DOCUMENT RESUME

ED 373 010 SO 024 326

AUTHOR Kiernan, Henry, Ed.; Pyne, John, Ed.
TITLE Measuring Student Performance: Assessment in the Social Studies. Theme Issue.
INSTITUTION New Jersey Council for the Social Studies.
PUB DATE 93
NOTE 28p.
AVAILABLE FROM The Docket, New Jersey Council for the Social Studies, P.O. Bos 4155, River Edge, NJ 07661.
PUB TYPE Information Analyses (070) -- Collected Works - Serials (022)
JOURNAL CIT The Docket: Journal of the New Jersey Council for the Social Studies; Win 1993

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Concept Formation; Decision Making; *Educational Assessment; *Educational Practices; Educational Principles; Elementary Secondary Education; Evaluation Methods; *Outcomes of Education; *Performance; *Social Studies; Standardized Tests
IDENTIFIERS *Authentic Assessment

ABSTRACT

The four articles in this theme issue provide an overview of assessment in the social studies and the rationale behind the movement for a more authentic assessment of learning outcomes. In the first article, "Thinking as an Unnatural Act," William T. Daly offers a clear rationale for social studies teachers to re-examine the methods of assessing student performance. In the second article, "Authentic Assessment in Social Studies," Jack L. Nelson conveys why standardized testing does not measure effectively skills taught in social studies and presents a strong case for social studies teachers to demand more authentic assessment approaches aimed at developing students' critical thinking, ethical decision-making, and the refinement of conceptual ideas. Pat Nickell in the third article, "Performance Assessment in Principle and Practice," offers guidance for developing performance tasks for classroom use and gives several examples. In the final article, "Performance Assessment in Social Studies: What CRESST Research Tells Us," Pam Aschbacher and David Niemi review present initiatives to emphasize deepening understanding, higher order thinking skills, and more authentic assessments of student outcomes in the social studies. They explore the research-based premises about developing performance assessments and provide examples of how to employ assessment tasks to improve instruction. (CK)

THE DOCKET

Journal of the New Jersey Council for the Social Studies

Winter 1993

THEME ISSUE:
MEASURING STUDENT PERFORMANCE:
ASSESSMENT IN THE SOCIAL STUDIES

CRITERIA FOR SUBMISSIONS TO THE DOCKET

EDITORIAL POLICY

The Docket is the official journal of the New Jersey Council for the Social Studies and reaches more than one thousand NJCSS members who work within social studies education from nursery school through graduate education. In an effort to act as a voice through which its members can share and express their ideas, thoughts, experiences, and research, The Docket publishes four types of articles.

1. Practical articles with ideas for teaching such as units, lesson plans, and reports of innovative practices.

2. Scholarly articles which cite current theory and research as a basis for making recommendations for practice.

3. Reviews of educational materials: books, textbooks, computer software, CD-ROM titles, laser discs, and others.

4. Informational articles about the efforts and activities of NJCSS and its members.

PREPARATION OF MANUSCRIPTS

Manuscripts should be between 750 and 1,500 words in length (3 to 7 manuscript pages), though the editors may consider longer manuscripts in exceptional cases.

Manuscripts should be typewritten and double-spaced, on single-sided, 8.5 X 11 inch white bond paper with margins of at least one inch all around. Authors should submit one original and two clear photocopies of their manuscript. A computer disk in either DOS or Macintosh format is welcomed. Illegible copies will be returned to the author. A **letter of transmittal** with your name and address should be included with your manuscript specifying that the article has not been submitted or published elsewhere. Do not include your name on the manuscript itself so that your identity can be concealed from the manuscript reviewers and thus insure impartial review. Please write in a concise, readable style free of jargon, stereotyping, and use of sexist language. Quoted material and references should be double-spaced and fully documented, using guidelines explained in *A Manual of Style* (13th ed., University of Chicago, 1982).

Receipt of your manuscript is acknowledged with a letter. Each manuscript is subject to impartial review by The Docket's editorial committee. The committee reserves the right to edit manuscripts for style, while changes in content are made with the consent of the author. The editors retain final responsibility for determining a manuscript's suitability for publication and their decision is final. **Materials shall be returned to authors only if a return envelope with adequate postage is included.**

TITLE PAGE

Name(s), titles and professional affiliations(s) should appear only on the title page so manuscript evaluators will not know the identity of the author(s). The title page must include the title of the paper, the name(s), complete mailing address(es), and telephone number(s) of the author(s).

QUOTATIONS

Quotations that are three lines or more in length should be indented four spaces and double spaced.

PERMISSIONS

The author(s) should obtain permission in writing from publishers for any copyrighted materials to be used in the manuscript, including any text, pictures, illustrations, and cartoons.

COMPLIMENTARY COPIES

Six complimentary copies of The Docket will be sent to authors when their articles or reviews appear.

MANUSCRIPTS AND CORRESPONDENCE

All manuscripts and correspondence should be sent to either of the co-editors:

Dr. Henry Kiernan
Humanities Supervisor
Southern Regional High School
600 North Main Street
Manahawkin, NJ 08050

or

Dr. John Pyne
Social Studies Supervisor
West Milford Township Public Schools
46 Highlander Drive
West Milford, NJ 07480

An Introduction and an Appreciation

Measuring Student Performance: Assessment in the Social Studies, is the inaugural issue for Henry Kiernan and John Pyne, who succeed William Fernekes and David Pierfy as the new co-editors of **The Docket**. Henry serves as the humanities supervisor at the Southern Regional High School District in Manahawkin, while John supervises the K-12 social studies program for the West Milford Township School District.

We would like to extend our appreciation to Bill Fernekes and Dave Pierfy for their devoted service to social studies educators and professionals around the state over the years, and particularly for making **The Docket** into the quality publication it has become. In addition, both Bill and Dave have eased our transition as editors by generously offering assistance and advice as we assumed operation of the journal and began preparations for our first issue. We extend to them our most heartfelt thanks for their help and our very best wishes for the future.

Bill was awarded the 1992 SIRS Academic Freedom Award at the 23rd Northeast Regional Social Studies Conference in Hartford. In the ceremony marking the occasion on March 13, 1992, Ed Reynolds, a past president of the New Jersey Council for the Social Studies, spoke in testimony to Bill's long devotion to the cause of human rights and his well-known efforts on behalf of social studies education, stating:

"{Bill} is the author of countless articles on individual rights, the Bill of Rights, human rights, the Holocaust, genocide, nuclear issues, global education as well as social studies education, scope and sequence, curricula trends, oral history . . . . etc.

He has served as a director of the New Jersey Council for the Social Studies and as editor of **The Docket**, the journal of that professional organization. He has been — in a very real sense — our conscience and a person who keeps us in touch with what we in social studies should and MUST be all about."

Finally, the editors welcome letters from our readers as well as suggestions for future issues. We are especially interested in articles dealing with Asian Studies, with new directions for social studies education at the elementary level, with the reaction of our readers to the recently mandated core proficiencies in U.S. and World History, and with innovative approaches to teaching and evaluating social studies instruction in the classroom.

In This Issue: Measuring Student Performance in the Social Studies

Henry Kiernan and John Pyne, editors

Concerns about our students' academic achievement vis-a-vis students in other industrialized countries and our country's competitive position in the world economy have led to an increasingly vocal and widespread movement to develop "world-class" standards in all academic areas. Currently national standards have been developed (i.e. mathematics), or are being developed in all academic disciplines, including civics - government, economics, geography, history, and social studies. In addition, educators have argued the necessity for amending the "two-tier" tracking system, which places too many of our students, especially minorities, in lower tracks where they pursue a no-win curricula that leaves them educationally disadvantaged and thus unable to compete successfully in the marketplace. Talk of national standards and heterogeneous grouping invariably raise the specter of some form of national standardized test in the minds of many educators and parents and arouses deep-seated suspicions concerning both the viability and desirability of another "paper and pencil" evaluation instrument.

Though evaluation has been a traditional part of learning and will continue to be in the future, it remains a controversial and often traumatic experience for students and teachers alike. While few would deny the importance and necessity for evaluation, the traditional reliance on paper and pencil tests raises questions of objectivity, understanding, equity, and usefulness.

Recognizing the necessity for assessment and evaluation, educators throughout the country have been busily at work developing new assessment instruments and strategies designed to measure learning objectives and what students actually know more appropriately than in the past. Though teachers are often apprehensive about the inadequacies of standardized testing, they nonetheless rely overwhelmingly on short answer tests and quizzes to evaluate their own students' performance.

Yet, many educators criticize such assessment procedures and methodologies for an overemphasis on discrete facts, the trivialization of important knowledge and skills, and the emphasis on one right answer. Students often learn "facts" without the corresponding conceptual framework and experience difficulty applying what they have learned to real life situations. As they race through the curriculum, they seldom have adequate time to process and internalize the information they are expected to know.

Performance assessment or authentic assessment has become the focus of new assessment strategies and methodologies, designed to promote outcome-based education and more appropriately parallel real situations students will experience in their daily lives. Our first issue of **The Docket** provides an overview of assessment in the social studies and the rationale behind the movement for a more authentic assessment of learning outcomes.

William Daly, a professor of political science at Stockton State College, offers a clear rationale for social studies teachers to re-examine the methods of assessing student performance. Bill is the former chairperson of the New Jersey Department of Higher Education's Task Force on Thinking Skills. He is the recipient of the Academy for Educational Development's national prize for educational innovation for his work in adapting thinking skills instruction to the needs of disadvantaged children and

served as a Visiting Fellow at Princeton University, where he studied the implications of recent cognitive science research for classroom teaching.

Jack Nelson, professor in the Graduate School of Education at Rutgers University, conveys why standardized testing, or at least the form that characterizes much of the testing we see today, does not effectively measure skills taught in social studies. Jack presents a strong case for social studies teachers to demand more authentic assessment approaches aimed at developing students' critical thinking, ethical decision-making, and the refinement of conceptual ideas.

Pat Nickell is the Director of Instructional Support Services for the Fayette County Public Schools in Lexington, Kentucky. She recently edited the section on *Student Assessment in the Social Studies* for Social Education (February 1992), the journal of the National Council for the Social Studies. Having served on the NCSS Board of Directors and Task Forces for Middle School, Teaching and Learning, and Curriculum Standards, Pat is currently the editor of the NCSS' *Classroom Teacher's Notebook*. In her article, she provides guidance for developing performance tasks for classroom use and offers several examples of ways in which this may be done.

Pam Aschbacher is the Project Director and David Niemi is a Senior Research Associate at the National Center for Research on Evaluation, Standards, and Student Testing (CRESST) at the Graduate School of Education, University of California, Los Angeles. In their article, they review present initiatives to emphasize deeper understanding, higher order thinking skills, and more authentic assessments of student outcomes in the social studies. They review the research-based premises about developing performance assessments and provide examples of how history teachers and others have employed their assessment tasks to improve instruction.

Along with the authors whose articles are featured in this issue, we believe real assessment and evaluation should reflect more accurately the outcome of classroom instruction and provide evidence of the success or failure of that instruction. Not only will our students benefit from a more honest and humane assessment and evaluative procedure, but teachers will begin to reclaim the control and choice of assessment in our schools. Real assessment and evaluation means knowing our students and knowing how best to support their social studies learning by providing excellence, equity and success for all students.

Thinking As An Unnatural Act

William T. Daly
Professor, Political Science
Stockton State College, Pomona, NJ 08240

The human foot was not built for ballet. Only with discipline, training, and pain can it endure the strain and produce beauty. The human mind was not built for independent thinking. Only with discipline, training, and pain can it endure the strain and produce knowledge. Such at least is the conclusion which seems to be emerging from our unfolding knowledge of how the human mind actually works.

An understanding of the implication of this unsettling conclusion for educators requires, first, an understanding of the kind of thinking skills now being demanded not only by educators but also by business and political leaders. It requires, secondly, an understanding of the growing evidence that much of this kind of thinking runs against the grain of that marvelous piece of mental equipment which our students, as members of the human species, bring from the primeval plain into our classrooms.

The Demand for Independent Thinking

The kind of thinking which is increasingly demanded of our students, both inside and outside of the academy, is independent thinking — thinking which will permit them to go beyond remembering the ideas of others to generate new ideas of their own.

This is not, of course, a new goal for those inside the academy. Since the time of Socrates, it has been the cherished hope of most teachers that they might develop at least some students who could one day add something to the store of human knowledge themselves. And it has always been the collective responsibility of teachers in a democracy to help develop a thinking citizenry, capable of independently evaluating the pronouncements and performance of public officials.

Most of the push behind the current emphasis on thinking skills, however, is not coming from these traditional academic concerns. It is coming from members of the national business and political communities who are concerned about the international competitiveness of the American economy and hence about the education of the national workforce. Their public statements on education reflect their widely shared belief that Americans, in the future, will not make their collective living primarily as mass producers of standardized industrial products. Instead they will have to make it as a source of continuing innovation in technology and services. In such an economy, they suggest, we will have to pursue the initial profits and jobs associated with each innovation, watch most of the economic benefits from its long term production gradually migrate overseas where labor and materials are cheaper, and hence confront the need for an endless series of such innovations.

The educational requirements of this kind of economy will be fundamentally different from those of the assembly-line industrial economy which has sustained American prosperity in the past. The success of this new kind of economy, they argue, will require the education of a larger professional level workforce and one with a substantial capacity for independent and innovative thinking. [1]

The Components of Independent Thinking

The practical meaning, for the classroom teacher, of this academic–economic convergence of opinion on the importance of independent thinking can best be understood by reviewing the kinds of instructional programs which have sprung up in response to it. In spite of variations in phraseology, most of those programs use a basic input–process–output model of thinking. That is to say, that they focus on the way in which students take in information when they read and listen, what they do with it between their ears, and how they put it back out again

in response to the demands of their teachers. Different programs focus on different parts of that three–part process but, viewed collectively, these instructional programs reflect considerable underlying consensus on the kinds of intake, process, and output skills which students must learn if they are to become independent thinkers. [2]

1. Abstract Thinking

This refers to the intake part of the process, and focuses on what students need to do when they read and listen in order to build the basis for independent thinking. Abstract thinking has been highlighted as a component of independent thinking primarily by "Piagetian" instructional programs, i.e., those based on Jean Piaget's famous distinction between "concrete" and "formal" thinking. What students most need, according to these theorists, is to move up a level of generality or abstraction from their instinctive tendency to memorize concrete bits and pieces of factual material in precisely the form in which they are initially presented.

Instead, students need to learn to abstract general concepts or principles from the welter of concrete detail and then to use those intellectual categories both to decide which specifics are worth keeping and recording, and to summarize and organize what is kept. In this way, the construction and use of abstract concepts can reduce the formless tidal wave of new information which school seems to offer to intellectually manipulable chunks of raw material, relevant to the thinking about the question at hand.

Beyond simply helping students to manage information, this capacity to build and use general concepts and principles is also a direct prerequisite to the first limited form of independent thinking — the capacity of students to independently apply what they have learned in one context to related materials which they encounter later. Only if they can abstract general ideas and principles from the concrete materials learned in one context, will they be able to carry those general principles forward and apply them to an understanding of related materials which they encounter subsequently — in a later portion of the same class, in later classes, or in the world of work after they graduate. [3]

For both these reasons abstract thinking is viewed as a crucial prerequisite to the next, more ambitious task — going beyond the management and application of others' ideas to create ideas of one's own. This is, of course, the most mysterious and prized component of independent thinking, and the second step of the input-process-output model used by most instructional programs in thinking skills.

2. Creative Thinking.

This refers to the "process" component of the three-step model, and focuses on what students need to do once they have extracted the information essential to their purpose and organized it under general concepts or principles.

This central component of independent thinking has been highlighted primarily by an explosion of self-help books and instructional programs on "creativity" and "problem solving." What students most need to do, according to these theorists, is to overcome their instinctive tendency toward immediate closure around the simplest or most familiar approach to a question. They need, instead, to wait, to consider a variety of approaches, to arrange the chunks of relevant information developed in the first stage in a variety of configurations — to give themselves, in short, the opportunity to see a new pattern, divine a new approach, generate a new idea.

No one pretends to know where creative insight comes from. But all of the instructional programs which pursue it seem to share the assumption that the appetite for immediate closure is its greatest enemy. And most of the instructional techniques they have devised are best understood as attempts to hold the mind open, and march students through the consideration of a number of alternatives, before permitting closure. [4]

3. Systematic Thinking.

This refers to the output stage of the thinking process, and focuses on what students need to do in order to elaborate on and validate any ideas generated by the first two stages.

Systematic thinking is the central concern of the instructional programs which focus on formal or informal "logic." According to these theorists, students need to be able to determine what follows logically from their ideas and from the available evidence — whether they are writing an essay for an English class or exploring a scientific hypothesis.

This third component of the capacity for independent thought implies the ability and the willingness to subject all ideas, even the most fervently held ones, to the tests of logical coherence and, where appropriate, empirical evidence. It is important to the more general capacity for independent thought for two reasons. First, it permits students to extend their knowledge into new areas by determining what follows logically from things they already know. Secondly, it permits them to validate their developing knowledge by constantly checking it for logical consistency and factual support. [5]

4. Precise Communication of Thought.

This final component of independent thinking refers to the ability to communicate the products of one's thinking to others, and is the central concern of the instructional programs which focus on the relationships between

language and thought. According to these theorists, students need to be able to communicate their thoughts not only orally but in writing. And they need to write with sufficient precision to be intelligible and persuasive not only to friends and teachers but also to audiences which are more distant, diverse, and skeptical.

Writing is emphasized in many thinking-oriented instructional programs both because of the belief that the writing process itself clarifies thought and because it is essential to the process by which knowledge is shared and becomes cumulative. [6]

Teaching Independent Thinking

To teach independent thinking effectively, we may have to seriously consider some approaches to instruction which are both more directive with respect to students and more demanding of teacher time and effort than we may like. A review of the abstract, creative, systematic, and precise thinking, delineated above as the essential underpinnings of the capacity for independent thought, produces the following list of pedagogical implications.

We may have to require constant practice in the construction and use of abstract concepts. And this should probably include not only more practice in the selection, organization, and manipulation of data generally, but also more work in the discipline most centrally concerned with the manipulation of abstract concepts — mathematics. State and national test scores have consistently confirmed what students have been saying for generations — that math, for most of them, is "harder" than other subjects. If the human brain evolved as a device which derives abstractions or generalizations primarily from a series of direct personal experiences, it is not surprising that most of us have considerable difficulty in constructing and manipulating abstract concepts which are more distant from such concrete experiences or which, in the case of mathematics, may have no concrete referents whatever. To help our students rise above their natural tendency to think only in concrete terms, we may have to accept the resource costs, and the teacher and student travail, which will be involved in an attempt to make all our students at least competent in basic quantitative reasoning.

We may have to construct our courses in such a way that students who wish to pass them have no choice but to create at least some ideas of their own — rather than simply demonstrating an understanding of the ideas we give them. If the human brain evolved primarily as a mechanism for providing instantaneous responses to a dangerous environment, it is not surprising that our students instinctively favor immediate closure around the most readily available approach or answer any given question. They seize upon any idea offered by the teacher or text rather than undergo the protracted uncertainty and the painstaking consideration of alternative approaches which would give them their best chance for creating a novel idea of their own.

To overcome this natural appetite for immediate closure, we may have to turn to those admittedly difficult and time-consuming pedagogies which are designed to initially deny students any authorized "right" answer and hence force them to fashion ideas of their own. Examples of this kind of approach include Socratic questioning, in which the teacher offers questions rather than answers, and "discovery" learning, in which the teacher provides a series of concrete experiences or experiments from which the students must derive general principles for themselves. Students who require more structure than such open-ended pedagogies provide might benefit from a "multiple perspectives" approach, in which the teacher presents a range of conflicting approaches or interpretations on each topic and requires the students to evaluate the relative merits of those approaches on route to constructing and defending views of their own.

We may have to insist that our students practice systematic thinking, by consistently requiring them to explain and justify their work logically and in terms of the available evidence. We have already discussed the argument that the human brain did not evolve as the kind of "logic machine" represented most clearly by digital computers, but as an organ which responds instantly to perpetually incomplete information from the environment by simply "filling in the blanks" based on past experience. If that is so, it is not surprising that our students are disciplined to subject their ideas to the painstaking and potentially corrosive tests of logic and evidence.

If we want them to develop this important but unnatural habit of mind, we may have to balance our legitimate efforts to build student "self esteem" (by trying to find something of value in all of their efforts) with a consistent demand that their ideas must ultimately stand the tests of logic and evidence. Even in areas (most of them) where there are a number of legitimate interpretations, we may have to impose some outside limits on our compassionate inclination to say "that's an interesting idea." Instead, we may have to join more often with a crusty old English teacher of my acquaintance in saying, "It is true, Mr. Daly, that there are a number of reasonable interpretations of this poem. Unfortunately, yours is not one of them."

If we want students to be able to express their ideas with sufficient precision to be intelligible and persuasive to a variety of audiences, we may have to require them to do a good deal more writing, and, in particular, a good deal more expository/argumentative writing than most of them currently do.

While the exact nature of the process is still very much in doubt, it is clear that the enormously complicated task of acquiring spoken language seems to come easily and naturally to most humans. In addition, our students are aided in face-to-face spoken communication by such factors as inflection, facial and physical gestures, the opportunity for mid-course corrections and elaborations, and the frequent familiarity of the listener with the thinking of the speaker even before he speaks. Perhaps the central reason why spoken communication seems to come so naturally and easily to our students is that the above factors create in the listener a considerable capacity to "fill in the blanks" even in this inevitably incomplete and ambiguous mode of communication.

As an experienced classroom teacher can attest, however, written communication is not nearly so natural for most students. There, the message must be much more complete and precise in the text, because the audience may not be personally familiar with the thinking of the writer, may lack any of the interactive aids available in face-to-face communication, and therefore may have a much more limited capacity to fill in the blanks.

The problem is compounded when students move from descriptive or narrative writing based on personal experience to the kind of expository writing which is central to success in school and in professional level work. Expository writing and argumentation often require the manipulation of information distant from the writer's direct personal experience. They also require skill in the use of logic and evidence. And both those sets of skills, as already noted, may come naturally to neither the writer nor the reader.

As a result, if we are to overcome our students' natural preference for oral communication and help them to develop the much more difficult skill of effective expository writing, we will have to require them to do lots of it. And we will have to be willing not only to grade their efforts but also to comment on the specific strengths and weaknesses of their work — quickly, repeatedly, and in detail.

Finally, the development of our students as independent thinkers will probably require us to substantially reduce our reliance on multiple-choice testing, and the instructional practices keyed to it, in favor of assessment methods which evaluate and reward the full range of the thinking activities necessary to independent thinking. The central problem with multiple-choice testing, as a way of measuring and rewarding independent thinking, is that it can effectively measure only one of the four elements discussed in this essay.

The time limitations on most multiple-choice tests preclude any assessment of abstract thinking — the construction of abstract concepts and their use to cull, organize, and manipulate large amounts of information. Similarly, the need for a single, predetermined right answer precludes any assessment of the capacity for creative or novel thinking which, by definition, might produce an answer unanticipated by the test makers. Finally, the use of prepackaged answers and machine-scorable answer sheets precludes any assessment of the capacity for the precise written expression of thought.

The multiple-choice format is, of course, well adapted to measure the recall of specific information. But, with respect to the elements of independent thinking, it is well adapted to measure only the third element of such thinking — systematic or logical thinking — because it is the only element which produces something approaching the single right answer which the multiple-choice format requires.

The same tendency to truncate the thinking process at both ends is characteristic of teaching methods used to prepare students for multiple-choice tests. Drill in standard math problems asks students only to recognize and apply the appropriate logically deductive steps to solve problems which have already been selected and set up for them. They are not required to select the critical elements of a real-world situation, and translate them into a mathematical representation of the problem, before carrying out the mathematical procedures on which they are being drilled. Nor are they normally required to reapply their answer back to the real-world problem, check it for reasonableness and usefulness, and explain it verbally to others.

Similarly, in the verbal area, standard courses in logic frequently focus on evaluating the logical coherence of an argument or series of statements constructed by someone else. As a result, students receive no training in the identification of questions worthy of investigation, in selecting and organizing relevant information, or in initially formulating a viewpoint which can then be subjected to the rules of logic and evidence. Nor, at the other end of the thinking process, do they receive any training in presenting and defending that view to others.

In spite of these weaknesses, the use of multiple-choice testing continues to expand rapidly, not only in the classroom but also in the state and national testing programs born of demands for greater educational "accountability." This popularity is based, of course, on the time, effort, and cost efficiency of multiple-choice tests as a method for evaluating large numbers of students. It is also based on the capacity of such tests to generate "objective" scores which are convenient for ranking the performance of individual students or groups of students.

But if independent thinking is as difficult and frightening for most students as this essay argues, we may have to employ a full set of rewards and punishments to induce students to make the requisite effort. And that will almost certainly require some movement away from the

convenience of multiple-choice testing toward evaluation and grading systems which actually measure and reward the difficult skills we want our students to develop.

Such are some of the classroom implications of the argument that independent thinking is an unnatural act. There are also major implications for those, a little higher up, who plan curricula, select instructional materials, and allocate the efforts of teachers.

Effective institutional efforts to develop the capacity for independent thought will probably have to involve thinking-oriented instruction across content areas. If independent thinking does run against the grain of our students' natural inclinations, they are likely to develop skill in this special way of thinking only if they confront the demand for it, and help in meeting that demand, in all or most of the classes they take.

This implies that instructional programs cannot rely on the creation of a few special classes in thinking skills. We will need to change the way we teach all or at least many of our content courses so that students practice thinking skills at the same time as they develop content mastery. Such a thinking-across- the-curriculum approach would ideally require students, for each major topic which the course covers, to: use general concepts to cull and organize the available information, construct ideas of their own, support those ideas logically and with available evidence, and express the results of this thought process effectively in writing.

Effective instructional efforts to develop the capacity for independent thinking will probably have to involve thinking-oriented instruction across grade levels. The difficulty of changing established ways of thinking has long been noted in folk wisdom and documented by psychological experiments. More recently, the "connectionist" theory of the brain has raised the possibility that the strength of such predispositions derives from the fact that they become rooted, by experience and practice, in the physiology of the brain itself. All of this would seem to suggest that attempts to develop the capacity for independent thought will have a much better chance of success if they are undertaken as early in a student's development as possible.

If success in the effort to develop a capacity for independent thought requires thinking-oriented instruction across content areas and across grade levels, then it will also require a major in-service and pre-service teacher training effort. Independent thinking is a difficult, acquired ability not only for students but for all of us. Many teachers will have to be trained, first, to do it and then to teach it.

Also, much more attention will have to be given to providing teachers with continuous content updating in the exploding knowledge areas in which they teach. Only such an effort to build and maintain their status as genuine up-to-date experts in the areas in which they teach will give them the classroom authority and self confidence necessary to free them from a desperate dependence on the textbook and teacher's guide, and to embolden them to use some of the more exploratory teaching methods necessary to thinking-oriented instruction.

A closing caveat. A model of independent thinking which emphasizes using abstract concepts to order information, subjecting ideas to the tests of logic and evidence, and communicating ideas by means of expository writing is probably more appropriate to the natural sciences and some of the social sciences than it is to the arts and some of the humanities. It is important to note explicitly the resultant limitations of this definition and to say that developing the capacity to think in this way is certainly not the only thing that American educators need to do for their students.

But both tradition-oriented and market-oriented academics should be able to agree that this is one of the most important capacities which we should help our students to develop. This way of thinking has been, after all, a central element of the Western intellectual tradition since classical Greece. And, as noted at the beginning of this essay, it is now sought with a growing sense of urgency by many of the potential employers of our students.

The central argument of this essay is simply that those of us who choose to pursue this important goal must adapt our teaching methods to the emerging evidence that we will have to overcome something no less fundamental than the way in which the unschooled human mind normally functions. The classic line from the Pogo comic strip should shape both our expectations and our instructional strategies: "We have met the enemy, and they are us."

NOTES

1. For representative and influential examples of the large and growing bodies of critical literature on schools, colleges, and workforce training, respectively, see: The National Commission on Excellence in Education, A Nation at Risk (Washington, D.C.: U.S. Department of Education, 1983); F. Newman, Higher Education and the American Resurgence (Princeton: Carnegie Endowment for the Advancement of Teaching, 1985); and W. Johnston and A. Packer, Workforce 2000 (Indianapolis: Hudson Institute, 1987).

2. The best overview of thinking-oriented instructional programs, and one to which this essay is heavily indebted, is R. Nickerson, et.al. , The Teaching of Thinking (Hillsdale, N.J.: Lawrence Erlbaum Associates, 1985). See also P. Chance, Thinking in the Classroom: A Survey of

Programs (New York: Teachers College Press, 1986); and J. Baron and R. Sternberg, Teaching Thinking Skills (New York: Freeman, 1987).

3. For a manageable introduction to Piaget's highly influential work, see J. Flavell, Cognitive Development (Englewood Cliffs, N.J.: Prentice-Hall, 1971). A good sampling of Piagetian instructional programs is provided in R. Fuller, et. al., Piagetian Programs in Higher Education (Lincoln, Ne.: ADAPT Program, 1980).

4. The apparent model for much of the large body of creativity/problem solving literature is G. Polya, How To Solve It (New York: Doubleday, 1957). Other influential texts are: J. Adams, Conceptual Blockbusting (Reading, Ma: Addison-Wesley, 1986); and J. Hayes, The Complete Problem Solver (Philadelphia: Franklin Institute Press, 1981).

5. Text on formal and informal logic are legion. An early and influential example is M. Beardsly, Thinking Straight (Englewood Cliffs, N.J.: Prentice-Hall, 1975). See also J. Cederblom and D. Paulsen, Critical Reasoning (Belmont, Ca.: Wadsworth, 1982); and D. Walton, Practical Reasoning (Totowa, N.J.: Rowman and Littlefield, 1990).

6. For overview of the literature which relates language and thought, see: L. Gregg and E. Steinberg (eds.), Cognitive Processes in Writing (Hillsdale, N.J.: Lawrence Erlbaum Associates, 1980); and E. Maimon, et. al., Thinking, Reasoning, and Writing (New York: Longman, 1988).

Authentic Assessment in Social Studies

Jack L. Nelson
Professor, Graduate School of Education
Rutgers University, New Brunswick, NJ 08903

Social studies is among the more complicated and sophisticated subjects taught in the schools. It relies on students having acquired a basic understanding of symbols such as written and spoken words and relative numbers. However, social studies is essentially an interpretive subject, one which expects students to go beyond rote recitation and the demonstration of performance skills to exhibit comprehension of concepts and ideas, understanding of ethical dimensions of human experience, and critical thinking. Conceptual ideas, ethical decision-making, and critical thinking are the foundations of significant social education knowledge. Individual and social improvement, twin goals of social education, depends upon these fundamental social studies concerns. Authentic assessment addresses these concerns; inauthentic assessment does not.

This social studies knowledge is interpretive in the sense that there may be no specific right answers to the most important topics, but there are many divergent views which should be examined in the process of becoming educated. Social studies knowledge is, then, always potentially controversial even though there may be general agreement on some information at certain points in time or in certain locations. For example, conceptual ideas like justice, equality, freedom, and responsibility are among the most important social studies knowledge, but well educated social studies students recognize that the defining qualities of these ideas vary in time, space, and situation. A test consisting of a vocabulary list of significant social studies concepts with single correct answers does not adequately judge social studies learning of these as interpretive knowledge.

Similarly, an adequate understanding of ethical dimensions of human experience requires that students interpret individual and social behaviors, using value bases which may vary by conditions. The test of what is right or wrong, good or bad, or proper and improper necesitates moral reasoning to explain positions and arguments. That essential test is missing in multiple-choice or short answer evaluations. Further, the standard form of essay tests which seeks conformist responses in a preset format may not provide for ethical reasoning opportunities and, thus, is likely to misrepresent the student's actual social knowledge of ethical issues.

Critical thinking, among the most valuable of social studies objectives, is a third area where typical assessment techniques are unsuited to the purpose. Sound critical thinking includes divergent views, contrasting evidence, and opportunity to present views which dissent from conclusions drawn by the majority. That does not fit well with instant assessment devices. As Banesh Hoffman in The Tyranny of Testing showed years ago, there is some evidence that students who engage in critical thinking on exam items actually suffer penalties and obtain lower scores.

It is on these and other grounds that social studies should demand more authentic assessment approaches. The ones currently in vogue in the United States are inauthentic in that they exempt, ignore, or demean the most important of social studies knowledge.

Assessment of students continues to be a knotty problem in education. And the further one gets from memorization or simplistic physical manipulation, the more problematic one assessment becomes. It is much easier to find an adequate assessment instrument to judge whether or not a child can recall the exact alphabetical order of a list of letters than to judge that child's reasoning talents. It is also far easier to measure the number of times a child puts a square block into a square hole than to measure what the child is thinking about that process. It is obvious that the most significant education task involves the development of thinking talents, the primary job for social studies, despite the fact that assessment of that area is the most difficult.

Does the problematic nature of such assessment lead to a decision to (1) continue the significant social education work of intellectual development regardless of assessment quality, (2) experiment to seek improved assessment techniques before requiring their use, or (3) tailor the social studies to meet currently available measures? Unfortunately, the tailoring of social studies to meet current and seriously deficient assessment programs seems to be winning. New Jersey and other states seem extra eager to provide single test scores which are supposed to represent student knowledge despite the extensive evidence that such devices are insufficient and misleading. Social studies and other subjects are bent to the measure, no matter its deficiencies. At loss is the continuation of social education aimed at developing critical thinking, ethical decision-making, and the refinement of conceptual ideas.

In the grand and convoluted scheme of things educational in the United States, although one would expect the more sophisticated mental processes to be those most valued and important for schools, the easily measured things have become the touchstones for judging schools, students, and teachers. This intellectual turn-around may suggest some very unpleasant things about the quality of education provided to those who are now demanding more testing and then using test scores to bash schools and teachers. Somehow, for this group, we have not been very good at social education that should have made them knowledgeable and, thus, skeptical of the quality of testing available. We have not given them sufficient education to make them critical thinkers about the political and ideological setting of the testing movement.

The current craze for standardized test scores continues a nearly century-long quest to find some measures of human mental ability (Wolf, et al. 1991). Relatively scientific efforts by psychologists in the early 20th century to ascertain differences in mental capabilities among people, often on racist or sexist grounds, were not a blinding success. We have some rough measures and approximations of base intellect, but little that provides sound educational reasons for precise separations among individuals. The best universities recognize that SAT or ACT test scores are not as good as predictors of student achievement in college as are previous high school grades. And motivation for comparable students, which may be a better predictor, is not assessed by standardized tests. For those not going on to higher education, we have virtually no valid assessment indicators that will help them in employment or life. We have some adequate measures of recall of specific information or select skills, but no adequate measures of critical thinking or reasoning about social or individual problems (Nelson, Carlson, Palonsky, 1993).

The scientific ethos under which early testing began has been replaced by the more powerful political and ideological effort to reduce educational accountability to a single test score. The politics of this testing phenomenon include appeals to taxpayer groups to keep down public spending on schools, attempts to become the "education-president" or "education governor" by showing how bad schools were under previous administrations, and stemming the political power of teacher unions. Ideological interests underlying the testing movement include essentialism in preserving the traditional cultural heritage against more contemporary multiculturalism or innovative conceptions of knowledge, a traditional anti-intellectualism in the society, positivism with its behavioristic and mechanical conception of humankind, and an effort to de-professionalize teachers (Apple, 1985; Nelson, 1983; Shwartz and Viator, 1990; Wolf, et al, 1991).

The underlying simple-minded concept that single score data show either the quality of the students, the quality of teaching, or the broad quality of education has been generally accepted in the public and the media. The concept has been challenged in esoteric and self-serving educational publications, such as **The Docket**, but it is very difficult to get the mass media to express such views because test scores are so clear and simple. And the public has been misled to believe that the test scores are valid and easily comparable; how do your children and your schools and your teachers compare on a single set of numbers?

The widely reported decline in average SAT scores over the past two decades hides the more complex factors which lie behind the scores, e.g., proportion and kinds of students taking the tests, social conditions which mitigate against the kinds of information tested, cultural bias, test validity in regard to real life, decline in real dollar educational support, long-term neglect of schools and short-term expectations. Similarly, the widely reported and uncritically accepted data that United States students rank near the bottom of average scores from some 19 industrial democracies on comparative tests of math, science, and basic information did not include the data that the U.S. proportional spending of Gross Domestic Product was even further down in rank among those industrial democracies. In those terms, we got more for our money than did the other nations, but that comment assumes the tests were valid, that the scores represent something more than instantaneous responses at a specific time, and it is educationally important to compare such scores. These assumptions are not well grounded.

In regard to social studies assessment using standardized instruments, one should be even more skeptical of currently available tests for several reasons: (l) as social studies is among the most complex of subjects because of its focus on human endeavors, it should be among the most difficult to properly assess; (2) social studies has a primary responsibility to educate for critical thinking and

ethical social action, two areas where tests are most weak; (3) social studies content is necessarily controversial with divergent views, making singular right-answer testing inappropriate; and (4) social studies is a long-term developmental activity designed to provide contributing members of a democracy, a purpose not easily subject to short response snapshot-like assessment. Currently available tests in social studies have many debilitating defects. Examine the several editions of the Mental Measurements Yearbook, the single best source of expert reviews of standardized tests, and you will find that standardized social studies tests generally lack suitable validity and have technical deficiencies well beyond those found in tests of math and reading skills. Even if a decent standardized test of social studies information could be found,the educational expectations for social studies noted above should give great worry to anyone who argued that those test scores adequateiy covered the nature of social studies.

The previous material is to make the case that we have a long tradition of inauthentic assessment in education, and that social studies is a field which has not lent itself well to the simplistic form of standardized testing that characterizes much of educational assessment today. There may be some approaches to more authentic assessement that could be more suited to social studies. They may be more complicated and more difficult than standardized tests, but they are probably worth it if we are to pursue social studies purposes. As argued at the start, social studies is a complicated and sophisticated subject, if taught properly. If taught improperly, it is mechanistic, jingoistic, silly, and useless.

Authentic assessment in social studies requires effort directed at its basic interests, operations, and purposes: conceptual ideas, ethical decision-making, and critical thinking about controversial topics related to individuals and society. That approach to assessment demands much more than multiple-choice or short answer tests of recall. Instead, such assessment must be multidimensional to attempt the measurement of how well a student understands and can use complex information to address a variety of individual and social issues in a critically reasoned and ethically grounded manner. The further social studies purpose of active participation in society also must not be lost in the rush to measure students. This approach to assessment is not properly made on a one-time or even once-a-year shot.

Some possible approaches to more authentic assessment:

1. Use long-term student activity as a basis for assessment; for example, Model United Nations, Mock court, student government. This would mean that, in order to give an appropriate assessment for all social studies students, these activities would become available for all students, not just an elite group who are intended to satisfy the ego of the school principal. The intent is to provide active settings for students to try out ideas and thinking about social issues.

2. Ethical decision-making discussion sessions. These would offer students in small groups the chance to explore and explain the values they use in considering ethical questions. Teacher domination needs to be avoided, but debates over the reasoning would be appropriate. The assessment may be progressive, indicating the quality of changes as students rethink their positions.

3. Portfolio of social experiences. This could include evidence of student participation in social life, community affairs, personal development, helping others, etc.

4. Sampling thought over time. This considers the quality of thinking shown by students on a number of topics at different times over the year to measure growth in thinking and evidence use. Comparisons are to judge progress, not singular right answers.

5. Episodic observation. Students are put into social studies issue situations (using such techniques as role-playing, simulation, media use, interactive video). Using journals, diaries, direct observation, and interaction, teachers and students assess qualitative aspects of the learning.

6. Case studies of multi-year development. Using groups of teachers, counselors, librarians, parents, other students, and others in a position to contribute insights, students are assessed individually over several years in terms of their development of social studies ideas, ethical decision-making, and critical thinking. Students themselves are asked for self-assessment and expected to dissent from or agree voluntarily to the views of others. Schools need to be cautious to protect individual students in their private views and to avoid oppressive conformity to peer or teacher views. This is meant to examine the quality of process, not the conclusions drawn.

7. Other forms of qualitative assessment. The ideas suggested, I believe, are closer to authentic assessment consistent with the most important social studies purposes and the long-term nature of adequate education.

References

Apple, M. and Teitlebaum K. (1985). "Ar Teachers Losing Control of Their Jobs?" Social Education. 49:372-375.

Nelson, J., Carlson, K., and Palonsky, S. (1993). Critical Issues in Education. New York: McGraw-Hill.

Schwartz, J. and Viator, K., eds., (1990). The Prices of Secrecy: The Social, Intellectual, and Psychological costs of Current Assessment Practices. Cambridge, MA: Educational Technology Center, Harvard.

Wolf, D. Bixby, J., Glenn, J., and Gardner, H. (1991) "To Use Their Minds Well: Investigating New Formsof Student Assessment" In Review of Research In Education - 17., G. Grant, ed., Washington, DC: American Educational Research Association.

Performance Assessment in Principle and Practice

by
Pat Nickell
Director of Instructional Support Services
Fayette County Public Schools, 701 E. Main St., Lexington, Kentucky 40505

By now, most teachers are familiar with the term "performance assessment." Grant Wiggins (Nickell, 1992) has defined it as "students performing with knowledge... turn the stuff of content into some product or performance." It is a form of student evaluation which requires more than recall of facts or information or replication of a skill; it asks students to apply knowledge and skill for a purpose. "Authenticity" of performance assessment requires further that the purpose closely imitate something the student might be required to do in the real world.

In a number of states, performance assessment is either recently adopted or currently under consideration as a possible replacement for traditional forms of large-scale standardized testing. Rather than becoming part of that debate, however, this author has chosen to give emphasis to classroom-level assessment, with the ideas set forth to guide classroom teachers as they go about the business of developing appropriate means of assessing student learning in particular classes about specific material.

Four principles will be offered which have emerged from the literature on performance assessment, and examples will be provided to enable teachers to recognize how these principles might guide the development of performance events or tasks for use in the classroom. Helpful tips are also included to assist teachers in evaluating student products and responses.

1. Performance events or tasks are virtually indistinguishable from authentic, activity-based instruction.

For years, we have treated instruction and assessment as separate entities. We teach and then we test; or we pretest, then teach, then posttest. The test has long been a discrete component of classroom activity. Performance assessment, for classroom purposes, virtually erases the distinct line between instruction and assessment. In practice, a good performance task for assessment will look just like what has long been recognized as a good activity designed to help students take what is being learned and apply it to the real world.

Ms. Martino has been teaching her seventh grade social studies class to interpret data from a variety of sources — maps, charts, graphs, etc. She incorporates the activity shown in Figure 1 for assessment purposes. Only the students know that this is their "test." An outsider might wonder why she is remaining mysteriously aloof from the students as they are completing the task, but otherwise assume that this is a regular classroom activity.

(Figure 1)
Global Grocery

In our nation's capital, Washington, D.C., the average price of a dozen eggs is ninety-nine cents. In Tokyo, those eggs would cost $1.95 and in Paris you would pay $2.75. A quart of milk in Washington averages fifty-eight cents, in Paris, eighty-seven cents and in Tokyo, $1.31. One would think that rice would be cheaper in Japan than in France or the U.S., but surprisingly, Washingtonians pay only $.49 per kilogram, while Parisians pay $.59 and those in Tokyo pay the most at $1.20! The most expensive item in anyone's grocery cart, however, is steak. A pound of boneless sirloin in the D.C. area averages $5.28. In Paris it runs $6.68, but in Tokyo? TWENTY-THREE DOLLARS AND NINETY-SEVEN CENTS! Now that's an expensive piece of cow!

You work for USA Today. Your editor wants you to present this information in a visual way so readers can understand it more easily and quickly than by reading text as shown above. Use the space below to show how you might do this.

Just in case your editor doesn't like the one you developed

above, think of another way to present the information if you can. It should still be quicker and easier to read and understand than the original paragraph, but somewhat different from your first idea.

A concern with this form of assessment is the degree to which teacher evaluation can be objective and fair. Thus, Ms. Martino has designed a scoring rubric (Figure 2) to enable her to grade student responses and report outcomes both to students and parents. She has previously discussed her expectations with students and thus they are not at all surprised at the grades they receive.

(Figure 2)
Scoring Rubric - Global Grocery

5
* Provides two distinctly different ways to present information
* Both presentations include all information accurately
* Both present information in unique and effective VISUAL ways (charts, diagrams, graphs, pictorials, etc.)
* Presentations are carefully and effectively rendered

4
* Provides two distinctly different ways to present information
* Both incorporate information accurately and completely
* Presentations are predictable, less creative, but clear and visual in format
* Presentation may be creative, but not absolutely complete (indicating lack of time to develop thoroughly)

3
* Provides two similar presentations
* Information presented accurately, but may not be complete
* At least one presentation is visual, both more concise than text
* May summarize text in one presentation, but does so in a clear, concise, accurate manner

2
* May explain how task would be accomplished without doing so
* Presentations are incomplete, inaccurate, or unclear
* May only include one attempt
* May simply summarize or rephrase text, without visuals, for both presentations

1
* No attempt or single statement lacking meaning or unintelligible
* Response indicates no understanding of task

2. Performance assessment is a better gauge of whether students know and can do what is important.

This principle arises from all the arguments against traditional fixed-response testing. Central among these is that simply the accumulation of knowledge may lead to astuteness in trivia games, but cannot guarantee astuteness in solving problems, making wise choices and decisions, figuring out the best means to an end, or evaluating information and occurrences.

Mr. Avery teaches eighth grade history. It bothers him that so few people are able to use historical information to analyze current situations. Thus, he focuses his instruction on applications of historical knowledge. He uses a performance task (Figure 3) at the beginning of the year to pretest his students' ability to recognize an inaccuracy based on lack of historical knowledge. He also uses it to get a sense of whether his students are able to take a position and defend it. Later, he will use the same assessment to reassess these abilities. He will then return both papers to let students see how they have progressed.

(Figure 3)
Rewriting History

A state Attorney General is interviewed on a television talk show. In the interview, he is asked to discuss what he thinks are the most serious problems state governments are having to deal with today that they did NOT have to deal with in the past. In his response, he names alcohol abuse as a serious problem governments have only recently had to address. He states that all sorts of criminal and irresponsible acts can be traced to this recent trend toward overconsumption of alcohol.

Use the information on the attached sheet to help you write a one-paragraph letter of response to the Attorney General. In your letter, you should do three things:

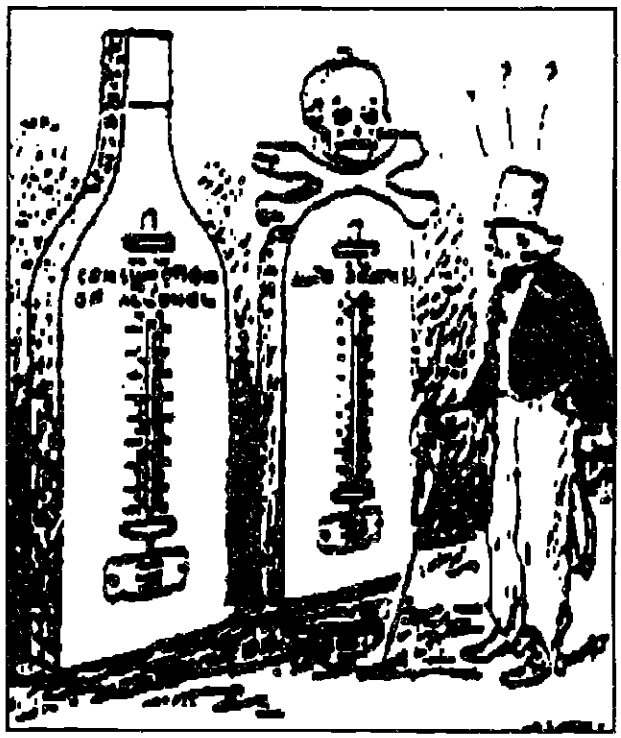
1. respond to the historic accuracy of his statement
2. state your position on the issue of alcohol abuse
3. suggest a solution to the problem

(figure 3 continued on next page)

William Shakespeare was born in 1564 and died in 1616. He wrote many plays and poems. In one of these he included the following line: "Oh God, that men should put an enemy in their mouths to steal away their brains!"

Abraham Lincoln (1809 - 1865) said: "Liquor might have defenders, but no defense. Whether or not the world would be vastly benefitted by a total and final banishment from it of all intoxicating drinks, seems to me not an open question."

A 1939 American Government textbook carried this editorial cartoon.



Going Up Together
What should we do about it?

In 1988, a drunken driver entered I-71 near Carrollton travelling north in the southbound lanes, drove head-on into a church bus carrying 67 adults and children coming home from an excursion to King's Island, killing 27. The news of this accident was carried nationwide over all major TV networks and news organizations.

(figure 3 continued)

Fixed-response test types -- multiple-choice, matching, fill-in-the-blanks -- rarely tell us whether students are able to sort through and interpret information for important purposes, and simply cannot gauge whether students are able to generate appropriately constructed responses, express in writing their evaluation of actions and situation, and justify opinions. Yet, these are some of the things we must all be able to do in order to function well in our everyday lives.

Like Ms. Martino, Mr. Avery has developed a scoring rubric (Figure 4) which he provides to his students when he returns their papers. From it, they are able to recognize that what he expected of them matched what he emphasized during instruction.

(Figure 4)
Scoring Rubric - Rewriting History

5
* Response is well-organized and presented in letter form
* Uses effective language and presentation style
* Uses historical information accurately, taking issue with the Attorney General's statement on this point
* States and develops a clear position
* States and develops one or more suggestions

4
* Response is clear and uses language and mechanics acceptably
* Addresses historic accuracy using supportive data
* States a position, with some explanation or support
* States one or more possible solutions with some explanation

3
* Response is readable but may include several language and mechanics errors
* Addresses historic accuracy, but supportive information may include inaccuracies
* States a position, but explanation is weak or absent
* States one or more possible solutions, weakly or not developed or explained

2
* Response is poorly organized, using ineffective structure, wording, and detail
* Historic accuracy of Attorney General's statement is not addressed or treated inaccurately
* States no position or a poorly developed one
* Offers no solution or a poorly developed one

1
* No attempt
* Response is unintelligible or unrelated to task requirements
* Only provides a single simplistic position or solution

3. Performance assessment is a better match with real-world tasks.

This principle is highly related to Principle 2, above, yet deserves special consideration as teachers attempt to design assessment tasks. Simply stated, no task is worth designing if it isn't worth doing. The test of a good task is whether it calls upon students to do something worthwhile and similar to something that, if done well, would make their lives easier in the real world. Ms. Martino's task is reflective of what we face when information is given to us out of order or in some format that is difficult for us to understand. We may reorganize it on paper or mentally, but only through this process does the information take on meaning for us. Mr. Avery's students will inevitably be required to develop positions and opinions and defend them. They must be able to evaluate opinions and statements of others in light of their own beliefs and values. Finally, it is important for students to recognize and use accurate historical data.

In the process of task design, it is most important to be sure that what has been taught has real-world application and that those applications are what guide the design process.

To put this principle into practice, readers may wish to try the following:

1. Think of what it is you are currently doing with your students. What is your topic? What are your objectives?

2. Why is this important? When have you used this information in the past year, outside the classroom and beyond you personal reading? (If you haven't, are you able to justify its importance?) What did you use it for?

3. How will students use it and for what purpose (other than within your classroom)?

You are now ready to design an assessment task.

4. Create a situation which calls for your students to use the material you are presenting. Write it up in an interesting and authentic fashion. Be specific about what is expected of their response or product. This is a performance task.

5. You need to be fair about grading their work and you want to be able to do it quickly. Therefore, decide what criteria are important for evaluating their success. What do you think they should demonstrate for you? Is it clear what you expect in the way the task is presented? If not, either edit the task or change your criteria. Describe what you think an outstanding response would look like; a good response; an average response; a barely acceptable response; and an unacceptable response. You now have a scoring rubric.

While this is a simplified and sketchy model, it should allow the teacher to recognize some of the critical differences between performance task and fixed-response test types, and to begin the process of task design.

4. Performance assessment should drive instructional as well as testing changes — even for lower performing students.

It is now considered a given that teachers "teach to the test." Shepard (1989), Wiggins (1989) and others claim that the trick is not to get teachers to quit doing so, but to make the test worth teaching. Performance assessment, these theorists claim, is a more worthwhile instrument toward which to gear instruction. It requires high-order thinking, generative responses, and can be formatted for assessing students independently or in groups. We cannot, however, reserve this type of instruction only for advanced learners. All students must learn to solve problems, make decisions, and discern the most appropriate ways to achieve an end. Critics of traditional testing argue that typical multiple-choice approaches serve to "dumb down" curriculum and instruction. "Conceiving instruction in the format of multiple-choice items...leads to endless drill and practice on decontextualized skills" (Shepard, 1989, p.5). This, it is argued, is especially true of instruction for lower-functioning students where test-taking preparation — raising scores — has been given central attention in the schooling process.

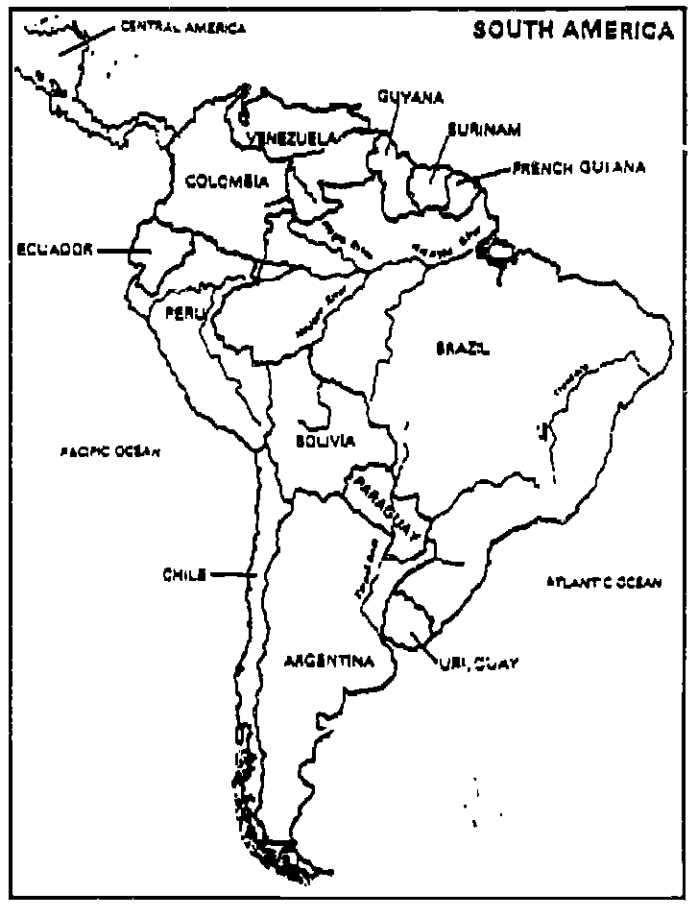
Mr. Spooner teachers a heterogeneous group of sixth graders. His most recent geography unit on Latin America also involved the development of an improved geographic vocabulary. Ongoing efforts have included assisting students in developing a mental map of the United States and providing multiple opportunities for students to make independent judgments and provide logical explanations for them in writing. All of these efforts appear in the assessment task he uses to evaluate their progress and the scoring rubric he has developed to assist him in grading student responses (Figure 5).

(Figure 5)
Landlocked

The Random House Dictionary of the English Language defines "landlocked" as "shut in completely, or almost completely by land; not having access to the sea." Look at the attached map of South America. Using your interpretation of the definition of "landlocked," would you say there are any landlocked countries in South America? If so, name them. Explain your answer.

Picture a map of the United States in your mind. Is Kentucky landlocked? Some people would say "yes," others "no." State YOUR opinion and support it with why you think as you do.

If a country is landlocked, what are several things it might do to gain access to the sea? Explain your response.



Scoring Rubric - Landlocked

5
* Responses are clear, accurate, and well-structured
* Responses supported by related and reasoned information
* Parts 1&2 consistent or differences fully justified
* Part 3 reasonable and fully explained, demonstrating understanding of political implications of sea access

4
* Responses are clear, accurate, and readable
* Responses are supported by related information
* Parts 1&2 are consistent or differences justified
* Part 3 offers one or more suggestions, explained, but may not exhibit understanding of political importance of sea access

3
* Responses readable and acceptably structured, but may include minor inaccuracies
* Responses to parts 1&2 may be inconsistent, but both are supported
** Response to part 3 provides one or more underdeveloped or weak idea, perhaps unreasonable
* Part 3 indicates little understanding of the importance of sea access, other than for saltwater resources or recreation

2
* Responses to parts 1&2 inaccurate and/or unsupported by related information
* Responses are poorly structured
* Parts 1&2 are inconsistent; one or both unsupported
* Part 3 poorly developed, unrelated, or shows no understanding of the need for sea access

1
* No attempt or responses irrational or unrelated to the task

Student responses to this task may look very different depending upon the students' educational background and prior knowledge. However, the skills required are ones that all students should have in order to read and understand a newspaper, participate in normal conversations referencing places in the U.S., posit an idea and provide an explanation, and communicate in writing. If all our students are "raised on" challenging tasks from early childhood, then by the time they are in middle school, even our "at risk" populations of normal intelligence will have little difficulty addressing tasks such as those offered here.

Conclusion

The principles presented here are designed to provide rationale and guidance for the development of performance tasks for classroom use. It is hoped that teachers will give these ideas serious thought, especially in terms of what we know is best for children: that they deserve not only interesting and high-quality instruction, but equally interesting and high-quality evaluation instruments; that they deserve to have us care about what they really know and can do; that they deserve practice in dealing with real-world challenges; and that all students deserve these things...not just those thought to have special gifts and promise.

Bibliography

Nickell, P. (1992). Doing the stuff of social studies: A conversation with Grant Wiggins. Social Education. 56(2). 91-94.

Shepard. L. A. (1989). Why we need better assessments. Educational Leadership. 46(7). 4-9.

Wiggins, G. (1989). Teaching to the (authentic) test. Educational Leadership. 46(7). 41-47.

Performance Assessment in Social Studies: What CRESST Research Tells Us

by
Pam Aschbacher and David Niemi
National Center for Research on Evaluation Standards and Student Teaching
UCLA Graduate School of Education, 405 Hilgard Ave., Los Angeles, CA 90024

New Goals, New Assessments

Over the past decade there has been a significant reexamination of the outcomes for which schools are to be held accountable. We have moved away from a strong, narrow focus on minimum competency in the basic skills of reading, writing, and math to talk of meeting world class standards across the full curriculum, including social studies. The importance of student achievement in social studies was underscored by the Governors' National Education Goals Panel and the President's America 2000 Education Strategy, both of which specifically mentioned history and geography as two of five subjects representing the core content that all students need to master in school.

Increasing Interest in Social Studies Assessment

Many states are now interested in expanding their statewide testing programs to include social studies content (Aschbacher, 1991) Currently, 19 states have representatives in the National Social Studies Assessment Consortium (NSSAC) and/or the State Collaborative for Assessment and Student Standards (SCASS), two organizations established by the Council of Chief State School Officers to collaboratively develop new performance assessments. To guide these collaborative efforts to develop new assessments, the staff of the National Center for Research on Evaluation, Standards, and Students Testing are currently compiling a consensus map of the NAEP and state curriculum and assessment frameworks in social studies (including history, geography, and civics), which should be available by early 1993.

Local education agencies have also generated a lot of interest and activity around performance assessment. Part of this interest stems from the growing belief that instruction across the curriculum should emphasize "deep understanding," "higher order thinking skills," and more "authentic," meaningful activity—not just rote acquisition of information and procedures. These new goals in turn require new forms of assessment that capture what we would like students to be able to do and understand.

A number of states, such as New Jersey, are encouraging districts to move ahead with development of their own performance assessments because the current economic situation hinders development at the state level. In states with new statewide performance assessment programs, like California, many districts are working to develop their own parallel classroom, school, and district level assessments. But developing good assessments requires expertise, time, and resources, all of which tend to be scarce. Thus it is important for those who have been working in the field of performance assessments for the past several years to share what works. Since we at CRESST have been researching performance assessments in social studies for the past six years, we have three kinds of information that may be useful. First, we describe some of our basic premises about developing performance assessments. Second, we present a specific example of one type of performance assessment we have developed in history and other content areas. Third, we describe how history teachers have used our assessment task as the basis for designing authentic instruction.

Some Basic Premises in Developing Performance Assessments

Although performance assessment implies new methods for looking at educational outcomes, the process for developing these assessments is based on decades of measurement research and involves the following tasks:

- specifying the nature of the skills and accomplishments students are to develop;
- specifying the kind of tasks that would require students to demonstrate these skills and accomplishments;
- specifying the criteria and standards for judging student performance on the task;
- developing a reliable rating process;
- gathering evidence of validity to show what kinds of inferences can be made from the assessment; and
- using the results to refine the assessment, improve curriculum and instruction; and provide feedback to students, parents, and the community (Herman, Aschbacher, and Winters, ASCD, 1992).

One of the major purposes of performance assessments is to help teachers make instructional decisions. This means that the assessment tasks and criteria for judging student performance should be aligned with our desired goals for student achievement. What do we want students to be able to accomplish in a unit, a course, a discipline, or across disciplines? The answers define our classroom priorities for both instruction and assessment.

A common procedure for developing performance assessments is simply to assemble interesting activities—preferably designed by teachers and other content experts—and expect that they will tell us what we want to know about students. Chances are, they won't, because of the hit-or-miss quality of the design process. Without deciding in advance what the tasks are supposed to assess, and without some preliminary thinking about how student responses will be scored, we may not find out what we want to know from the assessment.

The criteria we use to judge students' performance on an assessment task should help teachers know what to do next with students. For example, the scoring scheme for rating essays in history should not just label student work as "unacceptable," "adequate," or "advanced." Its dimensions and examples of scored work should give both teachers and students information that might help students elaborate and expand their understanding in history.

Our approach to assessment is based on a constructivist model of learning. In this model, students build personal understanding of a subject by relating new information to what they already know. They demonstrate this understanding by using their knowledge in a variety of contexts and ways.

Students do not come to understand history, for example, by memorizing or learning to paraphrase what teachers tell them or what they read in textbooks. Instead they need opportunities to reflect on and organize what they know, to talk and write about what they know, and to compare perspectives with other students and adults. (See Resnick and Klopfer, ASCD, 1989, for more information on this approach to learning and assessment.)

We know from many studies of cognition, including our own comparative research on essays written by history experts and novices, that students' understanding of history will develop in a certain way. As they move toward expertise, students will not only acquire more factual information, but they will begin to organize that knowledge around historical concepts or principles, such as "states rights" or "constitutionality." They will be able to integrate knowledge about many different historical periods and to relate that knowledge to their own personal experience, and to knowledge about other subject areas.

Research on expertise and how it develops, lies at the heart of our performance assessment model. At CRESST, we have been working for several years to construct a model for designing assessments of deep understanding in social studies, and both the model and tasks designed in accord with it have been validated in several large-scale research studies (e.g, Baker, Freeman, & Clayton, 1991).

The tasks we describe are the result of six years of research intended to develop and test alternative assessments in history, and are documented in our recent handbook (Baker, Aschbacher, Niemi, & Sato, 1992). Historians and high school teachers have helped us throughout the study as designers, reviewers, and scorers. So far, six complete sets of history assessments have been developed: two on the Revolutionary period; one on the Civil War; two on 20th century immigration; and one on the Depression Period. All of these topics connect to the *California History-Social Science Framework* (1988). Our work to date has involved students from grades 8 through 12, but we are expanding it this year to include 5th grade students.

Our assessment tasks require students to: 1) recall

prior knowledge in a particular area of history, such as the Civil War era or the Great Depression; 2) read primary source documents containing new information; and 3) write an explanation of important issues that integrates new and prior information. This three-step sequence takes one-and-one-half to two hours per topic.

In the first step, students take a 20-item, short-answer test (see example below) that measures conceptual and factual knowledge. Students are directed to write down whatever they know about terms like "sectionalism" or "bleeding Kansas." This is prior knowledge, or knowledge that students have before reading the texts we will give them. One of the purposes of this test is to activate knowledge that students might use later in their essays.

Next students read two primary source texts, typically speeches by historical figures, that express contrasting points of view. Finally, students are asked to write an essay that explains the positions of the authors or speakers of the texts, and to use what they know about U. S. history in their explanation.

An example of the writing assignment we have used for immigration is shown below. Prior to receiving this prompt, students will have read the two speeches referred to in the prompt.

Writing Assignment

Imagine that it is 1876 and you are an educated citizen living in California. Because you are interested in immigration, you make a special trip to hear the debate about Chinese immigration between Frank Pixley, the attorney representing the city of San Francisco, and B. S. Brooks, the attorney on behalf of the Chinese. When you return home, your cousin asks you about some of the concerns about Asian immigration in the country.

Write an essay in which you explain the most important ideas and issues your cousin should understand. Your essay should be based on two major sources: (1) the general concepts and specific facts you know about American History, and especially what you know about Asian immigration to this country; and (2) what you have learned from the readings.

Be sure to show the relationships among your ideas and facts.

Scoring Performance Assessments

Our scheme for scoring student responses to this assignment is based on a comparison of the writing of expert historians (high school history teachers, university professors, and graduate students) with that of novices (average high school students). It judges student understanding on the basis of six scales:

- General Impression of Content Quality (focused on the overall quality of historical understanding);
- Prior Knowledge (facts, information, and events that did not come from the primary source texts);
- Number of Principles or Concepts (the number and depth of description of principles);
- Argumentation (the quality of the argument, its logic and integration of elements);
- Text (the use of information from the text for elaboration);
- Misconceptions (the number and scope of misunderstandings in interpretation of the text and historical period).

Each of these dimensions is scored on a 0-5 point scale.

What CRESST Has Learned About the Uses of Performance Assessment

In the process of testing our explanation tasks with students, one of the first things we discovered is that many students were having trouble with them. A surprising number of students reported that they did not know what to do because they had never written a history essay before. Others seemed to have little or no understanding of the historical period we were asking them to write about. We decided that we had to try to do something about this. Our solution was to share what we knew about performance assessment with high school history teachers and to discuss with them how instruction might improve students' historical explanations. All teachers agreed that they would like their students to able to write well-reasoned, well-supported explanations.

To investigate ways that our assessment research might support instructional improvement, we have introduced eleventh grade history teachers in three school districts to our performance assessment model. With the teachers we designed two-day instructional sequences based on our assessment model and principles of cognitive learning. Even though we believe that our assessment tasks would also make good learning activities, we decided that we did not want to focus instruction on essay-writing, since we thought that teachers and students might concentrate on writing skills rather than historical concepts and information.

The instructional activities we designed included, among other things, concept-mapping (for a good introduction to this technique, see Novak and Gowin, 1984), small-group discussions of historical issues, role-

playing, and family history research. These activities were intended to help students access, organize, and elaborate on what they knew about the topic in a way that would lead to better performance on our explanation tasks. (More detailed lesson plans are available from CRESST.)

During the 1991-92 school year, eight teachers tested the experimental instruction and assessment tasks in two topic areas: the Civil War era and Immigration. Prior to assessment in each topic, half the students in the study participated in the experimental instruction activities we designed. The other half studied in a conventional textbook-based manner; this was our control group. After instruction, all students wrote 45-minute essays. Our analyses of these essays suggest that as little as two days of instruction based on principles of cognitive learning can measurably improve students' performance on explanation tasks. This is an extremely promising finding. It demonstrates not only that a performance assessment model can be a useful tool for teachers in designing instruction, but also that the model can help students to develop a more principled, more organized understanding of history, and to demonstrate that understanding in writing.

References

Aschbacher, P. R. (1991). Performance assissment: State activity, interest, and concerns. *Applied Measurement in Education*, 4(4), 275-288.

Baker, E. L., Aschbacher, P. R., Niemi, D. & Sato, E. (1992). *CRESST performance assessment models: Assessing content area explanations*. Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Baker, E. L., Freeman, M., & Clayton, S. (1990). *Cognitive assessment of subject matter: Understanding the marriage of psychological theory and educational policy in achievement testing (Tech Rep. No. 317)*. Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing (CRESST).

California Department of Education. (1988). *History-social science framework for California Public Schools, kindergarten through grade twelve*. Sacramento, CA: Author.

Herman, J. L., Aschbacher, P. R., & Winters, L. (1992). *A practical guide to alternative assessments*. Alexandria, VA: Association for Supervision and Curriculum Development.

Novak, J. D., & Gowin, D. B. (1984). *Learning how to learn*. New York: Cambridge University Press.

Resnick, L. B., & Klopfer, L. E. (1989). Toward the thinking curriculum: An overview. In L. B. Resnick & L. E. Klopfer (Eds.), *Toward the thinking curriculum: An overview* (1989 yearbook of the Association for Supervision and Curriculum Development) (pp. 1-18). Alexandria, VA: Association for Supervision and Curriculum Development.

THE DOCKET

NJ Council for the Social Studies

P.O. Box 4155 • River Edge, NJ 07661

Authors wishing to contribute to The Docket should submit manuscripts to: The Docket, P.O. Box 4155, River Edge N J 07661. Attention: Editor