DOCUMENT RESUME

ED 284 899   TM 870 450

AUTHOR   Aschbacher, Pamela R.; Herman, Joan
TITLE   R&D Needs for Assessment in the Content Areas.
                Testing Study Group: Content Assessment. Report on
                Research Needs.
INSTITUTION   California Univ., Los Angeles. Center for the Study
                of Evaluation.
SPONS AGENCY   Office of Educational Research and Improvement (ED),
                Washington, DC.
PUB DATE   Mar 87
GRANT   OERI-G-86-0003
NOTE   27p.
PUB TYPE   Reports - Evaluative/Feasibility (142)

EDRS PRICE   MF01/PC02 Plus Postage.
DESCRIPTORS   *Educational Assessment; Educational Testing;
                Elementary Secondary Education; *Intellectual
                Disciplines; *Needs Assessment; *Research and
                Development; *Research Needs; Research Projects;
                School Districts; School Surveys; Science Education;
                Social Studies; State Programs; State Surveys;
                Testing Programs

ABSTRACT
                This paper surveys the status of current state and
district level practice in content assessment, highlights related
research efforts currently underway, and identifies high priority
areas for subsequent research in content assessment. A needs
assessment for research in content area assessment was conducted
during 1986 by the Center for Research on Evaluation, Standards, and
Student Testing (CRESST). District level administrators identified
social studies and science as top priority areas for content testing.
About half the districts surveyed currently assess these areas,
primarily using locally produced tests. At the state level, there has
been little recent assessment of content areas other than the
National Assessment of Educational Progress, and that which exists
appears rather limited in scope and technique. A survey of state
directors of research and assessment in 1986 confirmed that science
and social studies are top priority areas for current and anticipated
testing for purposes of accountability, curriculum planning, and
student diagnosis. Most of current research on the content areas
focuses on issues in learning, instruction, and curriculum rather
than on assessment issues. Future research and development efforts
should: (1) identify which facts, concepts, and processes should be
assessed in each field; (2) address how best to assess the targeted
constructs and processes; and (3) determine how to facilitate the use
of new content area measures. (BAE)

Center for Research on Evaluation,
Standards and Student Testing

DELIVERABLE — March 1987

TESTING STUDY GROUP: CONTENT ASSESSMENT

Report on Research Needs

Study Director: Joan Herman

CENTER FOR THE STUDY OF EVALUATION

Graduate School of Education

University of California, Los Angeles

R&D NEEDS FOR ASSESSMENT IN THE CONTENT AREAS

by

Pamela R. Aschbacher and Joan Herman

4

Rationale

The l   les have witnessed explosive growth in
achievemer   and a concomitant belief in the power of
testing to   variety of purposes. The examples are many:
school board accountability concerns, state and local minimum
competency mandates, evaluation requirements for federal and
state programs, national assessments, and the growth of
curriculum-embedded and curriculum-based assessment systems. All
have converged to make testing a major enterprise in American
education, an emphasis which thus far ha.- focused almost
exclusively on student progress in the 3 R's. Now, however,
curricul* in the content areas are undergoing intense debate and
analysis, and new interest has emerged in the assessment of
student achievement in the sciences and social studies (e.g.,
Olson, 1984; Resnick, 1983) and in critical thinking skills
across all content areas (Costa, 1985).

Why the concern? The roots are both philosophical and
practical. While test scores in the basics show progress, many
educators fear that this progress has come at the expense of
higher order skills and knowledge in other areas. For example,
the Educational Commission of the States (1982) reported that:

> Today's minimum skills are demonstrated successfully by a
> majority of students. Higher order skills, however, are
> achieved only by a minority of 17-year-olds. If this trend
> continues, as many as two million students may graduate in
> 1990 without the skills necessary for employment in
> tomorrow's marketplace.

The extraordinary rate of emerging knowledge in today's world
makes it imperative that students be taught not just factual

material, but also the organizing, reasoning, conceptualization, problem solving, and analysis skills necessary to acquire and process information within our ever-expanding fields of knowledge.

Students' poor comprehension of concepts in science and social studies and inadequate application of knowledge to problem solving tasks have been documented by a number of researchers (e.g., Gabel, Sherwood & Enochs, 1984; Resnick, 1983; Texley & Norman, 1984; Yarroch, 1985). For example, Gabel examined general problem solving behavior of high school chemistry students by analyzing data obtained from students while they solved chemistry problems aloud. Results clearly showed that few students used reasoning skills in solving any problems whatsoever. Not only did they rely on memorized algorithms, but they attempted to fit them to inappropriate situations. The authors concluded that students relied on algorithms as a substitute for deep understanding of science concepts.

Some authors have suggested that the apparent lack of good critical thinking skills reflects the narrow emphasis in the typical classroom on lower order tasks (Goodlad, 1983; McTighe & Schollengberger, 1985). In Goodlad's study of schooling involving over 1,000 classrooms across the country, observers noted that less than one percent of teachers' instructional communication to students invited them to engage in anything more than mere recall of information. Much of the science and social studies curricula involve the transmission of a prodigious amount of very specifi. nowledge. Of course knowledge is important,

but even more critical in daily life is the capacity to organize and use that knowledge in classifying, analyzing, comparing, inferring, hypothesizing, and drawing conclusions.

Hirsch (1985a,b) and others (Olson, 1984), concerned with the importance of cultural literacy in a democratic society, have expressed the fear that many people in our country do not possess the requisite knowledge and skills. Hirsch makes the case that democracy depends on a literate populace that can communicate effectively through reading, writing and speaking. Those without basic cultural knowledge and the ability to think critically about issues cannot participate effectively in our democracy, especially as local, national and international issues become increasingly complex. Yet, our curricula have not been adequately preparing students for such participation. For example, science has been deemphasized in the elementary grades over a period of years. Historically, science courses have primarily prepared students for higher education and have given little attention to the necessity for widespread scientific literacy and little opportunity for students to experience the power and processes of scientific investigation. Social studies, too, has tended to emphasize memorization of people, events, dates and so forth rather than use of critical thinking skills applied to this knowledge. Thus many students today do not have adequate background information, nor the higher order skills of argument and evaluation, to act as informed citizens.

The focus on basic skills during a period of increasingly limited resources has tended to narrow the curriculum and preclude much instruction in the content areas. Advocates of

increases assessment in the content areas hope that such policies will be instrumental in reasserting the value of science and social studies and in securing their appropriate role and emphasis in the total school curriculum. Widespread assessment of the disciplines, they believe, may promote accountability and spur instructional efforts in the content areas just as it did in basic skills.

As educators, legislators, and others have become more concerned about improving curricula, particularly in science and social studies, the limitations of our current assessment techniques have become more apparent. For example, Morgenstern and Renner's (1904) analysis of commercially available, standardized high school science tests revealed that seven of the twelve tests in the sample and 90% of all the analyzed items required only recall of factual information. The authors were dismayed by the paucity of items dealing with higher order skills such as comparing, inferring, analyzing, evaluating, and synthesizing.

A number of researchers, disheartened by the preponderance of items assessing lower level skills, have attempted to fill the gap by creating multiple-choice tests of process skills in science and social studies, but they have met with varying degrees of success (Burns, Okey, & Wise, 1985; Lehman, Carter & Kahle, 1985; Ross & Maynes, 1983).

Other researchers, however, have questioned traditional methods of large-scale assessment and expressed dismay about the limits of multiple choice testing. Fredrikson (1984), for

instance, has urged measurement experts and the testing industry to move away from reliance on selected response tests to performance tests that simulate criterion tasks (cf. McClelland, 1973; Shavelson, 1985a), and to nontraditional tests that reveal something of the test taker's cognitive processes (cf. Curtis & Glaser, 1983; Haertel & Calfee, 1983; Linn, 1983; Shavelson, 1985b). Mirroring this view, attention to interviews and the think-aloud technique to illuminate students' problem solving capabilities has become more prominent within the last decade, particularly in math and science (cf. Finegold & Mass, 1985; Gabel et al., 1984). While these methods have the advantage of providing far more information about what students are actually thinking than can be obtained via paper and pencil tests, current techniques also have a number of potential disadvantages: possible inconsistencies across interviews and among interviewers, possible lack of reliability in coding the interviews, and the smaller sample size that can be used for a given amount of time and resources.

Assessment in the disciplines, in short, seems to be an area of increasing national interest and concern. In order to target research and development efforts to best serve that interest, this paper surveys the status of current state and district level practice in content assessment, highlights related research efforts currently underway, and identifies high priority areas for subsequent research.

Status of Current Practice

In an effort to determine what states and local education agencies are doing in response to this mandate and to guide the

direction of future research on assessment techniques, a needs
assessment for research in content area assessment was conducted
during 1986 by the Center for Research on Evaluation, Standards,
and Student Testing (CRESST). The results at the the district
and state levels follow.

District Level Efforts

District level administrators, principally directors of
research, evaluation and testing who are members of the Test
Directors of the Council of Great City Schools, were surveyed
during the spring and summer of 1986 regarding their districts'
top priorities for testing in the content areas. Two groups of
administrators, one with representatives from across the nation
and one with representatives from throughout California, provided
information in this survey.

The districts represented in the national group ranged in
size from approximately 14,000 to 570,000 pupils and tended to be
mainly urban or suburban districts. Districts were located in the
following states: Arizona, California, Connecticut, Florida,
Illinois, Minnesota, New Jersey, New York, Oregon, Pennsylvania,
and Texas. The districts in the California-only group ranged in
size from 8,000 to 20,000 pupils and included some rural areas in
addition to the urban and suburban areas spread across the state.
The two groups of respondents, national and California-only, were
rather similar in their responses.

All respondents were asked the following questions:

o What are your district's top priorities for testing in the
   content areas?

o At what grade levels are district assessments currently

conducted and/or anticipated within the next three years?

o What is the source of the tests that are given or planned (district developed, commercial off-the-shelf or custom)?

o Who are the primary users of the test results and what are the primary purposes for which the results will be used?

o What are the most important problems you anticipate or have encountered in this area related to test development, analysis reporting and/or use?

o What is the most important research and development that could be done in support of better assessment in these priority content areas?

The content areas listed on the survey were: art, business education, fitness/health, foreign language, literature, advanced math, music, science, social studies, and other. A summary of survey responses is reported below.

What are your district's top priorities for assessment among the content areas? Social studies and science were clearly identified by the majority of both California and national respondents as their district's top two priority areas for testing. About two-thirds of the administrators in the national group and about four-fifths of the California administrators ranked social studies and science as first or second priority.

The only other content areas ranked fourth or higher by several respondents were literature, advanced math, and foreign language. Several additional areas were written in by one or two administrators as having high priority in their districts: writing, critical thinking, language arts, reading, math, and basic skills.

In what grades does assessment occur and where is testing anticipated? Both science and social studies are currently assessed by about half the districts represented. The

distribution of grades in which these two content areas are
tested follow the same pattern in both the California and
national groups: most assessment occurs in grades 5-12, with a
peak in grade 10; little assessment is reported as occurring in
grades 1-4.

About half of the respondents anticipate expanding their
testing of social studies and/or science to more grades than are
currently tested. The districts that do not currently test in
social studies or science at all tend to anticipate testing in
approximately grades 7-12. The districts that currently do some
testing expect to expand their testing in both directions, such
as from currently testing in grades 5 and 7 to eventually testing
in grades 3,5,6,8, and 10.

What is the source of current and planned tests? Locally
produced tests, as opposed to commercial tests off-the-shelf or
custom-tailored, are used or planned in the great majority of
districts, especially in both science and social studies.

Who are the primary users and what are the primary uses of
the tests in the top two priority areas (science and social
studies)? The most frequently cited users of science and social
studies test data are district administrators, school
administrators and teachers. The primary uses of the data are
for curriculum planning, program evaluation, and diagnosis and
remediation.

What are the most important problems encountered or
anticipated in assessment of the priority content areas? The
primary problems cited by districts in mounting valid and useful

8

assessments in the content areas include:

    o Specification of outcome objectives and/or
      articulation of curriculum as basis for test design:

        - objectives often too broadly stated

        - philosophical differences in domain definitions

        - lack of consensus in identifying essential skills

        - isolating unidimensional traits

    o Development of test items

        - matching test item sets to new curriculum

        - shortage of good test items or approaches to
          developing them

        - assessing higher level skills

    o Shortage of resources

        - expense of changes in tests due to curriculum changes

        - lack of time, money and expertise for local test
          development

    o Establishing reliability and validity of measures

    o Lack of good data management system

    o Teacher resistance

What are the most important research and development ideas
that could be done in support of better assessment in the
priority areas? The following research and development ideas
were reported by districts as those most needed to support better
assessment in social studies and science:

1. New Approaches to Measurement:

    o Non-traditional assessment of student progress;
      alternatives to multiple choice items

    o Methods for measuring higher order skills, rather than
      just facts and information; help local professionals to
      develop such items

o Identification of benchmark items to allow for assesssment brevity

2. Item Banks and Clearinghouse:

    o List of available item banks in public domain or commercially available

    o Item banks with related item specification information

    o Standards for item banks, such as calibration of items, types of item statistics reported, measures of item sensitivity to context, and so forth

    o Computer storage and retrieval of items including appropriate high quality graphics

3. Integration of Curriculum, Instruction, and Cognition:

    o Analysis of core skills and concepts within particular content areas, particularly as they relate to learning and development

    o Development of stronger models of curriculum-based test development

    o Use of content area testing to improve instruction as well as standardize curriculum

    o The role of computers in improving testing, including computer adaptive testing, integrating instruction and assessment, and diagnostic measurement.

4. Training:

    o Help parents and teachers to use and understand test scores

    o Train district personnel to train others in district in test use and development.

State Level Efforts

There has been little recent state level assessment of content areas other than NAEP, and that which exists appears rather limited in scope and technique. For example, a review of state testing programs made by the Center for the Study of Evaluation's Quality Indicators project (Burstein, Baker & Aschbacher, 1985) found that only a few states assessed social

studies or science at that time, and most of these used a commercially available standardized test, the Comprehensive Test of Basic Skills (CTBS), which assesses a limited scope of content and skills via multiple-choice items.

In concern over the situation, the Council of Chief State School Officers are in the process of implementing a plan to have comparable indicators for measuring student achievement of both basic and higher order concepts and skills in science and social studies (as well as in math, reading and English) by the 1988-89 school year (Selden, 1986).

A survey of state directors of research and assessment by CSTES in 1986 confirmed that science and social studies are top priority areas for current and anticipated testing for purposes of accountability, curriculum planning, and student diagnosis. States' assessment problems and concerns focused on what to test, how to test it, and how to integrate curriculum with the various testing programs. For example, respondents mentioned the following:

o coordination with NAEP efforts (e.g., maintaining consistency with NAEP; determining the smallest number of NAEP items in each priority area that can be used to link with NAEP national results)

o selecting state-level measures that reflect individual district programs

o developing measures that reflect business concerns about student performance

o determining an appropriate mix of item types (e.g. multiple-choice versus production)

o getting adequate resources to develop performance tests (especially open-ended tests)

Related Research and Development

Against the background of field-based problems are a limited number of R&D projects which are currently being conducted. Most of the current research on the content areas focuses on issues in learning, instruction, and curriculum rather than precisely on assessment issues. However, we can look to this research for several types of assistance in our work on assessment.

In considering what is appropriate to test and how to define the relevant domains, we can consider the domains of knowledge used in current research. In addition, the analyses of learning processes and hierarchies in current research may suggest relevant subobjectives to assess. Comparisons of more and less successful learners may help us in yet another way: to identify useful distractors for multiple choice items or scoring guidelines for student discourse items.

The issue of how best to assess learning in the content areas is addressed indirectly by a number of researchers in that their empirical studies require some form of assessment, and these techniques -- be they interview, think-aloud, essay, multiple choice, short discourse, or other -- together with their relative success, will provide us with useful data on which to base development of future measures. This variety of techniques should provide information on which types of assessment strategies might be used successfully in moderate or large scale assessments, and which may be primarily instrumental in defining domains, identifying appropriate distractors, and constructing scoring guidelines to be used with other assessment strategies. A brief summary of several related R & D efforts follows.

Field-based research also may be able to provide some

guidance in a third area of concern -- how to facilitate the use of new content area measures. Research on barriers to curriculum implementation should be particularly germane.

The Council of Chief State School Officers, as mentioned above, is currently in the process of developing a state-by-state program to assess student achievement in several areas. As part of this endeavor, they are currently attempting to resolve several issues: articulating the core subject matter domains to be agreed upon across the states as the basis for developing instruments; selecting a scale by which results of the assessment program are analyzed and reported; and coordinating the administration of the program across states. These plans should have implications for deciding what to test and how to coordinate and facilitate use of new measures.

Recent successful efforts to judge student essays in the assessment of writing skills (Quellmalz & Burry, 1983) is expected to provide a basis for developing similar techniques to assess higher order skills via student discourse in the content areas.

The University of Pittsburgh is currently undertaking several related research studies within its Social Studies Learning Research Program (Glaser, Resnick & Thompson, 1985). Their focus is primarily on curriculum and instruction, but their research is also related to assessment in several ways: identifying what is taught, understanding higher order skills such as problem solving and reasoning, and understanding the effect on learning of the student's expectations about the testing situation.

McKeown and Beck, in their instructional practices study, are examining the content and structure of social studies basal texts and teachers' manuals to produce a fine-grained description of the content of third through eighth grade basal social studies programs and to provide an in-depth understanding of current practice. Based on their findings they plan to revise problematic instructional elements to enhance learning. Their results could prove useful in identifying appropriate domains and developing precise domain descriptions.

A project on problem solving and reasoning in the social sciences, directed by Voss, has three goals: to develop a better understanding of the nature of problem solving and reasoning found in the social sciences; to determine the processes by which students learn to solve problems and learn to reason in the context of social science; and to determine the processes by which students are able to evaluate their own and others' problem solutions and reasoning.

Voss and his colleagues use a model of effective problem solving in the social sciences involving knowledge of the subject matter, the ability to retrieve and organize that knowledge in the context of the problem at hand, and knowledge of what constitutes a high-quality solution and the ability to evaluate a solution. This model suggests types of appropriate test items to be developed, such as organization of given facts in the context of a specific problem, and critiqing given solutions to specific problems, in addition to the traditional items on factual recall.

An empirical study planned by Voss and his colleagues involves three types of tests of students' problem solving: a

multiple choice test of contents, retrieval of text contents in a hierarchical structure via an essay exam, and problem solving and reasoning via written discourse. A transfer task will determine the extent to which particular instructional and testing procedures that a student experienced are related to the criterion transfer performance. Although the emphasis of the study is on identifying instructional procedures that will enhance the ability to evaluate problem solutions, arguments, and reasoning, the study should also have implications for the value and construction of these three types of test items.

Further direction is given to assessment by Voss (in press), Chi (in press) and Hirsch (1985a, b), who agree that attempting to teach problem solving by providing practice in the use of "content-free" strategies cannot be expected to be very successful. Instead, problem-solving exercises in the context of specific and well-developed knowledge domains are more likely to be fruitful, and assessment should follow suit. That is, it makes little sense to speak of testing higher order skills divorced from specific content or without regard to the nature of the content knowledge to which they are to be applied.

The Science Learning Research Program at the University of Pittsburgh also includes several studies that may provide input for assessment of the content areas. Glaser and his colleagues are investigating forms of science instruction in which the learning of content knowledge is tied closely to reasoning. As part of this study, they are comparing the inferencing procedures and errors of more and less successful students. This aspect of

the study may provide useful information for the development of science assessment items, particularly in specifying appropriate distractors for multiple choice items and in specifying what is to be demonstrated in performance items. It may also be possible to begin to chart a hierarchy or sequence of concept attainment and reasoning development that might prove useful in diagnostic measurement. The study's own assessment strategies for assessing transfer of inductive reasoning skills may provide useful experience in computer-based testing.

A study by Champagne is focusing on the development of qualitative understanding of a set of generic relational concepts so pervasive in physical (and some social) science that mastery of them in one context can be expected to produce important gains in learning of other aspects of science and in general scientific reasoning. The core concepts used in this study, together with verbal and operational definitions and examples, may help us to define the pool of potentially important concepts to be comparably addressed across districts or states.

Chi is studying how successful students analyze, relate and integrate information rather than simply memorizing it. Her planned development of a taxonomy to characterize the nature of the contents of elaborations and inferences that good science learners make as they study may provide us with useful information in specifying subobjectives to assess.

The National Center on Effective Secondary Schools at the University of Wisconsin currently has a few projects that may also prove useful in specifying content domains and developing assessment techniques in the content areas (Newmann, 1986).

A project directed by Newmann entails analyses of
conventional and non-traditional testing in English and social
studies. By comprehensively documenting what is currently done,
thereby indicating (a) current organization and definition of
content and (b) gaps in content domains and testing methods,
their results will further enlighten areas of relative strength
and weakness in current test development practice.

Another project, directed by Newmann, Marrett and Schrag, is
comprised of several studies that are synthesizing research
related to higher order thinking and of an empirical study of the
teaching of thinking in five high school social studies
departments which emphasize this topic and serve a diverse range
of students. This project should provide useful input on what is
appropriate to test given adolescents' capacity to think. In
addition, the project's findings regarding barriers to the
implementation of higher order thinking curricula may generalize
to implementation of associated asses nent. This information may
imply modifications of assessment measures and techniques in
order to facilitate their use (such as format, content, length,
scheduling, roles of teachers and administrators in asssessment,
and so forth).

Future Efforts

Improved assessment of student learning in the content areas,
primarily science and social studies, is a clear and compelling
concern of many educators, researchers, and policymakers today.
Deep understanding of content, in particular, is inadequately
assessed by most current multiple choice tests. They simply do

17
21

not capture what we need to know about what and how students are learning. Better multiple choice tests that successfully assess higher order thinking skills as well as viable alternative testing techniques must be developed. To remedy the situation, future research and development efforts need to help identify what critical ideas and processes should be formally assessed and help determine how these can best be measured. The problem is a dual one: better assessing students' knowledge base in specific content areas and better assessing their thinking skills in applying, analyzing, synthesizing, and evaluating that knowledge.

Identifying, with consensus, which facts, concepts, and processes should be assessed will always be problematic to some extent because authorities in the sciences and social studies disagree, and the fields are subject to change. Nonetheless, this problem should not dissuade us from attempting to describe those constructs and processes in each field that appear most reasonable to assess formally, based on ethically and pragmatically defensible grounds. Research in this area will require extensive conceptualization and analysis with input from a wide array of content area specialists and others.

Embedded in the problem of deciding what to assess is the need to explore the balance between general constructs and processes and those which are tightly specific to a given discipline or topic. In seeking domains to test, we might want to consider including relevant concepts that may be learned, at least by some students, more as a function of classroom dynamics or total school environment than of intended classroom instruction. Examples include the concepts of power, demand and

shortage.

The second critical issue for future research to address is how best to assess the targeted constructs and processes. The plethora of multiple choice tests of factual recall suggests that future research be directed towards illuminating the possibilities and relative effectiveness of a variety of other techniques to assess the higher order thinking skills.

Several lines of research on cognition within the last ten years have produced learning models and evaluation techniques that may be able to provide new directions to the measurement of higher order skills in the social sciences, e.g., semantic networks (Dansereau & Holley, 1982), summaries (Wittrock, 1981), concept maps (Novak, et al., 1983). Up to now, such techniques were developed and used primarily for teaching and assessing knowledge structures learned in math and science courses, which may be characterized by well-defined problems. The challenge is now to apply such methods to areas such as social sciences where the objectives are diffuse and the problems ill-defined. Strategies derived from knowledge representation techniques such as concept mapping and summaries may provide criteria for deep understanding of concepts.

Recent work in the effective analysis of student essays (Quellmalz & Burry, 1983) should provide a basis for the development of methods to judge student discourse in the content areas. Future research will have to define and examine possible criteria such as accuracy, completeness, subtlety, and errors. Think-aloud and interview techniques used in some research may

19

suggest useful scoring criteria.

In addition to research on new methods of measurement, the
realities of large scale assessment programs necessitate future
research on improving multiple choice tests. To describe student
capabilities well, we need to base our measures on precise
specifications of elements of knowledge (facts, processes, and
skills) that cover the full range of important, relevant content
in the curricula. In addition, items must adequately sample
elements within the selected domains, have high reliability and
have good predictive validity. Work on alternative testing
methods, such as semantic network analysis of student prose,
ought to help generate domain specifications and diagnostic
distractors to use in improved multiple choice instruments.

A third issue towards which some future research and
development should be directed is determining how to facilitate
the use of new content area measures. A number of district and
state level educational administrators spoke directly to this
issue in the CSTES survey when they cited the following needs to
be addressed by future research and development:

  o a clearinghouse of information on assessment in the
    content areas

  o items banks

  o good data management systems

  o training (for teachers, administrators and parents) in the
    development, use and/or understanding of new measures

  o methods to overcome or finesse teacher resistance.

Without attention to this issue, potentially useful innovations
may never be widely accepted or integrated into the curriculum
assessment system.

20

24

REFERENCES

Burns, J. C., Okey, J. R. & Wise, K. C. (1985). Development of an integrated process skill test: TIPS II. Journal of Research in Science Teaching, 22, 169-177.

Burstein, L., Baker, E. L., Aschbacher, P. & Keesling, J. W. (1985). Using state test data for national indicators of educational quality: A feasibility study. Los Angeles, CA: UCLA Center for the Study of Evaluation.

Chi, M. T. H. (in press). Interactive roles of knowledge and strategies in the development of organized sorting and recall. In S. F. Chipman, J. W. Segal & R. Glaser, (Eds.), Thinking and learning skills: Vol 2 Research and open questions Hillsdale, NJ: Erlbaum.

Costa, A. L. (Ed.) (1985). Developing minds: A resource book for teaching thinking. Roseville, CA: Association for Supervision and Curriculum Development.

Curtis, M. E. & Glaser, R. (1983). Reading theory and the assessment of reading skill. Journal of Educational Measurement, 20, 133-147.

Dansereau, D. F. & Holley, C. D. (1982). Development and evaluation of a text mapping strategy. In A. Flammer and W. Kintsch (Eds.), Discourse processing. Amsterdam: North-Holland Publishing Company.

Education Commission on the States (1982). The information society: Are high school graduates ready? Denver: CO.

Finegold, M. & Mass, R. (1985). Differences in the processes of solving physics problems between good physics problem solvers and poor physics problem solvers. Research in Science and Technological Education, 3, 59-67.

Fredrikson, N. (1984). The real test bias: Influences of testing on teaching and learning. American Psychologist, 39:3, 193-202.

Gabel, D. L., Sherwood, R. D. & Enochs, L. (1984). Problem-solving skills of high school chemistry students. Journal of Research in Science Teaching, 21, 221-233.

Glaser, R., Resnick, L. B. & Thompson, J. H. (Eds.) (1985). echnical Proposal for the Center for the Study of Learning. Pittsburgh: Learning Research & Development Center, University of Pittsburgh.

Goodlad, J. (1983). A place called school. New York: McGraw Hill.

Haertel, E. & Calfee, R. (1983). School achievement: Thinking about what to test. Journal of Educational Measurement. 20(2), 119-132.

Hirsch, E. D. (1985a, Summer). Cultural literacy and the schools. American Educator, 8-14.

Hirsch (1985b, October). Cultural literacy doesn't mean "core curriculum." English Journal, 47-49.

Lehman, J. D., Carter, C. & Kahle, J. B. (1985). Concept mapping, vee mapping, and achievement: Results of a field study with black high school students. Journal of Research in Science Teaching, 22, 663-673.

Linn, R. (1983). Testing and instruction: Links and distinction. Journal of Educational Measurement, 20(2), 179-190.

McClelland (1973). Testing for competence rather than for intelligence. American Psychologist, 28, 1-12.

McTighe, J. & Schollenberger, J. (1985). Why teach thinking: A statement of rationale. In A. L. Costa (Ed.) Developing minds: A resource book for teaching thinking. Roseville, CA: Association for Supervision and Curriculum Development.

Morgenstern, C. F. & Renner, J. W. (1984). Measuring thinking with standardized science tests. Journal of Research in Science Teaching, 21, 639-648.

Newmann, F. (1986). Application for continuation: The program of activities. Madison, WI: National Center on Effective Secondary Schools, University of Wisconsin.

Novak, J. D., Gowin, D. B. & Johansen, G. T. (1983). The use of concept mapping and knowledge vee mapping with junior high school students. Science Education, 67, 625-645.

Olson, L. (1984, November 28). Social studies: Amid criticism, still in search of a clear rationale. Education Week, pp. 1, 14.

Quellmalz, E. S. & Burry, J. (1983). Analytic scales for assessing students' expository and narrative writing skills. Center for the Study of Evaluation Resource Paper No. 5. Los Angeles, CA: Center for the Study of Evaluation.

Resnick, L. B. (1983). Mathematics and science learning: A new conception. Science, 220, 477-478.

Ross, J. A. & Maynes, F. J. (1983). Multiple choice tests for seven science process skills. Journal of Research in Science Teaching, 20 (1), 63-75.

Selden, R. (1986). Strategies and issues in the development of comparable indicators for measuring student achievement. State Education Assessment Center: Council of Chief State School Officers.

Shavelson, R.J. (1985a). Evaluation of nonformal education programs: The applicability and utility of the criterion-sampling approach. Oxford: Pergamon Press.

Shavelson, R.J. (1985b, April). The measurement of cognitive structure. Paper presented at the annual meeting of the American Educational Research Association, Symposium on the "Psychology of Learning Science."

Texley, J. & Norman, J. (1984). The development of a test of formal operational logic in the content area of environmental science. Journal of Research in Science Teaching, 21, 589-597.

Voss, J. F. (in press). Psychology and instruction in the social sciences. In R. F. Dillon & R. J. Sternberg (Eds.) Psychology and curriculum design. Hillsdale, NJ: Erlbaum.

Wittrock, M. C. (1981). Reading comprehension. In F. J. Pirozzolo & M. C. Wittrock (Eds.), Neurophychological and cognitiv3e process in reading. New York: Academic Press.

Yarroch, W. L. (1985). Student understanding of chemical equation balancing. Journal of Research in Science Teaching, 22, 449-489.