Hands-On Elementary Science Curricula* (TC#600.3ASSFOH). This includes more detailed descriptions of the three hands-on tasks (paper towels, sow bugs, and electric mysteries) and computer simulations. Findings, in addition to those reported in the companion paper, include:

1. Although observers could be trained to be very consistent in their ratings, a major source of error is still in the tasks chosen. That is, the decision about the level of an individual's performance depends greatly on the particular task used.

2. Hands-on assessment provides different information than that provided by paper and pencil tests.

For additional information to those reported in this paper and its companion paper see the following references:

Baxter, Gail P., Richard J. Shavelson, Susan Goldman, and Jerry Pine. *Evaluation of Procedure-Based Scoring for Hands-on Science Assessment*. Journal of Educational Measurement, 1992, 29, pp. 1-17. (TC#600.3EVAPRB)

Shavelson, Richard J., and Gail P. Baxter. *What We've Learned About Assessing Hands-On Science*. Located in: Educational Leadership, Vol. 49, No. 8, May 1992, pp.20-25. (TC#600.3WHAWEL)

Shavelson, Richard J., Gail P. Baxter, and Jerry Pine. *Performance Assessments--Political Rhetoric and Measurement Reality*. Located in: Educational Researcher, Vol. 21, No. 4, May 1992, pp.22-27. (TC#600.3PERASP)

Shavelson, Richard J., Maria Araceli Ruiz-Primo, Gail P. Baxter. *On the Stability of Performance Assessments*. Located in: Journal of Educational Measurement, Spring 1993, 30, pp. 41-53. (TC#600.6ONSTAP)

(TC#600.3NEWTEF)

Small, Larry. ***Science Process Evaluation Model*****, 1992. Available from: Schaumburg Community Consolidated District #54, 524 E. Schaumburg Rd., Schaumburg, IL 60194, (708) 885-6700.**

This document contains a paper presented at a national conference in 1988 which briefly describes Schaumburg's science assessment system, and a set of tests for students in grades 4-6 contributed in late 1992.

The tests have three parts: multiple-choice to measure content and some process skills, self-report survey to assess attitudes toward science, and hands-on to assess science process skills.

The hands-on part attempts to measure 11 student science process skills: observing, communicating, classifying, using numbers, measuring, inferring, predicting, controlling variables, defining operationally, interpreting data, and experimenting. It consists of students using manipulatives to answer fixed questions such as "Which drop magnifies the most?" or "Which clay boat would hold the most weights and still float in the water?" Students respond by choosing an answer (multiple-choice), supplying a short answer, or, in a few cases, drawing a picture or graph. Complete tests for Grades 4, 5, and 6 are included.

No scoring procedures or technical information were included with the package. For additional information on this project see *Teamwork Testing* (TC#650.3TEATES)

(TC#600.3SCIPRE)

Small, Larry, and Jane Petrek. ***Teamwork Testing*****, 1992. Located in: <u>Science Scope</u>, 15, March 1992, pp. 29-30.**

The authors describe a model for performance-based assessment in chemistry (middle school) which emphasizes group cooperation and the process of doing science. One task was described in detail. Performance criteria were hinted at, but not described.

For other information on this project see *Science Process Evaluation Model* (TC#600.3SCIPRE).

(TC#650.3TEATES)

Surber, John R., Philip L. Smith, et al. MAP Tests: ***Structural Maps of Text as a Learning Assessment Technique: Progress Report for Phase I, Technical Report No. 1; Testing for Misunderstanding;*** **and** ***The Relationship Between Map Tests and Multiple Choice Tests, Technical Report No. 6.*** **Available from: John R. Surber, Department of Educational Psychology, University of Wisconsin--Milwaukee, Milwaukee, WI 53201, (414) 229-1122.**

These reports describe the development of map tests as an assessment technique to identify conceptual misunderstandings that occur when students learn from text. In this testing technique, concepts and their interrelationships are represented graphically. These graphic representations are called text maps. A training manual for constructing text maps is included. The manual introduces the symbols to be used in the concept map to indicate: 1) definitions; 2) characteristics or properties; 3) examples; 4) temporal relations; 5) causal relations; 6) similarity; and 7) greater or less than comparisons.

The papers present four methods of using maps to assess the structure of student knowledge. All involve various levels of deleting information from a completed text map and providing clues on content and structure. Students complete the missing information--similar to a cloze test.

Text maps and map tests can be constructed using any content area--science, social studies, etc. They can be used in study skills or reading classes. In these reports, the content of the training manual is drawn from chemistry and study skills.

Technical information on map tests can be found in the following document: Surber, John R., Philip L. Smith, and Fredrika Harper. *Technical Report No. 6. The Relationship Between Map Tests and Multiple Choice Tests.* March 1982.

Additional information can be found in: Surber, John R. and Philip L. Smith. *Testing for Misunderstanding.* Located in: *Educational Psychologist,* 1981, 16,, pp. 165-174.

(TC#150.6MAPTES)

Tamir, Pinchas and S. Glassman. *Laboratory Test for BSCS Students.* Located in: BSCS Newsletter, 42, Feb. 1971, pp. 9-13.

The authors provide information on six performance tasks designed to assess high school students' laboratory skills. The tasks include: photosynthesis, human respiration, grasshopper respiration, yeast fermentation, plant tissue and Daphnia activity. Descriptions of the tasks, directions to students, and scoring criteria are provided. There is no technical information.

(TC#640.3LABTEF)

Tamir, Pinchas. *An Inquiry Oriented Laboratory Examination.* Located in: Journal of Educational Measurement, 11, 1974, pp. 25-33.

The author discusses the need for performance assessments in laboratory skills, and presents a detailed description of one--pH variations. This task was used in a large-scale study; results are reported.

(TC#650.3INQORL)

Van Pelt, Jerry. *Discovering the Problem of Solid Waste: Performance Assessments* (ISBE Science Literacy Grant), 1991. Available from: Lake County Educational Service Center, 19525 W. Washington St., Grayslake, IL 60030, (708) 223-3400.

In this booklet, 17 performance tasks are presented for students in grades 3-6. The tasks are based on an instructional manual used to teach the topic of solid waste and assess knowledge of the topic and application of that knowledge in hands-on activities. Not all the tasks are appropriate for each of the grades.

Each performance task contains information about grade level, concepts being assessed (e.g., types of solid waste or recognizing changes in materials in a landfill), process skills needed to complete the task (e.g., classifying, measuring, observing, or ordering), and the objects/items needed for the task, directions, and questions to answer. Many of the tasks are completed at home or at a work station in the classroom.

Scoring emphasizes the correctness of the response; the scoring guides are different for each task. The guide provides information on the maximum points to assign for each question and for the entire task.

No information on staff training or technical information is provided.

(TC#620.3DISPRS)

Vargas, Elena Maldonado and Hector Joel Alvarez. Mapping Out Students' Abilities. Located in: Science Scope, 15, March 1992, pp. 41-43.

The authors use concept maps to assess the knowledge structures students have on various concepts in science. They give some brief help on how to design a concept map, and more extensive help on how to score maps. Two examples are given: matter and photosynthesis.

(TC#600.6MAPOUS)

Whetton, Chris, Marian Sainsbury, Steve Hopkins, Dorian Bradley, and Alan Greig. *National Assessment in England and Wales*, 1992. Available from: National Foundation for Educational Research (NFER), The Mere, Upton Park, Slough, Berks, S11 2DQ, England, UK.

This document is a series of papers presented at the American Educational Research Association meeting in 1992. It updates the status of the science assessment described in other entries for Whetton: *Science for Seven-Year Olds* (TC#600.6SCIFOS), *The Pilot Study of Standard Assessment Tasks for Key Stage 1* (TC#600.6PILSTO), and *Standard Assessment Tasks for Key Stage 1* (TC#100.3STAAST). For additional information see Harlen (TC#600.6PERTES).

The papers review the history of the assessment, describe and present a few examples of the assessment tasks for seven-year-olds, discuss the support needed to assist teachers to administer this large number of performance tasks, describe the changes that resulted for the 1992 assessment, and briefly describe plans for the 14-year-old assessment.

(TC#600.6NATASE)

Whetton, Chris. *Science for Seven-Year-Olds in England and Wales*, 1991. Available from: National Foundations for Educational Research, The Mere, Upton Park, Slough, Berks S11 2DQ, England, UK.

This paper reports on the development in England and Wales of performance assessments that are tied to their new National Curriculum. In spring, 1991, all seven-year-olds (600,000) were tested. This paper discusses the pilot that was carried out in 1990 and the changes made for the 1991 assessment. Although this paper addresses all subject areas, the examples are selected from the science portion.

Student performance was noted on over 200 "standards of achievement" observed during a series of specified performance tasks. In addition to these tasks, students also had a "science interview" to assess knowledge of specific facts.

Due to the pilot test, the full scale assessment for 1991 was modified so that:

1. Fewer attainment targets will be noted; 200 separate judgments was too many for teachers to make.

2. Not all attainment targets will be noted for each child; teachers will choose targets based on previous assessment results.

3. Certain "core" targets will be covered for all students. In addition, one extra target in science and math will be selected for each student.

4. Each task will focus on only one or two attainment targets.

5. Science interviews have been abandoned.

A related document, *The Pilot Study of Standard Assessment Tasks for Key Stage 1* (TC#600.6PILSTO) contains a complete description and analysis of the pilot, and *Standard Assessment Tasks for Key Stage 1* (TC#600.3STAAST), contains the complete 1991 assessment package for all content areas. For additional information see other entries for Whetton and Harlen (TC#600.6PERTES).

(TC#600.6SCIFOS)

Whetton, Chris, G. Ruddock, Steve Hopkins, et al. *The Pilot Study of Standard Assessment Tasks for Key Stage 1*, 1991. Available from: National Foundations for Educational Research, The Mere, Upton Park, Slough, Berks S11 2DQ, England, UK.

This set of two reports, describes the pilot test of the age 7 performance tests in England in more detail than that reported in *Science for Seven-Year-Olds in England and Wales* (TC#600.6SCIFOS). For other information see additional entries for Whetton and Harlen (TC#600.6PERTES).

(TC#070.3STAASTm--Inhouse use only)

Whetton, C., G. Ruddock, Steve Hopkins, et al. *Standard Assessment Tasks for Key Stage 1*, 1991. Available from: National Foundations for Educational Research, The Mere, Upton Park, Slough, Berks S11 2DQ, England, UK.

This package contains all the materials used by teachers for the age 7 *Standard Assessment Tasks*--administration handbooks, detailed description of tasks and scoring procedures, information recording booklets, and student worksheets. For related information see other entries from Whetton and Harlen (TC#600.6PERTES.)

(TC#100.3STAAST--In house use only.)

Wiggins, Grant. *The Futility of Trying to Teach Everything of Importance*. Located in: <u>Educational Leadership</u>, November 1989, pp. 44-48, 57-59.

Assessment has to reflect what we value. This article presents a philosophy for science instruction that has implications for assessment. Specifically, the author maintains that the goal of education should not be to teach every fact that we think students will need to know, because this will be impossible to do. Rather, we should concentrate on developing those habits of mind and high standards of craftsmanship that will enable students to be lifelong learners and critical thinkers. The article briefly mentions some of the implications for assessment of this philosophy.

(TC#600.6FUTTRT)

Yager, Robert E. and Alan J. McCormack. ***Assessing Teaching/Learning Successes in Multiple Domains of Science and Science Education.*** **Located in: Science Education, 73, 1989, pp. 45-58.**

This article describes the authors' view of the proper targets for instruction in science (knowing and understanding, exploring and discovering, imagining and creating, feeling and valuing, and using and applying), goes on to describe the STS (Science-Technology-Society) approach to teaching science, and then lists some tests (mostly multiple-choice) that attempt to measure the targets. The paper is included on this bibliography mainly for the first two points.

(TC#600.5ASSTEL)