DOCUMENT RESUME

ED 272 583 TM 860 545

AUTHOR Kearney, C. Philip; And Others

TITLE Assessing Higher Order Thinking Skills.
INSTITUTION ERIC Clearinghouse on Tests, Measurement, and Evaluation, Princeton, N.J.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
REPORT NO ERIC-TME-R-90
PUB DATE Apr 86
CONTRACT 400-83-0015
NOTE 67p.
AVAILABLE FROM ERIC Clearinghouse on Tests, Measurement, and Evaluation, Educational Testing Service, Princeton, NJ 08541 ($7.50).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Information Analyses - ERIC Information Analysis Products (071)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Academic Achievement; *Cognitive Ability; *Cognitive Measurement; *Cognitive Processes; Creative Thinking; Critical Thinking; Decision Making; *Educational Assessment; Educational Policy; Educational Trends; Elementary Secondary Education; Logical Thinking; *Measurement Techniques; Problem Solving; State Programs; *Teaching Methods; Time Management
IDENTIFIERS Blooms Taxonomy; Connecticut; Michigan; *Thinking Skills

ABSTRACT

Four presentations from the 1985 Annual Michigan School Testing Conference on "Assessing Higher Order Skills" are offered in this paper, and the chairman of the First General Session provides an introductory section. The papers individually and collectively address the problem of defining higher order thinking skills. A second major question facing those interested in teaching and testing thinking skills involves whether such skills should be taught and tested as a separate subject area or embedded and infused in existing subject matter. The paper by Michael H. Kean offers a concise treatment of the major questions facing those who would embark on the teaching and testing of higher order thinking skills. Edward D. Roeber and Betty L. Stevens describe the activities in Michigan during the planning and development stage for testing higher order skills, and outline the alternative approaches being considered by state level decision makers. Joan Boykoff Baron's paper provides an analysis of Connecticut's experiences in implementing a higher order thinking skills component in an ongoing assessment program. John Fremer and Mark Daniel provide a recapitulation of problems and prospects through a discussion of several recent developments in the assessment of higher order thinking skills. (LMO)

ASSESSING HIGHER ORDER THINKING SKILLS

TME REPORT 90

ERIC®

TESTS MEASUREMENT EVALUATION

EDUCATIONAL TESTING SERVICE
PRINCETON, NEW JERSEY 08541-0001

ERIC/TME Report 90

ASSESSING HIGHER ORDER THINKING SKILLS

by

C. Philip Kearney
The University of Michigan

Michael H. Kean
CTB/McGraw Hill

Edward D. Roeber, and
Betty L. Stevens
Michigan Department of Education

Joan Boykoff Baron
Connecticut State Department of Education

John Fremer, and
Mark Daniel
The Psychological Corporation

This report is based on papers presented at the 1985 Annual Michigan School Testing Conference held in Ann Arbor, Michigan during February, 1985.

The material in this publication was prepared pursuant to a contract with the Office of Educational Research and Improvement, U.S. Department of Education. Contractors undertaking such projects under government sponsorship are encouraged to express freely their judgment in professional and technical matters. Prior to publication, the manuscript was submitted to qualified professionals for critical review and determination of professional competence. This publication has met such standards. Points of view or opinions, however, do not necessarily represent the official view or opinions of either these reviewers or the Office of Educational Research and Improvement.

ERIC Clearinghouse on Tests, Measurement, and Evaluation
Educational Testing Service
Princeton, NJ 08541

This publication was prepared with funding from the Office of Educational Research and Improvement, U.S. Department of Education under contract No. NIE-400-83-0015. The opinions expressed in this report do not necessarily reflect the positions or policies of OERI or the Department of Education.

TABLE OF CONTENTS

WHAT'S GOING ON IN THE ASSESSMENT OF HIGHER ORDER SKILLS?

C. Philip Kearney
The University of Michigan

A simple and straight-forward answer is, "A great deal!" As Fremer and Daniel, the authors of the fourth paper in this collection, point out, concern with the teaching and testing of higher order thinking skills is fast taking on the characteristics of a major educational reform movement. Several states are developing and implementing assessment programs aimed at higher order thinking skills, textbook publishers and testing companies are becoming increasingly active in this arena, and conferences centered on this topic are springing up across the Nation. One of these conferences was the 1985 Annual Michigan School Testing Conference which took as its theme, "Assessing Higher Order Skills." The First General Session of the Conference was built around four presentations which addressed the title question of this piece, "What's Going on in the Assessment of Higher Order Skills?" ERIC/TME felt that these four presentations merited a wider audience and, consequently, asked the four presenters to prepare for inclusion in this present ERIC/TME publication.* Because I chaired the session, I was asked to prepare this brief introductory piece.

The four papers, properly, do not attempt to provide definitive answers on what constitutes so-called higher order thinking skills, on how they should be taught, or on how they should be assessed. But the papers do offer a base of

information from which the reader can begin to form her or his own tentative answers to these questions.

The papers individually and collectively address, but do not resolve, the first major problem facing those interested and involved in this arena, namely, the problem of defining higher order thinking skills. As each of the authors implicitly or explicitly demonstrates, there is no firm consensus on what should be included or excluded under the higher order thinking skills rubric. For the parent, as Kean suggests, the answer is easy: "What I want is for you to teach my child to think." For the professional, the answer is much more complex. It includes such notions as a habit of reflective thinking: a disposition or willingness to think critically, assertively, and habitually; more difficult subject matter content; critical reasoning skills; skills that go beyond straight recall or learning of facts; and a literal laundry list of other cognitive activities.

Neither do the papers offer a definitive resolution of a second major question facing those interested in teaching and testing higher order thinking skills, namely, whether they should be taught and tested as a separate subject area or embedded and infused in existing subject matter and tested in like fashion. While the papers appear to have a bias toward the embedded and infused approach, it appears still to be a question lacking a clear cut answer. Kean strongly advocates embedding thinking skills in every subject. Baron tells us that Connecticut has embraced both approaches in its assessment

efforts. Roeber tells us that Michigan has yet to resolve the question completely. Fremer and Daniel point out that there are still clear differences of opinion on the question in the instructional and measurement communities.

The reader also will become aware of a number of other questions facing those who would develop programs to assess higher order thinking skills, including whether to use a "one-tiered" or "two-tiered" approach in fashioning the program, whether the benefits of using multiple approaches to measuring these skills outweigh the costs, whether every-pupil testing or matrix sampling is called for, whether there is a need for considerable test development work, or whether a number of instruments that could be used in measuring these skills is already available.

The paper by Roeber nicely summarizes the basic differences in the answers being provided to these questions by those in Michigan who advocate the teaching and testing of higher order thinking skills. Roeber's paper also offers a picture of what is going on in a State which, while it has had a state assessment program for a number of years, is only now setting out in a systematic way to include higher order thinking skills in its assessment program. Baron's paper capitalizes on Connecticut's experiences with the inclusion of higher order skills in a state assessment program and shares with the reader the lessons learned from those experiences. Taken together, the two papers offer succinct descriptions of what's going on in the

assessment of higher order skills in two states.

Kean's paper and the paper by Fremer and Daniel provide the reader insights gained from persons vitally interested in the teaching and testing of higher order thinking skills because of their current roles with major test publishing firms, as well as their ongoing roles as members of the professional measurement community. Their experiences in working with instructional and measurement practitioners charged with developing large-scale assessment programs lend a practical flavor to their views on this important topic.

We suggest that the reader read the four pieces in the order that they are presented. Kean's paper, in our view, offers a concise treatment of the major questions facing those who would embark on the teaching and testing of higher order thinking skills. Roeber's paper describes the activities of a State still in the planning and development stage, and the nature of the alternative approaches being considered by state level decision makers. Baron's paper provides the reader benefit of Connecticut's experiences in implementing a higher order thinking skills component in its ongoing assessment program and the lessons that were learned from those experiences. Fremer's and Daniel's paper, while not necessarily written for that purpose, provides a good recapitulation of problems and prospects through its discussion of several recent developments in the assessment of higher order thinking skills.

As we stated above, the reader is not offered definitive answers to the questions of what constitutes higher order thinking skills, of how they should be taught, or of how they should be tested. The careful reader, however, is offered a solid base of information from which she or he can draw some tentative--and we would stress tentative--answers to these questions.

*Thomas H. Fisher, Director, Student Assessment Program, Florida State Department of Education, was one of the four presenters at the 1985 Annual Michigan School Testing Conference. Unfortunately, because of other demands, he was not able to prepare a paper on his presentation for inclusion in the present collection. Edward Roeber, who is immediately responsible for Michigan's efforts in this area, graciously consented to fill in for Dr. Fisher and prepared a paper describing Michigan's current activities in the assessment of higher order thinking skills.

ASSESSING HIGHER ORDER THINKING SKILLS: AN OVERVIEW OF THE ISSUES

Michael H. Kean
CTB/McGraw-Hill

In discussing higher order thinking skills, I plan to address:

1. What they are
2. How they might be taught and measured, and
3. Whether all the attention being paid to them will result in changes of substance in education or result in just another passing fad.

Before I do all that, though, I should like to reference several comments that I think are pertinent to the issues at hand. The first is from Bill Honig, the California State Superintendent of Public Instruction, and has to do with how we arrived at what now seems to be a crisis in the teaching of higher order skills.

Dr. Honig says:

> In the '60s and '70s, we told kids, "you make up your mind as to what's relevant and fun and study that." That was an abdication of our role as educators. Then, when people didn't think kids were learning anything, we went back to basics. The public never misinterpreted what back to basics meant--history, literature, science, writing, high expectations, homework, order in the classroom--but educators did. What educators did was narrow the curriculum down to basic skills.

And what was the result of that narrowing? Ray Cortines, the Superintendent of the San Jose (California) Unified School District, characterized it rather nicely, when he stated: "With the return to basics, we screwed off the kids' heads, poured in the information, and asked them to regurgitate the information by asking questions at the end of the week. But we didn't teach them how to use that information."

Public expectations are difficult to gauge. When we taught students what they said they wanted to learn, the public wasn't happy. When we taught the students what we thought the public said we should be teaching, it turned out not to be sufficient. Now we are being asked to teach something called "higher order skills."

1. What are higher order skills?

Following is a brief list of some of the skills and attributes that various authorities have identified as constituting higher order thinking skills:

- o Comparing and contrasting
- o Making inferences
- o Analyzing events
- o Synthesizing information
- o Drawing conclusions
- o Identifying the problem
- o Analyzing the problem
- o Suggesting possible solutions to the problem
- o Testing consequences of possible solutions
- o Assessing the reliability, relevance, sufficiency, validity, and meaning of data
- o Analyzing arguments
- o Judging credibility of sources
- o Observing and judging observations and reports
- o Induction

- Deduction
- Assumption identification
- Prediction
- Identification of fallacies
- Definition of problem
- Distinguish between differences of kind and differences of degree
- Understanding verbal analogies
- Selection of a solution process
- Selection of a way of representing a solution
- Selection of a problem-solving strategy
- Allocation of processing time
- Sensitivity to feedback
- Translation of feedback into an action plan
- Implementation of an action plan
- Testing hypotheses
- Linear reasoning
- Data gathering
- Decision making
- Classifying
- Organizing
- Identifying alternative points of view
- Recalling
- Grouping/labeling
- Classifying/categorizing
- Ordering
- Patterning
- Prioritizing

The list is even longer, but I do not think the point needs to be belabored: there is a certain lack of consensus among educators as to what higher order skills are.

Probably the average lay parent would have less trouble defining what she or he thinks should go on in public schools. "What I want," a parent might say, "is for you to teach my child to <u>think</u>."

What that average parent might not say, but what they would almost surely <u>also</u> want, is for the child to be taught to think <u>critically</u>, <u>assertively</u>, and <u>habitually</u>; that is to be a thinking being, not just a pliant subject capable of displaying certain behaviors on cue in an academic environment.

Harvey Siegel, Professor of History and Philosophy of Education at the University of Nebraska, had some interesting things to say about critical thinking in the November 1980 issue of <u>The Educational Forum</u>:

> "...it is not enough for a student to be able to evaluate claims on the basis of evidence...In order to be a critical thinker, a student must be <u>disposed</u> to do so. A critical thinker must have a <u>willingness</u> to conform judgement to principle, not simply an ability to so conform."

In the same article, Dr. Siegel says that students have a "right to question, to challenge, and to demand reasons and justifications for what is being taught." Those two quotations have some interesting implications.

The first suggests that the apparent failure of our schools to produce thinking beings may have at least as much to do with the

general environment they provide, as with the specific curricula they teach; for surely "disposition" and "willingness" are not explicitly taught commodities. I'll return to that point shortly, but would first like to examine some of the implications of that second quotation, the one about the students' right to question, to challenge, and to demand.

I want to suggest that there may be less of a constituency out there in the world at large, and even within the educational community, for rational, thinking beings, than we as educators might like to believe. If we succeed in teaching students to think, we cannot expect that they will limit their thinking to prescribed subject matter. We must expect, rather, that they will question us and challenge us and demand of us that we justify our positions on any number of issues from curriculum content to dress codes. Thinking students can, in short, be very inconvenient students.

If we as educators, whose business it is to train young minds, are not entirely sure we want to deal with rational beings, how much more likely is it that very considerable segments of society at large may in fact be angry rather than grateful if we should ever succeed in graduating a generation of truly rational students? I do not mean to be overly negative; nor do I mean to suggest that efforts to improve students' thinking skills are either undesirable or impossible. Quite the contrary--if there is in fact some degree of anti-rational bias both in our education system and in society at large, it is all the more incumbent upon

us to find ways to teach students to overcome that bias. It is important, though, that we be honest with ourselves about what we are trying to accomplish. If we delude ourselves that we can teach higher order thinking skills as just another chunk of curriculum, to be drilled like the multiplication tables, our efforts will fail.

2. How then should thinking skills be taught?

At this point, I'd like to compare and contrast two subject areas that have been getting quite a bit of press lately: higher order skills and computer literacy. At the moment, both computer literacy and higher order skills curricula are rather trendy subjects. Both have many buzz words associated with them, and both have a certain air of newness.

Of the two, though, only computer literacy is genuinely new. No one has, to my knowledge, suggested that the public schools did a better job of teaching computer literacy a decade or a generation ago than they do now. The public fear, in connection with computer literacy, is that the schools may be failing to keep pace with brand new developments, not that they are becoming deficient at something that they used to do well.

Higher order thinking skills are an entirely different matter, however. It is suggested that a decade or a generation ago schools did a better job of teaching than they do now. And yet, I do not think that any large number of public schools ever explicitly taught thinking skills until quite recently. I don't

think it occurred to very many people that thinking skills needed to be taught.

It is a rare person indeed who can progress very far in computer literacy without at least some formal instruction. We see in the microcomputer a device with definite characteristics that must be explicitly learned. We do not blame ourselves if, in the absence of instruction, we are unable to make much use of computers.

Thinking, on the other hand, is something that most people do remarkably well without any formal instruction. That is not to say, of course, that we could not all improve our thinking skills with formal instruction; but it is to suggest that, in the case of thinking skills, we are dealing with something quite different from other subjects in our curricula.

If the schools of the past did not explicitly teach thinking skills, yet managed to turn out reasonably good thinkers, what did they do that the schools are not now doing? For one thing, they simply <u>existed</u> at a time when reason was held in higher repute than it now is. I cannot prove that, but it's worth considering. They also required a lot of writing, and writing is notorious as an instrument of thought.

However, schools of the present cannot, in a direct and immediate way, control the spirit of the times in which they must function. Writing, for all its utility as a tool of thought, cannot be expected by itself to overcome students' deficiencies in thinking skills. So we are left with the proposition that

something must be done to teach thinking skills in the public schools.

There are two fairly obvious ways to go about it. You can introduce into the curriculum a new subject with a new and trendy name, or you can embed the teaching of thinking skills throughout the existing curriculum. There is ample evidence in the literature that either approach can be made to work. There are several dangers in the first approach. For example, it is easy to overload the system itself. There is only time in the day for so many subjects; introducing a new one may cost an old one.

In addition, although there is evidence for the efficiency of teaching thinking by teaching <u>about</u> thinking, by making thinking itself a subject like English or math, the risk is run that teachers and students alike will treat thinking in the same unproductive ways that they have sometimes treated other subjects. The teacher will drill into the students' heads the fourteen steps of critical thinking, and the students will dutifully list those steps on the next quiz, without bothering ever to apply them to any other aspect of their lives, in or out of school. Finally, by isolating thinking skills as a separate item in the curriculum, you make them a likely target for the first "no frills" budget cutter who comes along.

By embedding the teaching of thinking skills in every subject, on the other hand:

1. You are likely to take less time away from subject-area studies;

2. You give students more opportunity to apply the thinking skills they learn in diverse situations;

3. You effectively forestall the possibility that thinking skills will be deleted from the curriculum the moment they are no longer a "hot item."

Embedding the teaching of thinking skills into every subject area also increases the likelihood that all students, regardless of achievement level, will benefit. Higher order skills should not be considered the special province of the gifted and talented.

Whatever method is used to place thinking skills into the curriculum, it is important that we not lose sight of environment and attitudes. No amount of explicit teaching of thinking skills will ever overcome implicit environmental clues telling students that independent thinking, far from being valued, is likely to get them into trouble. A teacher or an entire school system that is unwilling to entertain serious questions about its goals and methods, or allow open discussion of issues of importance raised by students--such as censorship of school book lists--is unlikely to produce a crop of questioning students. Thinking skills must not only be explicitly taught, they must be practiced; they must be exemplified in the behavior of teachers and administrators; and they must be valued in students. When these criteria are met, we may expect to see students disposed to evaluate claims on the basis of evidence and willing to conform judgment to principle.

Assuming then that we are agreed that higher order thinking skills can, at least in some degree, be taught, we come to the question of measurement.

3. Can thinking skills be tested?

There is no reason why they cannot. Though there are considerable differences between thinking skills and other skills taught in the schools, the fact remains that thinking skills, though perhaps not themselves observable, when exercised, produce observable outcomes; and observable outcomes can be measured.

Admittedly, not all authorities take the same view on the subject. In the March 1984 issue of Phi Delta Kappan, for instance, Barry K. Beyer stated that: "The best measure of students' ability to think may be their behavior as they sift through data to arrive at a conclusion or as they go about solving a problem. The development of instruments or observation techniques that can measure such behavior ought to be a major priority of test makers." There is no reason to believe, however, that standard multiple-choice items cannot be constructed in such a way that they can only be correctly answered by engaging in the kinds of higher order thinking skills that have been discussed. Why shouldn't students' ability to engage in those skills be assessed with existing instruments and with instruments that can be fairly readily produced?

Nevertheless, I think development of test instruments of the type Dr. Beyer advocates might be a very good thing indeed. Such instruments might well provide a more useful degree of diagnostic information than is presently available. Having the information that a student is deficient in, say, deduction, is of limited value if you do not know what actual subprocesses to attack in remedying the deficiency. I do not think it is either necessary

or advisable, however, for the education community to wait for instruments that may or may not soon be available, given that we are faced with a critical problem and already have at our disposal some useful tools with which to begin to attack that problem.

I would suggest, for example, a norm-referenced achievement test such as the California Achievement Tests (CAT), Forms E and F. Even though these tests are designed primarily to measure the most commonly taught basic skills, there are many items throughout the series that measure higher order thinking skills.

These items measure more than recall of facts or answering questions based on the information provided. The items were developed to require students to analyze, to synthesize and to interpret the information provided. Studies will be done during the standardization of CAT E and F to determine what kinds of valid scores or results can be reported on these items. In addition, CAT E and F has been developed so that there is a better probability that reliable and valid information can be obtained for higher scoring students. Additional items have been included at the difficult end of the range to minimize the chance of students "topping out." CAT E and F will also provide End-of-Course tests at the secondary level for students taking specific courses in algebra, geometry, physics, biology, world history, American history, computer literacy, and consumer skills.

While CAT E and F is still primarily designed as a measure of basic skills, procedures and information have been built in to also provide useful and valid information on higher order skills.

In Conclusion

To close, it may be useful to very briefly reiterate several of the points made earlier.

Is it necessary explicitly to teach higher order thinking skills? Given the mounting evidence that our students are deficient in such skills, I think the answer is clearly yes.

Is it, in fact, possible to teach and test such skills? Considerable evidence suggests that it is.

Is it sufficient explicitly to teach higher order thinking skills? Absolutely not! I think the education community needs to take a hard look at whether or not it provides an environment in which thinking skills, once acquired, can flourish. Providing that environment is in the long run at least as important as any formal teaching, testing, and remediation we can provide.

DEVELOPING MICHIGAN'S ASSESSMENT OF THINKING SKILLS: DEVELOPING MICHIGAN'S PROGRAM

Edward D. Roeber and Betty L. Stevens
Michigan Department of Education

In July 1984, the Michigan Department of Education was funded to investigate and plan a higher-level assessment program. Specifically, the Department budget bill included the following language: "...develop advanced skills tests for use in grades four, seven, and ten in the areas of language arts and mathematics..." Although Legislative intent was clearly to develop more difficult assessment tests, staff of the Michigan Educational Assessment Program (MEAP) have also explored the possibility of including tests of higher order thinking skills. The following is a description of the current MEAP status of and an examination of how it might be changed, what might be changed, what might be tested in the future, and issues which must be addressed. As with any developmental project, what emerges in a year or two may bear little resemblence to current plans.

The Current Assessment Program

The current MEAP program assesses all students in grades four, seven, and ten in the areas of mathematics and reading. This program has been in existence since 1969-70. Results of the MEAP Program are used to help students make up skill deficiencies, as well as to provide schools with a point of departure in reviewing and revising their curricula in these

areas. Scores of individual students are not used in promotion or graduation decisions. Over the years, scores on the tests have improved considerably, most notably in ' areas of reading.

Because results are reported in the newspapers, school personnel, parents, and the general public are very sensitive about information which may reflect negatively on individual schools or local districts. This concern often stimulates school districts to take steps to improve student performance by making changes in school programs. Staff of the Department (assessment, instruction, compensatory education, and so forth) spend a considerable amount of time assisting local districts to use the results appropriately, as well as to report them in a useful manner.

Forces For Change

A major force for change of MEAP, of course, has been the spate of reports on the condition of education nationally and in Michigan. A number of these have proposed using testing not only as a vehicle to monitor student achievement but also as stimulus for educational reform. In Michigan, for example, a special report (Sederburg & Rudman, 1984) was prepared that examined changes in performance for various subgroups of students, particularly at the high school level, where comparative data on students in Michigan and the nation is available using college-entrance tests such as the SAT. This

report was written in response to A Nation At Risk and Michigan State Board of Education plan for the future (A Blueprint for Action, 1984), which included recommendations made by the Michigan High School Commission. The following is taken from the summary of the Sederburg and Rudman report:

> Over the past few years, state and federal educational policy has targeted the lower achieving student. This targeting of funds and effort has yielded results. However, it is apparent that, at the same time, we may have neglected the better achieving student. In contrast to the prevailing belief, the brightest students have not succeeded regardless of the educational system.
>
> Consequently, we are calling for a shift in educational policy. We must create an educational system that challenges all young people and develops students to the best of their abilities. Emphasis on testing for basic skills for high school graduation and grade promotion reinforce the attitude that teachers and administrators should be most concerned with the lower achieving student. While it is worthwhile to insure that all students possess "essential" skills before graduation, we must not overlook the student who is not challenged by minimal objectives.
>
> The recent proposals made by the State Board of Education go a long way toward accomplishing the goals outlined here. However, the entire focus must be shifted away from minimal skills which tend to bring high achievers down while trying to bring everyone up to the highest level possible. The State Board and the legislature will need to clarify their philosophical direction as well as set specific goals for whatever educational reform they wish to achieve in the 1980's.

Proposals For Change

The Sederburg and Rudman paper contained the first proposals for developing a higher-level test. Although the State Board of Education's report included changes for the assessment program, such changes dealt only with broadening the scope of MEAP to

include periodic, every-pupil testing of other subject areas including Health, Science, Career Development, and Social Studies. The Sederburg-Rudman article, however, dealt specifically with higher-level assessment by suggesting, among other things, that:

1. The testing program of the State Board of Education should be changed to adequately measure all Michigan students, not just those below the achievement level determined by the State.

    a. The State Board should establish a qualified task force to develop such a testing program.

    b. The legislature should mandate this testing program through the budget of the Department of Education.

2. The State Board of Education set achievement goals to be attained by all achievement classifications by a specific date. In their "Blueprint for Action" the State Board calls on local boards to initiate a 3-5 year plan to improve achievement. Similarly, the Board should set state goals to improve all categories of Michigan youngsters.

3. State policy should reflect an effort to pressure local school districts to provide programming for the entire spectrum of students. The state testing program should be used to validate or accredit local school diplomas for all students.

    a. Achievement tests administered as early as the tenth grade should point to areas for potential remediation. The 10th grade test should emphasize reading, language and basic math skills.

    b. An 11th grade exam should include physical science, biological science, and social science. The 12th grade year would be used to assist students who did not meet essential skills in the 10th and 11th grade exams.

    c. The State Board of Education should use these tests as the basis for accrediting high school diplomas.

A response to the Sederburg and Rudman paper by the Department of Education suggested possible direction for the MEAP Program:

> The other way in which MEAP may change in coming years is to assess students beyond the basic skill level. This discussion presumes that (1) testing basic skills is valid and will still be carried out, (2) testing higher-level skills should emphasize the same purposes as the regular MEAP program (i.e., individual student assistance, curricula review and revision, reporting to various audiences), (3) students should be identified based on their basic skill achievement, (4) such higher-level skills are either more difficult subject matter content, critical reasoning skills or higher-level thinking skills (e.g., analysis, synthesis and evaluation from Bloom's Taxonomy), and 5) the students identified can be offered a school program which meets their educational needs, even as schools are helping students who have not as yet achieved the minimums. The presumption is that schools (and the State) can emphasize both "basic" skills and "advanced" skills and not have to choose one over the other (Roeber, 1984).

MEAP staff proposed a plan that included a two-tier approach, with all 4th, 7th, and 10th grade students taking the basic skill level and those that passed, the higher-level examination. It was proposed that advanced tests be developed at three levels (grades 4-6, given in seventh grade; grades 7-9, given in 10th grade; and grades 10-12, given in grades ten, eleven, and twelve). Staff also developed a list of technical and policy issues for testing beyond the basic skills.

The Department plan was presented to the State Board of Education in early 1985. After considerable discussion, the State Board approved the MEAP staff plan for a higher-level assessment program and directed that a study group be convened

to examine issues and to develop a tentative assessment plan.

Developing the Plan for The Higher Level Assessment Program

Since late 1984, Department staff have been meeting with a planning group consisting of local and intermediate district educators, college and university specialists and others. Represented on the group are gifted educators, assessment and curriculum specialists, content area specialists (e.g., science, reading), and administrators.

The Higher Level Assessment Committee has spent a considerable amount of time discussing methods to address student needs, particularly those of students who already pass the current basic skills tests. Very early in these discussions, it was apparent that there were sharp differences of opinion regarding the direction MEAP should take. Some members of the advisory group, for example, proposed toughening the current content standards tested in MEAP. Others, however, suggested that tests of critical thinking, critical reasoning, or thinking skills be used.

The group has been pursuing both options. Discussions have focused on what "tougher" standards really mean, how higher-order thinking could be tested and how this program could mesh with the current basic skills program. Others have been examining various approaches to teaching thinking skills, looking particularly at how thinking skills are defined and the implications for testing. While viewed originally as an

alternative to the current basic skill program (or, at least, a more difficult extension of it), thinking skills is now viewed as a logical complement to the current program, plus any new program which might be developed.

With this background in mind, the committee began to examine alternative approaches to the new assessment program. Members of the committee were challenged to develop a new assessment program model. Thus far, two plans have been suggested. The first (Rudman, 1985) is much different than the second (Downing, Johnson-Lewis, Leddick, Lohr, Stevens, 1985). Each is described more fully below.

The Rudman plan proposed a different approach than proposed earlier by Sederburg and Rudman. The new plan is predicted on seven assumptions:

1. The power of state-mandated assessment programs has been convincingly demonstrated to be a force in instituting instructional change within the schools.
2. Higher order reasoning skills can best be taught as an integral part of some specified body of knowledge.
3. There is a demonstrable relationship between focused instruction and student performance at all levels of ability.
4. Schools can be effective if the mandate given them is strong enough and if adequate resources are available.
5. There is a limit to the amount of resources--human and fiscal--that are available for education.
6. The schools are an important instrument in affecting social and economic policies of a nation.
7. Recommendations for reform of any institution, including the schools, must be based on a reasonable expectation of stability of public policy (pg. 25).

Rudman goes on to make four recommendations:

1. The State should develop a plan which incorporates a two-tier evaluation of student academic status;

2. The State should establish a standard setting advisory committee;

3. The State should assume a major portion of the funding for mandated assessments;

4. The State Department of Education should establish a technical advisory committee to determine test specifications, set criteria for selecting tests, and recommend tests or test contractors.

Rudman suggests that the current Michigan Educational Assessment Program should be "mandated on a matrix-sampling basis rather than an every-pupil requirement....Matrix sampling could yield useful information for public monitoring of minimal achievement within the state's schools while at the same time reducing the amount and testing time....He further suggests that this program be administered at grades 4, 7 and 9, with concentration on the first two levels of Bloom's taxonomy (Knowledge & Comprehension).

Rudman also recommends that a second tier of testing should be undertaken by the Michigan Department of Education. This testing should be mandated on an every-pupil basis in grades 1, 3, 5, 8, 10 and 11. The content of these tests should include much more than Reading and Mathematics. It should measure the language skills, social sciences, science, and listening skills at the appropriate grade levels. The content of these tests should consist of levels 1 to 4+ of the Bloom Taxonomy

(depending upon the grade level at which the test is administered). It should be so constructed that there are sufficient items at a variety of difficulty levels from .30-.90.

The other proposal under consideration by the Higher-Level Assessment Committee was prepared by a subgroup of the Committee (Downing, Johnson-Lewis, Leddick, Lohr and Stevens, 1985). This proposal is considerably different from the Rudman plan, in that it suggests that every pupil should be included in the same testing program.

Five basic assumptions underlie this approach. They are:

1. The Processing of knowledge is critical to the current information society, therefore the accumulation of information must be accompanied by increased emphasis on problem finding, problem solving, critical thinking and decision making.
2. For some students, employing higher order thinking may lead to more successful acquisition of basic skills.
3. A state mandated assessment program can and does drive curriculum in new directions.
4. Focused instruction results in acquisition of identified skills.
5. Test construction should not be attempted without consideration of program implementation and acceptance factors. Any new program must be built on what currently exists.

The subgroup has recommended that the State continue a one tier assessment program to evaluate student academic progress. Within this program, however, it is recommended that the existing assessment program be expanded to include:

1. "essential skills" as these not only subsume basic skills but can expand to include areas of greater

difficulty;

2. a writing component which focuses on higher order thinking;

3. an indicator test to assess content of specific skills involved in problem finding, problem solving, critical thinking and decision making.

In order to articulate the new MEAP for educators and the public in general, it is recommended that changes occur in a phased approach as follows:

In Phase I, the current MEAP would be expanded to include a measurement of thinking processes identified in accordance with Bloom's taxonomy and essential skills in the areas of Reading, Mathematics and Writing. Students would be tested in grades 4, 7 and 10.

Phase II would replicate Phase I and, in addition, an indicator test would be administered to students in Grades 3, 6 and 9. The indicator would delineate the skills of problem-finding, problem-solving, critical thinking and decision-making.

Phase III would replicate Phase II (i.e., the indicator test would continue to be administered) but the indicator skills would be measured on the Grades 4, 7 and 10 essential tests.

Phase IV would be the same as the preceding Phase III. This final phase may include a Grade 12 test where high school subject content would be assessed.

As the Committee discussed these two plans along with the original two papers, a list of issues has emerged. These issues are shown below, as well as the initial "votes" of the

Committee. The list of issues will form the basis of future discussions of the Advisory Committee.

	ISSUES	OPINIONS OF COMMITTEE	
1.	Should thinking be tested?	12-Yes	0-No
2.	If yes, should thinking be tested:		
	a. as as separate content area?	0-Yes	7-No
	b. within the subject matter content? i.e., within science, social studies, etc.?	8-Yes	0-No
	c. a combination of a & b?	5-Yes	4-No
3.	Should there be a two tier test? (If no, go to 6)	7-Yes	4-No
4.	If Yes to 3:		
	a. Should tier 1 test only essential subject matter and tier 2 test only thinking skills?	0-Yes	5-No
	b. Should tier 1 test essential subject matter and thiking skills and tier 2 test harder subject matter?	0-Yes	5-No
	c. Should tier 1 test essential subject matter and thinking skills and tier 2 test harder subject matter and thinking skills?	7-Yes	0-No
5.	If Yes to 3:		
	a. Should all students take both tiers?	7-Yes	1-No
	b. Only those "passing" the 1st level take the 2nd level?	0-Yes	5-No
	c. Test all students on level 1 and sample test level 2?	0-Yes	3-No
	d. Sample test level 1 and test all students on level 2?	1-Yes	2-No
6.	Should there be a writing assessment? If yes, how should it fit with the above?	12-Yes	0-No

Responses:

Level 2 Test
Compose a persuasive statement which relates thinking to content
Given to all students in tier 2
Every-pupil testing
It could be done 1 of 2 ways:
1. essays within content area
2. in separate content area

Completing the Plan

The Higher Level Advisory Committee hopes to finalize the plan for a higher level assessment program by October, 1985. Once completed, the plan will be submitted to the State Board of Education for review and action. If the State Board of Education approves the plan, staff will immediately begin to present it to local educators throughout the State, and at the same time will begin to develop the specific list(s) of skills to be measured. It is anticipated that it will take at least two years to finalize the list of skills and appropriate measures of them, and that it will be at least three years before a revised assessment program can be implemented.

REFERENCES

Bloom, Benjamin S. (Editor). Taxonomy of Educational Objectives. Handbook I: Cognitive Domain. New York: David McKay Co., Inc., 1956.

Downing, Gloria, Sharon Johnson-Lewis, Linda Leddick, Seibert Lohr and Betty Stevens. Report of the MEAP Design Committee. Michigan Department of Education, 1985.

Roeber, Edward D. Michigan Educational Assessment Program: Proposed Plans for the Future. Michigan Department of Education, 1984.

Rudman, Herbert C. "Testing Beyond Minimums." Paper presented at the 25th Michigan Testing Conference, The University of Michigan, Ann Arbor, Michigan, 1985.

Sederburg, William A. and Herbert C. Rudman. "Educational Reform and Declining Test Scores," Michigan School Board Journal, 30:8-10,24 (April, 1984).

State Board of Education. Better Education for Michigan Citizens: A Blueprint for Action. Lansing, Michigan, 1984.

ASSESSING HIGHER ORDER THINKING SKILLS IN CONNECTICUT: LESSONS FROM CONNECTICUT

Joan Boykoff Baron
Connecticut State Department of Education

Since 1982, the Connecticut Assessment of Educational Progress (CAEP) program has been systematically integrating higher order thinking skills into its assessment of subject matter domains in grades 4, 8 and 11. To complement these efforts, the new Connecticut Mastery Testing Program has incorporated many inferential and evaluative comprehension skills into its fourth grade reading test, and conceptual understanding and problem-solving skills into its fourth grade mathematics test (see Tirozzi et al. 1985). This paper will first summarize what we have learned about students' thinking skills when measured in the context of social studies and English language arts. Then, it will summarize what we have learned about how to measure higher order thinking skills, discussing some of the current methods being explored and the challenges which lie ahead.

The Performance of Connecticut Students

In general, Connecticut students perform either the same as or slightly better than the national sample tested by the National Assessment of Educational Progress. Furthermore, Connecticut students in the early and mid 1980's are performing about the same as they were five years earlier. It is against this backdrop of rather typical and stable performance that we

are confronting the disappointing results found on higher order thinking skills between 1982 and 1985.

On our 1982-83 Social Studies Connecticut Assessment of Educational Progress (CSDE, 1984) students performed poorly on many items measuring higher order thinking skills. For example, "students had difficulty in recognizing associations such as cause and effect when more thought than immediate recall was required, in drawing conclusions from evidence and in interpreting data." Five of the nine statewide recommendations in social studies pertained to thinking and are presented below:

- Provide students with as many opportunities as possible to interpret information rather than merely recite it back in an identical form.
- Encourage students to interpret information depicted in graphs, charts, and tables rather than simply read it.
- Emphasize the personal relevance and modern day implications of social studies concepts.
- Place greater emphasis on cause and effect relationships.
- Incorporate problem solving and logical analysis in the context of social studies.

Similar findings from our 1983-84 English Language Arts Connecticut Assessment of Educational Progress were reported to the Connecticut State Board of Education:

> One finding that pervaded the assessment in reading, literature, listening, study skills, writing, and computer literacy was that students do well on the literal comprehension level and not so well at the higher levels of thinking. Our students have learned a lot. They have many facts

at their command and they can solve simple one-step problems. However, when they are asked to infer, integrate, and evaluate, performance drops. Furthermore, when they are asked to solve more complex problems involving the application of knowledge to new situations, the condensation of information, the synthesis of several pieces of information or the solving of problems requiring several steps, performance drops. In addition, when students are asked to develop and maintain a point of view and support it with reliable and sufficient evidence, performance is poor.

The rest of this paper is devoted to some lessons we've learned on how to assess thinking skills and to a brief description of some of the challenges that lie ahead.

Using Multiple Approaches to Assess Thinking Skills

The first lesson we learned is about the importance of using multiple approaches to assess thinking skills. Frederickson (1984) alerts us to the bias inherent in relying solely on multiple choice items. In our two most recent CAEP assessments we used multiple approaches. In English language arts, for example, we measured writing skills with five approaches which included more than one hundred multiple-choice items, two direct measures of writing requiring writing samples from narrative and persuasive discourse modes, a dictation test, a note-taking exercise, and a revising and editing test in which students had to correct errors made by others. Furthermore, in an attempt to be eclectic as well as thorough, we used holistic, primary trait, and analytic scoring rubrics to score our writing samples. Our experience clearly demonstrated that these three

scoring rubrics provide information of such varying levels of specificity about both content and mechanics that the choice of scoring methods should be dictated by the purpose of the test and the degree to which the information will be used by teachers to influence the instructional process (See Baron, 1984).

It is unfortunate but true that different measures of the same trait using different methods often provide different results. For example, spelling results using multiple choice items requiring students to select the one misspelled word from among four alternatives, differ dramatically from spelling results generated from a paragraph densely laden with an unspecified number of errors which students have to locate and correct. (See Baron, et al. 1985 and CSDE 1985 for some examples.)

Fortunately, sometimes different approaches yield corroborative results, a particularly reassuring finding when one is preparing to embark on a major effort to remedy a problem. One example of corroborative data was found on our English language arts test when we used three approaches to measuring students' ability to recognize and provide good support in writing. Student performance was consistently disappointing. On a persuasive essay, at all three grade levels tested (4, 8, and 11), fewer than 5 percent of the students were judged to have provided enough support to convince a television critic to either write more editorials like the one he had written or to take back what he had written. Fewer than five

percent of the grade 8 students provided support that was judged adequately deep, sufficiently credible, or amply numerous. The grade 11 students performed slightly better with 17 percent providing sufficient support to validate their position and 31 percent explaining the stated reasons with ample explanation. On the revising test, where students were specifically asked to provide support for purchasing school computers, the grade 8 students outperformed the grade 11 students with just over a quarter of the grade 8 students providing two or more credible facts, examples and/or reasons as support. (The corresponding number of eleventh grade students was 12 percent.) On a multiple choice item requiring students to recognize an essay's greatest weakness, only 40 percent of the grade 8 students identified the correct answer, "It does not provide enough supporting examples." (See Baron & Wallick, 1985 for some examples and CSDE, 1985 for a more detailed description of the findings.)

Our recently completed science assessment also used multiple assessment approaches. We measured the same concepts using multiple choice items, short essay questions, and a practical test which included short tasks like focusing a microscope, wiring an electrical circuit to light a bulb, weighing, measuring and sorting objects, and conducting an experiment. In examining the data, the importance of using multiple approaches was quite evident. Consider the multiple choice item provided in Exhibit 1. What conclusions might be drawn from the data

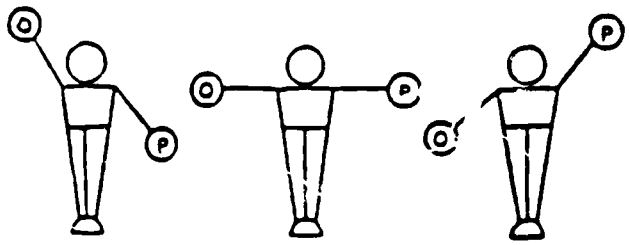
which shows that 71 percent of the fourth grade students answered this item correctly, as compared with just over half of the eighth and eleventh grade students? Our state advisory committee generated lots of hypotheses ranging from skeptical suggestions like "it's an anomaly" to more optimistic ideas like "these fourth grade students must be getting the 'hands-on experience' that many of the science experts in the state have been advocating." Imagine the

Exhibit 1

Percentage of Students Selecting Each Option

Gr. 4	Gr. 8	Gr. 11	
17	31	30	A
71	56	57	B*
8	8	9	C
4	3	3	D
1	1	0	No. resp.

Suppose that you want to drop a penny and a quarter at exactly the same time and have them hit the floor at exactly the same time. Which picture BEST shows how you would hold the penny and the quarter just before you drop them?



D. I don't know.

Example of science choice item measuring higher order thinking skill.

committee's surprise when they turned to the results of the practical exercises in which fourth grade students had been asked to predict what would happen if a penny and a quarter were dropped together and discovered that only 5 percent gave the correct response that they would fall at the same rate. It then became clear that the students did not really understand the physics concept being measured by the multiple choice test item in Exhibit 1.

To Infuse Thinking Skills into Subject Areas or Keep Them Separate?

The second major lesson we've learned concerns the debate over whether to infuse thinking skills into the curriculum and the test or whether to teach and assess thinking skills separately. Perkins (1986) refers to this as the "generality-power tradeoff." If you teach a broad skill, it will have wide generality to many areas but not much specific applicability to any particular subject area. On the other hand, if you teach a narrow skill, it will boost performance in that narrow area but have little applicability to other areas.

Infusing Thinking Skills into Subject Area Curricula and Tests.

If one chooses to infuse thinking skills into the assessment of subject areas, Bloom's taxonomy (1956) can be very useful in designing test items if the taxonomy is used systematically. We

initially used Bloom's taxonomy to help us create a test that would be balanced across social studies disciplines. On prior assessments, students generally performed more poorly on some subsets of skills than others. There was always the temptation to conclude that students knew less about those areas. However, when experts scrutinized the various groups of items, often they could explain the results on the basis of the cognitive skills demanded. In order to avoid drawing inaccurate conclusions, we assigned to each item on the test a Bloom's taxonomic level and equally distributed the levels across the subsets of items that would be reported. In this way, if differences among the item groups emerged, the differences would not be a function of different cognitive skills.

One of the findings worth noting is that contrary to popular belief, knowledge items are often the most difficult items on the test because of their sensitivity to instruction and recall. In order to get a knowledge item right, the student has both to have learned the information and to be able to recall it. Because there is no standardized statewide social studies curriculum or list of approved textbooks used in Connecticut, the likelihood of all students being exposed to any particular piece of information is low. Even if they had been exposed to the material, they would still need to recall it, often after a period of several years. By contrast, some of the higher order thinking skills items were developed with generally well known information that students were required to apply. Had we not

tried to do this, if a student got an item requiring higher order thinking skills wrong, there would be no way to know whether it was because the student did not have the requisite knowledge or whether the student had the knowledge but could not use it. An example of such an item is provided in Exhibit 2. In this item, the only knowledge required concerns the concept that two nations would be more likely to work together when they each had abundant resources that the other needed. What we learned from items like this one is that these concept application items are often not as difficult as knowledge items assessing less commonly known information.

Exhibit 2

Below are the names of some imaginary countries that are neighbors. In which of the following situations would the two neighboring countries be MOST LIKELY to work together?

	Percentage of Students Selecting Each Option Gr. 8
A. Lam has coal but not enough wheat. Alf has coal but not enough cotton.	5
B. Dara has oil but not enough food. Mondo has food but not enough oil.	81*
C. Clow has food but not enough water. Tarm has food but not enough wood.	9
D. Kant has sugar but not enough copper. Nale has potatoes but not enough fish.	5

Example of science multiple choice item measuring higher order thinking skill with familiar concept.

For further clarity, we used a more systematic procedure called "nesting". In nesting, several items were created to cover the same topical areas, but at different levels of conceptualization. In this way, when students perform poorly on questions requiring higher degrees of conceptual thought, it can be determined more accurately whether that weakness was due to a lack of factual knowledge or whether the problem lay elsewhere. It is often the case that students can provide factual information, but they lack the skills necessary to successfully apply the information to problems using those same facts. An example of nesting is found in Exhibit 3. In the first question 59 percent of the eighth grade students indicated that when presented with the definition of supply and demand, they

Exhibit 3

	Percentage of students Selecting Each Option
	Gr. 8
The price of a product is determined by the relationship between people's wants and needs, and the availability of the product. This is called	
A. supply and demand.	59*
B. price fixing.	19
C. black market.	12
D. bartering.	9
If the law of supply and demand works, the farmer will obtain the highest price for crops when	
A. both supply and demand are great.	41
B. both supply and demand are low.	9
C. supply is great and demand is low.	11
D. supply is low and demand is great.	38*

Example of "nesting" using two social studies items.

could label it. However, on the second question, only 38 percent of the grade 8 students could apply that definition to a hypothetical situation requiring the understanding of an inverse correlation.

Several psychologists and philosophers have discussed the importance of integrating critical and creative thinking and reasoning skills into subject areas. (See Glaser, 1984 and McPeck, 1981.) It therefore seems unconscionable to not devote our psychometric energies to continuing to develop compatible assessment strategies. (This applies not only to state tests but to local and national tests as well.)

Teaching and Testing Thinking Skills Separately.

As noted earlier, there are also many experts in the field of thinking skills who advocate testing higher order thinking skills separately. Because he has authored several tests on critical thinking, Robert Ennis is often cited as one such expert. This is only partly true. In July, 1985 at a presentation at the University of Massachusetts Critical and Creative Thinking Program Summer Lecture Series, Ennis made clear his position that critical thinking skills should be assessed in both ways--as infused and isolated. This is the current position of many of the experts who have been identified with "isolating" thinking skills, and it is the position of the Connecticut State Department of Education as well. At the present time, we have a statewide committee overseeing the development of a variety of approaches to assess higher order thinking and reasoning skills in the elementary and secondary grades. This is part of a larger effort to develop appropriate

objectives, suggested instructional activities and learning strategies, and staff development activities from kindergarden through high school. Two aspects of this larger program were pilot tested in the fall of 1984 in grade 4. These included a set of multiple choice items based upon Sternberg's triarchic theory of intelligence and a multiple choice test of critical thinking skills developed by Ennis (1985).

When Sternberg was asked to help develop multiple choice test items based upon his triarchic theory of intelligence, he selected the following 12 objectives as being appropriate for fourth grade students using a multiple choice format (see Sternberg and Baron, 1985):

1. Standard verbal analogies
2. Counterfactual verbal analogies
3. Standard number series
4. Figural classifications
5. Everyday inference
6. Counterfactual everyday inferences
7. Inferences about advertisements
8. Linear syllogisms
9. Spotting contradictions
10. Learning from context
11. Route planning
12. Mathematical and logical insights

The objectives developed by Ennis are presented below. The preliminary results of the test as well as a description of some protocol analyses are described in Ennis, 1985.

DEFINE AND CLARIFY

1. Identify central issues and problems
2. Identify conclusions
3. Identify reasons

4. Identify appropriate questions to ask, given the situation
5. Identify assumptions

JUDGE INFORMATION

6. Determine credibility of sources and observations
7. Determine relevance
8. Recognize inconsistency

INFER: SOLVE PROBLEMS AND DRAW REASONABLE CONCLUSIONS

9. Infer and judge inductive conclusions
10. Deduce and judge deductive validity
11. Predict probable consequences

It might be interesting to note that Ennis' items were pilot tested both as reading items and as listening items where the students also saw the item as it was read aloud. On average, the students performed about 6 percentage points better when they heard and saw the items, although there were some items where there were no differences between the two presentations and others on which students performed better when they read the items without hearing them read. This motivated us to ask Ennis to develop a cartoon version of the test to be used for children in elementary school. The cartoon version of the test is designed to reduce the reading load and be more motivating for elementary school children. We are currently pilot testing the cartoon version on the grade 6 test in Fall 1985.

The Challenge Ahead

It has become increasingly more apparent that there is a larger payoff in teaching learning strategies than in teaching specific knowledge (Perkins, 1986). Furthermore, teachers

should "teach for transfer", looking for applications of the same skills in a variety of contexts. We hope our assessment instrument will develop in parallel ways, with attention paid to measuring the same thinking skills and strategies as applied to different subject areas.

In trying to develop assessment approaches we recognize the need to develop activities and items that have "ecological validity" or a high degree of verisimilitude. In other words, the items should be similar to those that students will have to face in their lives. For example, one of the desirable traits of good thinkers is that they persist in the face of failure. We are therefore looking at ways to incorporate persistence and sustained thought into our assessment. Certainly writing exercises can be ecologically valid and incorporate sustained thought. And certainly, our science tests incorporated these traits into the practical section of the test requiring students to design and conduct an experiment.

As described in Baron (1986) another fruitful area for evaluation is in the assessment of students' dispositions as they relate to students' thinking. Brandt, (1985); Costa, (1984); Duckworth, (1978); Ennis, (1986); Feuerstein, (1980); and Nickerson, (1986) have provided lists of dispositions of good thinkers. Efforts are currently underway in Connecticut to develop an inventory of thinking skills dispositions that can be used by teachers and administrators to monitor students' attitudes and dispositions.

In the past few years, apart from discovering the inadequacies in students' thinking skills, we are beginning to better understand the issues, problems, and needs related to assessing thinking. If assessing thinking skills becomes a national priority, we can look forward to the collective wisdom of psychologists, philosophers and psychometricians assisting educators to develop instruments that will more accurately determine the extent to which our students are becoming better thinkers.

References

Baron, J. "Writing Assessment in Connecticut: A Holistic Eye Toward Identification and an Analytic Eye Toward Instruction." Educational Measurement: Issues and Practice". 3, no. 1 (1984): 27-28, 38.

Baron, J.B. "Evaluating Thinking in the Classroom." In Teaching Thinking Skills: Theory and Practice, edited by J.B. Baron and R.J. Sternberg. New York: W.H. Freeman, Inc. In press.

Baron, J. and Kallick, B. "Assessing Thinking: What Are We Looking For? And How Can We Find It?" In Developing Minds: A Resource Book for Teaching Thinking, edited by A.L. Costa. Association for Supervision and Curriculum Development. In press.

Baron, J., Quellmaz, E. & Court, S. "Alternative Approaches to Measuring Writing." Paper presented at symposium, Research in Large-Scale Writing Assessment. American Educational Research Association Annual Meeting. Chicago, Illinois.

Bloom, B.S., ed. Engelhart, M.D., Furst, E.J., Hill, W.H., and Krathwohl, D.R. Taxonomy of Educational Objectives the Classification of Educational Goals. Handbook I: Cognitive Domain. New York: David McKay Co., Inc., 1956.

Brandt, Ronald. "A Preliminary Synthesis of Teaching Thinking." Mimeograph. Distributed in Fairfield, Connecticut. Winter, 1985.

Connecticut State Department of Education (CSDE). Connecticut Assessment of Educational Progress 1982-83 Social Studies Summary and Interpretations Report. Hartford, Connecticut, 1984.

Connecticut State Department of Education (CSDE). Connecticut Assessment of Educational Progress 1983-84 English Language Arts Summary and Interpretations Report. Hartford, Connecticut, 1985.

Costa, A.L. "Thinking: How Do We know Students Are Getting Better At It?" Mimeograph. Distributed at The Harvard Conference on Thinking, Cambridge, Massachusetts, 1984.

Duckworth, A. The African Primary Science Program: An Evaluation and Extended Thoughts. Grand Forks, North Dakota: University of North Dakota, 1978.

Ennis, R. "Large Scale Assessment of Critical Thinking in the Fourth Grade." Paper presented at symposium, Issues in the Development of a Large-Scale Assessment of Critical Thinking Skills. The American Educational Research Association Annual Meeting. Chicago, Illinois, 1985.

Ennis, R. "What is Critical Thinking?" In Teaching Thinking Skills: Theory and Practice, edited by J.B. & R.J. Sternberg. New York: W.H. Freeman, Inc. In press.

Feverstein, R. Instrumental Enrichment: An Intervention Program for Cognitive Modifiability. Baltimore: University Park Press, 1980.

Frederiksen, N. The Real Test Bias. American Psychologist, 39, no. 1 (1984): 1-10.

Glaser, R. (1984) "Education and Thinking: The Role of Knowledge." In American Psychologist, 39, no. 1 (1984): 93-104.

McPeck, J.E. Critical Thinking and Education. New York: St. Martin's Press, 1981.

Nickerson, R. "Why Teach Thinking?" In Teaching Thinking Skills: Theory and Practice, edited by J.B. Baron & R.J. Sternberg. New York: W.H. Freeman, Inc. In press.

Perkins, D.N. "Thinking Frames: An Integrative Perspective on Teaching Cognitive Skills." In Teaching Thinking Skills: Theory and Practice, edited by J.B. Baron & R.J. Sternberg. New York: W.H. Freeman, Inc. In press.

Sternberg and Baron. "A Triarchic Approach to Measuring Critical Thinking Skills: A Psychological View." Paper presented at symposium, Issues in the Development of a Large-Scale Assessment of Critical Thinking Skills. The American Educational Research Association Annual Meeting. Chicago, Illinois, 1985.

Tirozzi, G.N., Baron, J.B., Forgione, P.D., & Rindone, D.A. "How Testing is Changing Education in Connecticut." Educational Measurement: Issues and Practice, 4, no. 1 (1985): 12-16.

THE ASSESSMENT OF HIGHER-ORDER THINKING SKILLS: RECENT DEVELOPMENTS

John Fremer and Mark Daniel
The Psychological Corporation

This paper identifies thirteen developments related to the assessment of higher order thinking skills (HOTS). In our listing and analysis we attempt to bring together testing, curriculum, and instructional points of view because we are convinced that it is the users of test results who have the greatest potential to help students and improve programs. The best a test can do is to provide information on a sample of student skills. It is the teacher in the classroom, the curriculum supervisor, the school administrator or the program evaluator who must apply this information in an effective way.

It will be useful to comment briefly on terminology. We view "higher order thinking skills" as those skills that go beyond straight recall or learning of facts. They encompass a wide range of activities including problem identification and problem solving, evaluation of information and of arguments, deduction, inference, taking alternate points of view, creating reasonable arguments in support of a position, and making decisions. The term "critical thinking" often is used interchangeably with higher order thinking, but it also has a specialized meaning that denotes a formal approach to problem analysis and argument evaluation.

Dispositions, or motivational factors, also are central to higher order thinking skills, because without the desire to make

good decisions and the willingness to consider new ideas, the reasoning abilities listed above are unlikely to be called into play. A primary goal of higher order thinking instruction is to create a habit of reflective thinking.

INSTRUCTION AND TESTING

A Major Trend

Major attention is being devoted to the development of curriculum materials and tests directed at higher order thinking skills at all educational levels. According to Edward Glaser, a founder of the critical-thinking movement and author of the Watson-Glaser Critical Thinking Appraisal, the current interest in critical thinking is stronger and more widespread than at any time in the last 35 years.

One way of tracking an educational movement is to look at press coverage. US News & World Report, January 14, 1985, had an article, "Think. Now Schools Are Teaching How." The article reported on an American Federation of Teachers survey indicating that six states out of 23 responding had passed laws mandating instruction in critical thinking. It described available programs in the schools. On February 6, 1985, the Hartford Courant ran a story that was headed "State Says Johnny Can Read - He's Ready to Reason." The Sunday, May 19, 1985, Cleveland Plain Dealer ran an article that had originally been written for Harper's magazine, "Why Johnny Can't Think."

At the Michigan Testing conference in February 1985, state testing staff from Florida and Connecticut described higher order thinking skills projects in their states. The emphasis in Florida is on testing higher levels of the Bloom taxonomy within content areas. The state is working on developing realistic standards for average-ability and high-ability students. The State of Connecticut asked the Psychological Corporation to build a higher order thinking skills test for grades 4, 6, and 8 to be part of the statewide mastery testing program. Connecticut has engaged in very thoughtful and careful planning, bringing together ideas from many sources including Robert J. Sternberg of Yale and Robert Ennis of the University of Illinois.

Textbook and test companies are also very active. The Metropolitan Achievement Test Sixth Edition (MAT-6) that is coming out in August 1985 has a Higher Order Thinking Skills score. In addition, conferences on higher order thinking skills are springing up all over the country. The Connecticut State Department of Education recently ran one, and it was oversubscribed by 100%! The International Conference on Critical Thinking and Education Reform attracts a growing number of participants from around the country, as does the Conference on Thinking (held at Harvard in 1984).

Not Just a Reaction

Part of what is happening can be interpreted as a reaction to the back-to-the-basics movement, but other factors are clearly at

work. Some of the focus on higher order skills is a direct reaction to the amount of attention devoted to basic skills, survival skills, and minimum competency testing. The concern of the Basic Skills/Minimum Competency movement was to bring as many students as possible up to a specified minimum level of achievement. Exactly which group was being focused on varied somewhat from place to place, but it usually was something like the bottom 25% of developed ability. However, all along people have been asking: "What about the average, above average, and gifted student?"

On the other hand, one of the reasons we feel that the higher order skills movement is more than just the opposite of the basic skills movement is that we see evidence that curriculum developers and educators want to improve the reasoning skills of ALL students, not just the more able ones. In general, instructional programs are not targeted toward special groups, such as gifted and talented students. Instead, attention is being paid to cultivating student reasoning skills over the entire range of student ability. Programs that do focus on special groups may be aimed at low-ability rather than high-ability groups, such as Instrumental Enrichment, which is intended to give low-achieving students the learning skills that will help them perform better.

Recent scientific work in the field psychology has provided a foundation for the current HOTS movement. There has been a great deal of attention paid to the processes involved in

problem-solving and learning. Better thinkers may be seen to differ from less effective thinkers largely in how they approach problems, rather than in their "mental hardware." This research base has naturally encouraged efforts to add training in cognitive processes to the school curriculum.

Test data are another impetus to the HOTS movement. Test data have seen an extraordinary amount of use in virtually all recent analyses of education. Some reasonable interpretations of National Assessment of Educational Progress (NAEP) and state assessment data are being combined with misinterpretations of the Scholastic Aptitude Test score decline data. NAEP's Reading, Thinking, & Writing report (1979-80) points out that many students seem to lack the skills to evaluate the ideas that they take away from something they read. We believe that the NAEP data do indicate a need for better training of thinking abilities. However, people have also used the score declines in the College Board's Scholastic Aptitude Test and in the ACT to draw unwarranted inferences about thinking skills.

Modifiable Skills

An important characteristics of the HOTS assessment movement is its emphasis on modifiable skills. This is in contrast with conceptions of human ability as somehow being fixed (a variation on the "nature versus nurture" controversy). We see people wanting to improve the reasoning skills of students, as opposed to merely using tests to classify students as being at different

levels. The HOTS movement appears to be based on the assumption that virtually all children can be taught certain problem-solving techniques, strategies, principles, dispositions, and habits of thought that will improve their ability to deal with problems they encounter as students and as members of society.

This emphasis on teaching rather than classifying is a very positive development. It relates to some other current trends:

- preparing people to do better on tests (the computer SAT by Harcourt Brace Jovanovich was a smash hit); and
- the interest in diagnosis in testing (the Stanford Diagnostic Reading and Mathematics Tests are very popular).

Defining Content

Tests can help to define the content of a movement. We did an extensive review of existing tests in the course of developing the Metropolitan Achievement Test Sixth Edition Higher Order Thinking Skills score. We found that some available tests seemed to go beyond what was appropriate for achievement testing; that is, they included figural analogies or syllogistic reasoning materials that are not generally part of the elementary or high school curriculum. Other materials were exclusively taxonomies that had not attempted to integrate the taxonomic terms with the various subject matter disciplines. Still others seemed too inclusive, labeling as higher order thinking almost anything that went beyond the initial knowledge stage in any taxonomy. Existing tests of thinking skills are too varied to serve as a guide to the content of the higher order thinking skills movement.

Perhaps the main challenge involved in developing tests of higher order thinking skills will be to select objectives from the broad domain. It is technically possible to test for a large variety of reasoning and problem-solving skills. However, not all of these skills can be covered in any single test or perhaps even in any test program. Some of these skills are more important than others, in the sense that they have wider applications in school and work. For example, to borrow from Robert Sternberg's terminology, executive-processing skills such as planning and strategy selection may be more important than individual performance-component skills. Further, objectives differ in how well they can be addressed in existing school courses. The selection of test content will be closely linked to experiments in teaching higher order thinking skills.

Everyday Applications

A good deal of work is being devoted to how thinking skills can be cultivated so that students can analyze television news advertising, political speeches, and other everyday presentations of positions and arguments. Part of Edward Glaser's motivation in developing the Watson-Glaser Critical Thinking Appraisal in the early 1940's was to help students of all levels of ability to think more critically about important issues. This test, which has been revised in 1966 and again in 1980, measures Inference, Recognition of Assumptions, Deduction, Interpretation, and Evaluation of Arguments. The test is sensitive to instruction in

critical thinking, and use of the instrument is expanding rapidly.

Basic Subjects

The curriculum areas where greatest attention appears to be going to higher order thinking skills are reading, writing, mathematics, social studies and science. Higher level thinking is being addressed in the basic subjects, not primarily in highly specialized and advanced subjects. The curriculum is being expanded by being given depth, not by adding new subjects.

Wide Age Range

One area of HOTS instruction and assessment that requires substantial exploration and research is the proper ages and developmental levels for teaching various thinking skills. As new methods of teaching are tried in elementary schools, we will learn more about the capacity of children of various ages to handle such things as designing experiments, analyzing the structure of an argument, and identifying relevant and irrelevant information. Lipman's Philosophy for Children program has shown that young children can not only learn some basic philosophical principles but also take an active interest in discussing them. It would be a mistake to assume that children of certain ages are unable to acquire particular reasoning skills without having made an effort to teach those skills in an appropriate fashion. The door is open to experimentation on this issue.

The extent to which thinking skills training will be accessible to students at all ability levels depends in large part on how content is defined, and on whether thinking skills instruction is embedded in subject areas or treated as a separate subject.

Multiple Assessment Techniques

Both objective multiple-choice tests and more open-ended tests are playing a role in the assessment of higher-order thinking skills. Multiple-choice tests are uniquely suited to certain assessment needs, such as monitoring the performance of large numbers of children, or measuring change over time. Many of the thinking skills are well suited to measurement by multiple-choice item types. However, it is also true that some of the more complex thinking-skill objectives can best, or only, be assessed by other means. When teaching a child to analyze an argument, there is no better way of evaluating learning than asking the child to analyze an argument, orally or in writing. Similarly, if one wants to know whether students have developed the habit of selecting a problem-solving strategy before trying to solve the problem, the best approach probably is to observe the process. Thus, although objectives tests can provide useful information on thinking skills, there will also be a need for a considerable amount of classroom-level assessment.

Validation

One of the most important issues a test developer in this field faces is how to validate a test of higher order thinking skills. There is no easy answer to the problem of validating new thinking-skills tests, because we lack easily-available criteria for "good thinking." We suspect that the task calls for a "bootstrap" approach. On the one hand, tests of thinking skills programs are effective. On the other, the increase (or lack of increase) in test scores following instruction in thinking skills indicates whether the test is a measure of the thinking abilities being taught. The closer the test content is to the target behaviors of the training, the more confidence we can have in the test's construct validity. With experience we will discover which types of tests are sensitive to certain types of training.

Separate or Integrated?

There are clear differences of opinion among those who favor separate instructional units on reasoning skills and those who insist that such skills need to be addressed within existing curriculm areas. Schools have a number of "free-standing" instructional packages to choose from, such as Philosophy for Children or Instrumental Enrichment. A 1984 Educational Testing Service (ETS) report Focus 15: Critical Thinking describes a number of programs in schools and colleges around the country. One of the people quoted in this article is Vincent Ruggiero, textbook author.

Ruggiero argues, "You have to have a special course for students to learn the full range of critical thinking, and other courses should reinforce what is learned." He compares the critical thinking course to freshman English as a course teaching a fundamental skill necessary to succeed in all college courses. He also insists that, like writing, critical thinking should be a part of every other course, "No one argues that because freshman English is taught in college, no one else has to each writing."

Ruggiero's course would cover problem solving and decision making, principles and techniques of creative thinking, overcoming attitudes that handicap thinking, and developing techniques for critiquing one's own arguments. The course would also introduce students to the techniques and principles of persuasive writing and provide them with practice in the detailed expression of their ideas.

On the same page, the argument is presented that critical thinking should be integrated into every subject in the curriculum and that establishing a separate course is unnecessary and, in many cases, impractical. The Critical and Creative Thinking Program at the University of Massachusetts at Boston, for instance, prepares teachers to incorporate critical thinking into established courses. "I don t think you need to introduce a new course to teach critical thinking," says Robert Swartz. "Perhaps the best approach is to introduce critical thinking into the existing curriculum - to make it part of existing courses. Certainly, to introduce critical thinking as a separate course

without making it part of the rest of the curriculum sends a mixed message to students."

In support of this position the ETS report quotes Barry K. Beyer, in an April 1984 Phil Delta Kappan article, as saying, "Research suggests that skills taught in isolation from subject matter are not likely to transfer easily to other situations where they can be used productively. Research also suggests that skills taught in isolation from one another are not likely to become functional. Furthermore, research suggests that massed practice of skills is not as effective in promoting learning as intermittent practice and reinforcement over a long period of time. Thus the research that has been conducted seems to argue for sequential instruction in thinking skills across all subject areas and throughout all grades, K-12. Few such curricula exist, but they should be developed."

The ETS report goes on to describe an "integrative approach" that is being pioneered in the junior high schools of Pennsylvania's Neshaminy School District. Each of the district s three junior high schools employs a specialist who comes into regular classrooms to present units in critical thinking and philosophy that are coordinated with the subject matter of the standard curriculum.

Textbook publishers are working hard to emphasize the units on higher order skills in their existing materials and to make new materials such as Thinking Boxes and Packages.

Choosing a Program

If you have to choose a program, the approach recommended by Dr. Robert Sternberg may be helpful. He argues that the research that he and a colleague, Janet Davidson, have carried out supports the effectiveness of well-executed theory-based training in higher order intellectual skills. He presents thirteen general principles for selecting and offering training programs.

1. Clarify your purposes and needs for training.
2. Choose programs with some real-world content, not all abstract materials.
3. Choose programs that are motivating to teachers.
4. Teach for transfer.
5. Have an instructional theory.
6. Address broad-ranging intellectual skills, not narrow test-item content.
7. Teach children how to learn, so they can keep on growing.
8. Use multiple teaching approaches.
9. Provide an integrated program.
10. Use socioculturally appropriate materials.
11. Be responsive to individual differences.
12. Find children's strengths and capitalize on them. Help children recognize and deal with their weaknesses.
13. Eliminate barriers to using intellectual skills.

Testing Teachers

The issues of upgrading teacher as well as student thinking skills is receiving attention in instruction and assessment. One of the places we sought help in preparing this paper was ETS.

ETS is now receiving requests to help upgrade the thinking skills of teachers. Some propose that there be course work for teachers in reasoning skills, followed by certification testing in reasoning.

Breadth of Movement

The number of different currents of thought and research that are being brought together under the higher order thinking skills banner is quite remarkable:

- philosophers - (formal and informal logic, Philosophy for Children, dialectical thinking)
- state assessment staff
- curriculum designers
- cognitive psychologists
- test developers
- veterans in the area and newcomers
- people working at all levels of education

Cognitive psychology, in particular, has had some influence, but we think it has potential for a great deal more. Often it takes a long time for the findings of cognitive psychological research to be applied to educational practice. We have already alluded to several questions concerning higher order thinking instruction and assessment that need to be addressed by research. If we will work harder in testing and instruction to involve the research community in our development activities, we see substantial payoffs being possible.

RECENT TITLES
IN THE ERIC/TME REPORT SERIES

#90 - Assessing Higher Order Thinking Skills, by C. Phillip Kearney, and Others. 4/86, $7.50.

#89 - Microcomputer Programs for Educational Statistics: A Review of Popular Programs, by Paul M. Stemmer and Carl F. Berger. 3/86, $7.50.

#88 - Computer-Assisted Test Construction. A State of the Art, by Tse-chil Hsu and Shula F. Sadock. 12/85, $7.50.

#87 - The Statewide Assessment of Writing, by Peter Afflerbach. 8/85, $7.50.

#86 - The Effects of Testing on Teaching and Curriculum in a Large Urban School District, by Floraline Stevens. 12/84, $6.00.

#85 - Reporting Test Scores to Different Audiences, by Joy A. Frechtling and N. James Myerberg. 12/83, $7.00.

#84 - Assessment of Learning Disabilities, by Lorrie A. Shepard. 12/82, $6.50.

#83 - Statistical Methodology in Meta-Analysis, by Larry V. Hedges. 12/82, $7.00.

#82 - Microcomputers in Educational Research, by Craig W. Johnson. 12/82, $8.50.

#81 - A Bibliography to Accompany the Joint Committee's Standards on Educational Evaluation, compiled by Barbara M. Wildemuth. 12/81, $8.50.

#80 - The Evaluation of College Remedial Programs, by Jeffrey K. Smith and others. 12/81, $6.50.

#79 - An Introduction to Rasch's Measurement Model, by Jan-Eric Gustafsson. 12/81, $5.50.

#78 - How Attitudes Are Measured: A Review of Investigations of Professional, Peer, and Parent Attitudes toward the Handicapped, by Marcia D. Horne. 12/80, $5.50.

#77 - The Reviewing Processes in Social Science Publications: A Review of Research, by Susan E. Hensley and Carnot E. Nelson. 12/80, $4.00.

#76 - Intelligence Testing, Education, and Chicanos: An Essay in Social Inequality, by Adalberto Aguirre Jr. 12/80, $5.50.

#74 - Intelligence, Intelligence Testing and School Practices, by Richard DeLisi. 12/80, $4.50.

#72 - Methods of Identifying Gifted Minority Students, by Ernest M. Bernal. 12/80, $4.50.