

EMBARGOED-ADVANCE COPY Not for Public Release Before Thursday, July 16, 1998, 12:00 p.m. EDT
Please Cite as a Report of the National Research Council

NATIONAL RESEARCH COUNCIL

COMMISSION ON BEHAVIORAL AND SOCIAL SCIENCES AND EDUCATION

2101 Constitution Avenue Washington, DC 20418

BOARD ON TESTING AND ASSESSMENT
Michael J. Feuer, Ph.D., Director

202-334-3087
202-334-3584 FAX

July 16, 1998

Secretary Richard W. Riley
U.S. Department of Education
Room 6263
600 Independence Avenue, SW
Washington, DC 20202

Dear Mr. Secretary:

As you know, the National Research Council (NRC) has been conducting an independent evaluation of certain technical aspects of the development of items for the proposed Voluntary National Tests of fourth-grade reading and eighth-grade mathematics. To carry out this mandate, specified in P.L. 105-78 (November 1997), the NRC appointed us as co-principal investigators. We are working with NRC staff under the auspices and oversight of the NRC's Board on Testing and Assessment (BOTA) and soliciting input from a wide range of outside experts. Please note that we interpret our mandate as a request for technical review only, and we take no position on the overall merits of the Voluntary National Tests.

Under the NRC contract with the Department of Education, we are scheduled to issue a report of our first-year findings in September 1998. However, we have identified an issue that we believe can benefit from immediate attention, and we are sending this letter in the hopes of ensuring the best possible test development process. *Our principal concern is with the timetable for item review and revision, which, if relaxed somewhat, would allow more time for the full benefit of those activities to be realized.* This letter is based on our evaluation of the materials available to us as of early June 1998: we realize that because the development process is ongoing, some of the issues identified in this letter may have already been addressed by the time it is released. In the interest of providing constructive feedback, however, we choose to express our concerns now even if this proves to have been unnecessary, rather than delay until September and risk being too late to effect positive change.

EMBARGOED-ADVANCE COPY Not for Public Release Before Thursday, July 16, 1998, 12:00 p.m. EDT
Please Cite as a Report of the National Research Council

Our letter is organized in three main sections: a brief discussion of the scope and purpose of the evaluation; a description of the workshop that focused on item quality; and our findings, conclusions, and recommendations. We trust this information will be helpful and interpreted in the spirit we intend, as a suggestion that more time be allowed for item review and revision.

Scope and Purpose

In his January 1997 State of the Union address, President Clinton called for the development of Voluntary National Tests (VNT) in fourth-grade reading and eighth-grade mathematics as a means of providing information on the academic progress of American youth in relation to national and international benchmarks. Congress gave the National Assessment Governing Board (NAGB) exclusive authority to develop the VNT and asked the NRC to provide an independent evaluation of the development activities.

In this first phase of our evaluation (which is scheduled to culminate in a report in September 1998), we have focused largely on three aspects of the test development process and products: issues surrounding the test specifications and the NAEP frameworks; plans for the development and implementation of the pilot study scheduled for spring 1999; and preliminary evidence of the quality of possible test items. To date, we have observed laboratory-based talk-aloud item try-out sessions with students, reviewed design and development plans and reports, examined draft test items and scoring materials, and conducted three workshops in which additional experts with a wide range of skills and experience have contributed unique and invaluable input. This letter pertains exclusively to our concerns about the item review and revision schedule for the next few months; it is based on our most recent workshop (June 2-3, 1998).

VNT Item Development and Review Schedule

It is important to note that NAGB and its contractors for test development have not completed the development of candidate items for the VNT. Because of a 6-month delay in initiation of the project, NAGB, its prime contractor (the American Institutes for Research), and the subcontractors for reading and mathematics test development have faced a daunting and compressed schedule for test design and development. (Although the schedule for development, review, revision, and approval of materials has not been rigidly set, approximate dates have been specified.) Revised test specifications were approved by NAGB at its March meeting and item development continued thereafter. The subcontractors for reading and mathematics development relied on the NAEP frameworks, the VNT specifications, released NAEP items, and bias and sensitivity review guidelines in writing items. Each subcontractor appears to be following the development and quality control procedures used in its proprietary testing programs. After initial drafting, items then undergo review and revision.

The item review and revision process includes a set of sequential and overlapping steps:

1. initial review of items by the subcontractors for reading and mathematics development;
2. content review by the prime contractor and its consultants;

EMBARGOED--ADVANCE COPY Not for Public Release Before Thursday, July 16, 1998, 12:00 p.m. EDT
Please Cite as a Report of the National Research Council

3. review by outside content experts;
4. trial evaluation of a subset of items in one-on-one talk-aloud sessions with students (called cognitive laboratories);
5. to the extent possible, provision of recommendations for item revision to item writers based upon summaries of information obtained from steps 1-4;
6. review of items for bias and sensitivity by consultants to the contractors;
7. revision of items by the item writers;
8. review of test items by NAGB; and
9. NAGB sign-off on test items.

Steps 1-5 above were to be completed by June 30th, so that revisions to at least a subset of the approximately 2,600 items were possible prior to bias review that was scheduled to take place July 6-8. This bias review and the previously provided review information was to be the basis for additional item revision prior to NAGB's review and approval. NAGB is scheduled to review items in three waves, one beginning July 15, the next beginning on July 22, and the final set beginning on July 29. Approval of items by the Governing Board will be sought at the November 1998 NAGB meeting. Approved items then will be assembled in draft test forms for the proposed pilot administration of the VNT in spring 1999. It is unclear what activities NAGB plans to undertake for its review between July 15 and the November meeting.

Workshop on VNT Item Development

Our concern with the scheduling of item review and revision activities results primarily from data obtained in our third workshop. On June 2 and 3, a group of outside experts with experience in the development and evaluation of conventional and performance-based test materials (see the list of outside experts at Attachment 1) examined and rated a subset of secure test items in their area of expertise. The test developers supplied a list of items that represented the VNT specifications for content coverage and item formats. We selected a total of 60 mathematics items from a pool of 120 items provided by the prime contractor. Similarly, we selected 6 reading passages with a total of 45 items from a pool of 12 reading passages with roughly 120 questions. The sample of items we selected matched the VNT specifications for the length of a test form as well as for content coverage and item formats.

These materials and the workshop participants' observations provided the data that led to this report. Because the items we examined in June were products of the initial stages of item development, we did not expect them to reflect the complete development process. Many items had been through content review, and a number were being tried out with small numbers of students to assess item clarity and accuracy. However, the items had not yet been reviewed for bias or sensitivity, revised by test writers, or submitted for NAGB review and approval.

Before examining the test materials, the experts, co-principal investigators, and NRC staff signed nondisclosure statements promising to protect the confidentiality of the materials. Consequently, specific illustrations of our findings cannot be provided without breaching the security of these materials. When analysis for the broader evaluation is complete, however, we and the NRC staff will be happy to share specific examples with NAGB members and staff who already have access to the secure test materials.

EMBARGOED-ADVANCE COPY Not for Public Release Before Thursday, July 16, 1998, 12:00 p.m. EDT
Please Cite as a Report of the National Research Council

At the workshop, the experts independently identified the knowledge and skills likely to be measured by each item and attempted to match these to the content and skill outlines for the VNT. They also appraised item quality and identified ambiguities that might lead students to invalid responses (correct or incorrect). After the item rating exercise, the principal investigators and staff met jointly with the experts, NAGB, and the developers of the VNT to discuss issues of item quality and coverage and to discuss plans for item review and quality assurance. The principal investigators and staff also met separately with the experts for further discussion of the item materials.

Findings, Conclusions, and Recommendations

Our findings, conclusions, and recommendations are based on our own evaluation of the information provided at the workshop, subsequent discussions, and the process and products of item development. Although we benefited greatly from the views of the experts with whom we worked, we stress that this report is solely the responsibility of the authors and the NRC. We also reiterate that these findings apply only to the materials available to us by early June.

- *The review plans for VNT items are appropriate and extend beyond procedures typically employed in test development.*

The plans for content review, student try-outs of items, and bias and sensitivity review appear rigorous and thorough. The plans for student try-outs, in particular, go well beyond item review procedures found in most test development programs. These tryouts include extensive probes to determine the validity of the students' scored responses and to identify problems that may lead to correct answers when students do not have the targeted knowledge or skill or lead to incorrect answers when they do.

- *The draft items we examined were at an early stage of development, and many of them need improvement.*

We and our experts found items with ambiguities and other problems of construction. In the case of reading items, for example, there were items for which there was not a clearly correct response, some with two possibly correct responses, and some with distracter options that might signal the correct answer. In addition, some items could possibly be answered without reading the associated passages, and others appeared to ask students to use supporting information for their response that was not in the text. Expert panelists flagged roughly half of the items available for our examination as requiring further review and possible revision.

It is critical to keep this finding in proper perspective. First, it is common to find problems with a significant number of items early in the item development process: why else conduct rigorous item reviews if not to weed out items that do not pass muster? The VNT developers, in fact, expect that 15 percent of the items will be eliminated from consideration even before pilot testing and also expect that only one-third or one-fourth of the piloted items will be used in the initial forms to be field tested for possible operational use. Moreover, as specific items were discussed at the workshop, the test developers who were present largely agreed with our assessment of item problems and in several cases reported that they had earlier come to the same conclusions.

EMBARGOED-ADVANCE COPY Not for Public Release Before Thursday, July 16, 1998, 12:00 p.m. EDT
Please Cite as a Report of the National Research Council

Thus, we conclude that there has not been sufficient time for the test development contractors to act on weaknesses in the test items. This conclusion about the unrefined state of the items we reviewed is by no means a final assessment of their quality. Rather, it signals that a significant amount of review and revision will be required to achieve a final set of high quality items.

- *The items examined did not appear to represent the full range of knowledge and skills targeted by the VNT content and skill outlines.*

Although there were items that represented varied content areas and many of the less complex skill areas, few of the items in the sample were likely to assess higher order thinking skills, as required by the approved test specifications. It will not be easy to revise items to cover these important parts of the skill specifications or the subareas of knowledge that might be underrepresented.

- *NAGB and its development contractor have also not yet had time to determine the extent to which the pool of items being developed will enable reporting of student performance in relation to NAEP's achievement levels, a central goal of VNT development.*

NAGB has developed specific descriptions of the skills associated with NAEP's "basic," "proficient," and "advanced" achievement levels for fourth-grade reading and eighth-grade mathematics (see the Achievement Level descriptions at Attachment 2). The validity of the achievement levels, which has been a topic of considerable discussion, depends on whether the description of each achievement level matches the skills of students classified at those levels.¹ Comparing candidate VNT test items to the achievement level descriptions is an important step to ensure coverage of each achievement level. In the items we and our experts reviewed, there appeared to be a shortage of items tapping higher order skills. We urge NAGB and its contractors to determine whether additional time and development are needed to produce enough items that test skills at the advanced achievement level.

- *The current schedule does not provide sufficient time for the provision of item review and try-out results to item authors and for the revision of item materials to ensure their accuracy, clarity, and quality.*

Because of the current schedule, a number of review activities that are more logically conducted in succession are being conducted simultaneously. These include content reviews by the prime contractor and its consultants, reviews by content experts, and the item try-outs.

¹ See:

Leigh Burstein, Daniel Koretz, Robert Linn, Brenda Sugrue, John Novak, Eva L. Baker, and Elizabeth Lewis Harris (1996). Describing performance standards: Validity of the 1992 National Assessment of Educational Progress achievement level descriptors as characterizations of mathematics performance. *Educational Assessment* 3(1):9-51.
Robert L. Linn (1998). Validating inferences from National Assessment of Educational Progress achievement-level setting. *Applied Measurement in Education* 11(1):23-47.
National Academy of Education (1996). *Quality and Utility: The 1994 Trial State Assessment in Reading*, Robert Glaser, Robert Linn, and George Bohmstedt, eds. Panel on the Evaluation of the NAEP Trial State Assessment. Stanford, CA: National Academy of Education.
U.S. General Accounting Office (1993). *Educational Achievement Standards: NAGB's Approach Yields Misleading Interpretations*. GAO/PEMD-93-12. Washington, DC: U.S. General Accounting Office.

EMBARGOED-ADVANCE COPY Not for Public Release Before Thursday, July 16, 1998, 12:00 p.m. EDT
Please Cite as a Report of the National Research Council

Furthermore, very little time is available to act on the results from each step in the review process. For the student try-out results to be of full use in further item development, reports on specific items must be summarized and generalizations applied to the larger set of items not examined in student try-outs. Yet the schedule allowed less than a week between the conclusion of the cognitive laboratory sessions and the provision of feedback to the item writers. It also provided less than a week for revision (if this, indeed, occurs) of a large number of items prior to the bias and sensitivity review planned for July 6, and only another week between the bias and sensitivity review and submission of the first wave of items to NAGB for its final review beginning July 15.

Given the large volume of items being developed, it appears unlikely that any of these steps—summarizing review and tryout results, responding to bias and sensitivity reviews, and item revision—can be adequately completed, let alone checked, within the 1-week time frame scheduled for each of them. In other testing programs with which we are familiar, such as the Armed Services Vocational Aptitude Battery, the Medical College Admission Test, the Kentucky Instructional Results Information System, and the National Assessment of Educational Progress, this review and revision process takes several months.

These findings lead to our central conclusion: *While the procedures planned for item review and revision are commendable, the current schedule for conducting review and revision appears to allow insufficient time for the full benefit of those procedures to be realized.* This conclusion leads to our central recommendation: *We urge NAGB to consider adjusting the development schedule to permit greater quality control, and we suggest that it may be possible to do so without compromising the planned date for the administration of pilot tests (spring 1999).* Specifically:

1. We recommend that NAGB consider whether the remaining time for refinement of item materials by VNT developers and for item review and approval by NAGB should be reallocated to allow more time for the developers' careful analysis of item review information and for the application of this input to the entire set of items. The period of time allocated for NAGB's review of item materials might be reduced correspondingly, to allow for full and complete attention to item revision and quality assurance by the test development contractors.

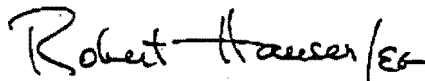
2. We recommend that NAGB and its contractors consider efforts now to match candidate VNT items to the NAEP achievement level descriptions to ensure adequate accuracy in reporting VNT results on the NAEP achievement level scale.

3. We recommend that NAGB and its contractors consider conducting a second wave of item development and review to fill in areas of the content and skill outlines and achievement level descriptions that appear to be underrepresented in the current set of items.

EMBARGOED-ADVANCE COPY Not for Public Release Before Thursday, July 16, 1998, 12:00 p.m. EDT
Please Cite as a Report of the National Research Council

We appreciate the opportunity to examine the development of the Voluntary National Tests and hope this interim letter provides formative and constructive feedback. We believe that if our recommendations are implemented, additional activity this summer and fall could improve the item review and revision process without delaying the planned pilot test in spring 1999. We would be happy to provide additional information on issues raised in this letter.

Sincerely,

Handwritten signature of Robert Hauser in black ink, followed by a forward slash and the initials 'EG'.

Robert Hauser
Co-Principal Investigator

Handwritten signature of Laureess Wise in black ink, followed by a forward slash and the initials 'EG'.

Laureess Wise
Co-Principal Investigator

c: Mr. Mark D. Musick
The Honorable William F. Goodling
The Honorable William L. Clay
The Honorable James M. Jeffords
The Honorable Edward M. Kennedy
The Honorable Arlen Specter
The Honorable Tom Harkin
The Honorable John Porter
The Honorable David Obey

EMBARGOED-ADVANCE COPY Not for Public Release Before *Thursday, July 16, 1998, 12:00 p.m. EDT*
Please Cite as a Report of the National Research Council

ATTACHMENT 1

Workshop on VNT Item Development

The Foundry Building
June 2-3, 1998

Outside Experts

Peter Afflerbach, Curriculum and Instruction, University of Maryland
Lizanne DeStefano, Bureau of Educational Research, University of Illinois
Roberta Flexer, Department of Education, University of Colorado
John Guthrie, Department of Human Development, University of Maryland
Patricia Kenney, Learning Research and Development Center, University of Pittsburgh
Marjorie Lipson, Department of Education, University of Vermont
William Tate, School of Education, University of Wisconsin-Madison

Other Participants

Rebecca Adamson, Mathtech, Inc.
Stephen Baldwin, National Research Council
Carol Benjamin, National Assessment Governing Board
Meryl Bertenthal, National Research Council
Clayton Best, American Institutes for Research
Mary Lyn Bourque, National Assessment Governing Board
Mary Crovo, National Assessment Governing Board
Marilyn Dabady, National Research Council
Larry Feinberg, National Assessment Governing Board
Michael Feuer, National Research Council
Ray Fields, National Assessment Governing Board
Robert Hauser, Co-Principal Investigator, University of Wisconsin
Cadell Hemphill, National Research Council
Mark Kutner, American Institutes for Research
Archie LaPointe, American Institutes for Research
Diane Leipzig, National Assessment Governing Board
Karen Mitchell, National Research Council
Patricia Morison, National Research Council
William Morrill, Mathtech, Inc.
John Olson, American Institutes for Research
Audrey Pendleton, U.S. Department of Education

EMBARGOED-ADVANCE COPY Not for Public Release Before *Thursday, July 16, 1998, 12:00 p.m. EDT*
Please Cite as a Report of the National Research Council

Elizabeth Rowe, American Institutes for Research

Terry Salinger, American Institutes for Research

Sharif Shakrani, National Assessment Governing Board

Gary Skaggs, National Assessment Governing Board

Roy Truby, National Assessment Governing Board

Don Wise, Mathtech, Inc.

Laurens Wise, Co-Principal Investigator, Human Resources Research Organization

EMBARGOED—ADVANCE COPY Not for Public Release Before Thursday, July 16, 1998, 12:00 p.m. EDT
Please Cite as a Report of the National Research Council

ATTACHMENT 2

Description of Reading Achievement Levels for Basic, Proficient, and Advanced Fourth Graders

Basic performance in reading should include:

- Determining what a story/informational text is about (i.e., topic, main idea)
- Determining the main purpose for reading a selection
- Identifying character(s), setting(s), conflict(s), or plot(s) in a story
- Supporting one's understanding of a story/informational text with appropriate details
- Explaining why one likes or dislikes what they have read [a reading]
- Connecting material from a story/informational text to personal experiences
- Making predictions about situations beyond the confines of the printed material
- Maintaining a focus over the entirety of a story/informational text

Proficient performance in reading should include:

- Summarizing a story/informational text
- Recognizing an author's intent or purpose
- Making simple inferences based on information provided in a story/informational text
- Drawing a valid conclusion from a story/informational text
- Determining the meaning of key concepts in the story/informational text and connecting them to the main idea
- Recognizing relationships in a story/informational text (time order, cause/effect, compare/contrast)

Advanced performance in reading should include:

- Explaining an author's intent, using supporting material from the story/informational text
- Describing the similarities and difference in characters, settings, and plots
- Demonstrating an awareness of the use of literary devices, such as figurative language
- Applying inferences drawn from a story/informational text to personal experiences
- Extending the meaning of a story/informational text by integrating experiences and information outside of the text
- Making and explaining a critical judgment of a story/informational text
- Demonstrating an ability to adapt reading purpose to a variety of printed materials and/or writing styles

EMBARGOED-ADVANCE COPY Not for Public Release Before *Thursday, July 16, 1998, 12:00 p.m. EDT*
Please Cite as a Report of the National Research Council

Description of Mathematics Achievement Levels for Basic, Proficient, and Advanced Eighth Graders

The five NAEP content areas are (1) numbers and operations, (2) measurement, (3) geometry, (4) data analysis, statistics, and probability, and (5) algebra and functions. Skills are cumulative across levels--from Basic to Proficient to Advanced.

Basic 256	Eighth-grade students performing at the basic level should exhibit evidence of conceptual and procedural understanding in the five NAEP content areas. This level of performance signifies an understanding of arithmetic operations—including estimation—on whole numbers, decimals, fractions, and percents.
------------------	---

Eighth graders performing at the basic level should complete problems correctly with help of structural prompts such as diagrams, charts, and graphs. They should be able to solve problems in all NAEP content areas through the appropriate selection and use of strategies and technological tools--including calculators, computers, and geometric shapes. Students at this level also should be able to use fundamental algebraic and informal geometric concepts in problem solving.

As they approach the proficient level, students at the basic level should be able to determine which of the available data are necessary and sufficient for correct solutions and use them in problem solving. However, these 8th graders show limited skill in communicating mathematically.

Proficient 294	Eighth-grade students performing at the proficient level should apply mathematical concepts and procedures consistently to complex problems in the five NAEP content areas.
-----------------------	--

Eighth graders performing at the proficient level should be able to conjecture, defend their ideas, and give supporting examples. They should understand the connections between fractions, percents, decimals, and other mathematical topics such as algebra and functions. Students at this level are expected to have a thorough understanding of basic level arithmetic operations—an understanding sufficient for problem solving in practical situations.

Quantity and spatial relationships in problem solving and reasoning should be familiar to them, and they should be able to convey underlying reasoning skills beyond the level of arithmetic. They should be able to compare and contrast mathematical ideas and generate their own examples. These students should make inferences from data and graphs; apply properties of informal geometry; and accurately use the tools of technology. Students at this level should understand the process of gathering and organizing data and be able to calculate, evaluate, and communicate results within the domain of statistics and probability.

Advanced 331	Eighth-grade students performing at the advanced level should be able to reach beyond recognition, identification, and application of mathematical rules in order to generalize and synthesize concepts in the five NAEP content areas.
---------------------	--

Eighth graders performing at the advanced level should be able to probe examples and counterexamples in order to shape generalizations from which they can develop models. Eighth graders performing at the advanced level should use number sense and geometric awareness to consider the reasonableness of an answer. They are expected to use abstract thinking to create unique problem-solving techniques and explain the reasoning processes underlying their conclusions.

EDUCATION DAILY

The education community's independent daily news service

NRC Letter, Goodling Spark National Testing Dispute

The House's leading education lawmaker is distorting findings about the validity and racial impact of voluntary national tests in reading and math, researchers charged yesterday.

Rep. Bill Goodling, R-Pa., issued a statement yesterday declaring that voluntary national tests—as proposed by President Clinton last year—“have the potential to be invalid, unreliable and racially biased” because the National Assessment Governing Board (NAGB) is developing individual test items too quickly.

Goodling, chairman of the House Education and the Workforce Committee, attributes this finding to the National Research Council (NRC), which is studying the proposed exams for Congress and the administration (ED, Nov. 13, 1997).

Missing The Mark

NRC researchers, who reported on the issue in a July 16 letter to Education Secretary Richard Riley, told *Education Daily* yesterday that Goodling missed their point.

“There is absolutely nothing in our report upon which to base the conclusion that the tests have the potential to be invalid, unreliable and racially biased,” said Michael Feuer, NRC's director on testing and assessments.

“The report says, first of all, that the review plans for the test are appropriate, and actually extend beyond procedures that are typically employed in test development,” he added.

In fact, Feuer said the “principal message” in NRC's letter to Riley is that a modest change in the review schedule between now and November—when NAGB wants to approve pilot exams—would allow “the full benefit of the review and revision activities to be realized.”

“We think that a small shift in the schedule could go a long way towards getting more out
(more on p. 2)

Vol. 31, No. 135 • Thursday, July 16, 1998

House Would Block Viagra
To Cover Kids' Mental Health Page 2

Children's Welfare Improves,
Interagency Report Says Page 4

Ga. Teacher Prep Initiative
Raises Bar For Requirements Page 5

Regional Education Labs Miss The Mark, Report Says

If lawmakers are serious about cutting bloat in the federal government, they should look no further than the Education Department's 10 regional education laboratories, a scathing new report from a conservative think tank says.

The regional labs—tasked to research effective teaching techniques and disseminate successful research and practices—have failed to perform their mission and should be overhauled by Congress, said Nina Shokrafi, education policy analyst for the Washington, D.C.-based Heritage Foundation.

The labs “do little if anything to promote proven models of teaching that lead to higher academic outcomes,” Shokrafi charged. Instead, they issue shoddy research, lack objectivity, promote fads and are rife with waste and bureaucracy, she said.

But Bob Stonehill, who directs the lab program for ED's Office of Educational Research and Improvement, countered that the labs are cost-efficient and indeed make a difference through effective research and outreach.

(more on p. 3)

30
Years of Excellence
EDUCATION
DAILY

NRC Letter, Goodling Spark National Testing Dispute (Cont.)

of this process, without compromising the November deadline," he said.

NRC researchers stressed that they're not taking a position on whether Congress should approve national tests; instead, they recommend an improved timeline for test development.

Bob Hauser, research professor of sociology at the University of Wisconsin, one of two co-principal NRC investigators, said "NAGB should consider taking more time to digest all the [information] ... in order to ensure that it gets used effectively."

NRC reported last month that existing state and local tests could not validly be compared to each other (ED, June 15).

In his statement, Goodling also said it would take four to five years to develop valid, racially fair tests.

While that's "a fair estimate," Feuer noted that such a long timeline includes lengthy quality reviews and field testing, so NAGB is not departing wildly from a regular schedule because

it's not starting from scratch. Instead, it is using test frameworks and items from the National Assessment of Educational Progress (NAEP), which "gives the developers of the VNT a tremendous head start ... It's not as if they're trying to do in one year what is normally done in five years," Feuer said. "That's a fairly, strong exaggeration of the facts."

Jay Diskay, a spokesman for the House education panel, defended Goodling's statement as a fair interpretation of NRC's message to Riley.

"From our point of view, they're claiming that there are some real technical problems with this," Diskay said. "We have been told over and over ... that it takes four to five years [to develop] what is valid and free of racial bias."

He added that Goodling opposes national tests even if they are valid and unbiased.

NAGB yesterday called the report "helpful" and said it would act on NRC's recommendations, providing two more months for test item development while still aiming to ready pilot tests for March 1999. —William J. Cahir

House Would Block Viagra To Cover Kids' Mental Health

The House spending committee voted along party lines Tuesday for its fiscal 1999 education funding bill, embracing a half-hearted plan to provide mental health coverage to students with behavioral problems.

Before passing its \$30.55 billion fiscal 1999 funding bill for the Education Department, 32-23, the House Appropriations Committee approved an amendment by Rep. David Obey, R-Wis., to ban federal payments for the anti-impotence drug Viagra.

The measure would direct any savings—up to \$100 million—to school-based mental health services for children at risk of violence.

Obey explained that he wants to prevent school violence. But it's unclear if Obey's approach would direct any funds to schools.

The Health Care Financing Administration, which runs Medicaid jointly with states, lets state plans cover Viagra when the drug is "medically necessary." Obey's plan may redirect little or no funding to schools.

House appropriators also approved an amendment to delete \$89.7 million from the increase in funding for Pell Grants, cutting the boost in student financial aid to \$693.7 million (ED, July 15). President Clinton has threatened to veto the bill. —William J. Cahir



Aspen Publishers, Inc.

Published every business day by Aspen Publishers, Inc. ISSN: 0013-1261, P.O. Box 1452, Alexandria, VA 22303-2052; Editorial: (703)683-4100; Subscriber Services: (800)655-3597; Publisher: Michael Brown, Executive Editor; Annette Lickona, Managing Editor; Dave Harrison, Deputy Managing Editor; Kim Sanderson, Editorial Staff; William J. Cahir, Gina Farrell, Jonathan Fox, Darren Goode, Bob Jennings, Eli J. Lata, Rebecca S. Wilson, Pam Moore, Deborah Givens. Annual subscription price: \$595. Six-month subscription rate: \$302. Copyright 1998 by Aspen Publishers, Inc. Federal copyright law prohibits reproduction of material in this newsletter without written permission. For reprint permission please contact Permissions Dept., Aspen Publishers, Inc., (703)739-6444, fax (703)739-6348 or e-mail: customercare@aspenpub.com. For editorial questions or suggestions, e-mail: editor@aspenpub.com or letterstoeditor@aspenpub.com.

For reprint permission please contact Permissions Dept., Aspen Publishers, Inc., (703)739-6444, fax (703)739-6348 or e-mail: customercare@aspenpub.com. For editorial questions or suggestions, e-mail: editor@aspenpub.com or letterstoeditor@aspenpub.com.



**COMMITTEE ON EDUCATION & THE WORKFORCE
U.S. HOUSE OF REPRESENTATIVES**

**FOR IMMEDIATE RELEASE
July 15, 1998**

**CONTACT: Jay Diskey or
Bill McCarthy
202/225-4527**

**STATEMENT OF CHAIRMAN BILL GOODLING (R-PA)
ON NATIONAL RESEARCH COUNCIL REPORT ON
DEVELOPMENT OF THE VOLUNTARY NATIONAL TESTS**

Goodling: "The tests have the potential to be invalid, unreliable & racially biased."

"This week Congress learned that the Clinton Administration's national tests are technically unsound. According to a National Research Council report to Secretary of Education Richard Riley, development of the tests is moving too quickly and because of it, the tests have the potential to be invalid, unreliable and racially biased.

"The National Research Council's findings aren't surprising. Throughout last year's debate on the President's federal testing plan we said the Administration was moving too quickly in trying to develop tests in a year. After all, the nation's leading test publishers and education researchers say that it takes four to five years to develop sound tests. What makes the Administration and the National Assessment Governing Board think that sound tests can be developed in a year?

"It is time for the Clinton Administration to give up its national tests. This proposal has nothing going for it. Only seven states have accepted it. On two occasions the U.S. House of Representatives has acted in a bipartisan fashion to stop it and the U.S. Senate voted once to prohibit it. Now, the National Research Council says its time to slow down. I say its time to stop for good. The American taxpayer should not have to spend another penny for this nonsense. Congress has no plans to authorize new national tests or appropriate money for them. More testing is no solution. Instead, let's focus on teacher training, quality in the classroom, parental involvement, basic academics and dollars directly to the classroom - not bureaucracy."

###



**COMMITTEE ON EDUCATION & THE WORKFORCE
U.S. HOUSE OF REPRESENTATIVES**

**FOR IMMEDIATE RELEASE
July 15, 1998**

**CONTACT: Jay Disky or
Bill McCarthy
202/225-4527**

**STATEMENT OF CHAIRMAN BILL GOODLING (R-PA)
ON NATIONAL RESEARCH COUNCIL REPORT ON
DEVELOPMENT OF THE VOLUNTARY NATIONAL TESTS**

Goodling: "The tests have the potential to be invalid, unreliable & racially biased."

"This week Congress learned that the Clinton Administration's national tests are technically unsound. According to a National Research Council report to Secretary of Education Richard Riley, development of the tests is moving too quickly and because of it, the tests have the potential to be invalid, unreliable and racially biased.

"The National Research Council's findings aren't surprising. Throughout last year's debate on the President's federal testing plan we said the Administration was moving too quickly in trying to develop tests in a year. After all, the nation's leading test publishers and education researchers say that it takes four to five years to develop sound tests. What makes the Administration and the National Assessment Governing Board think that sound tests can be developed in a year?

"It is time for the Clinton Administration to give up its national tests. This proposal has nothing going for it. Only seven states have accepted it. On two occasions the U.S. House of Representatives has acted in a bipartisan fashion to stop it and the U.S. Senate voted once to prohibit it. Now, the National Research Council says its time to slow down. I say its time to stop for good. The American taxpayer should not have to spend another penny for this nonsense. Congress has no plans to authorize new national tests or appropriate money for them. More testing is no solution. Instead, let's focus on teacher training, quality in the classroom, parental involvement, basic academics and dollars directly to the classroom - not bureaucracy."

###



National Assessment Governing Board

National Assessment of Educational Progress

HOLD FOR RELEASE
July 16, 1998, 12 Noon EDT

Contact: Lawrence Feinberg
(202) 357-6942

STATEMENT BY MARK D. MUSICK, CHAIRMAN NATIONAL ASSESSMENT GOVERNING BOARD

Comments on National Academy of Sciences Interim Report on Item Development for Proposed Voluntary National Tests

The National Academy of Sciences has issued a very helpful interim report on the work of the Governing Board and its contractors in developing questions for the proposed Voluntary National Tests in 4th grade reading and 8th grade mathematics. We intend to act on the recommendations, as the Board continues to carry out the responsibilities, assigned to it by law, to develop the proposed exams.

We agree with the Academy that the highest priority must be placed on developing good questions that measure the full, rich content of the National Assessment of Educational Progress. As the report says, the pool of items must also insure that the tests can reliably report whether individual students meet the NAEP standards for basic, proficient, or advanced performance.

The Board staff is working now with American Institutes for Research, our test contractor, to provide more time for item development while still meeting the Board-approved schedule for proposed pilot testing in March 1999. The Board has already increased the planning and development time for the proposed exams by one year since it was given responsibility for this project.

The NAGB staff and contractor are also making arrangements to convene review panels to make sure that both the NAEP content frameworks and achievement levels are well covered by the available test questions. If more questions are needed in any particular category or level of difficulty, more will be written.

Altogether, about 2,600 test questions will be prepared for the pilot-testing. As the Academy notes, its investigators and outside experts reviewed just 105 items at their workshop, which was held six weeks ago. This was at an early point in item development. Since then, many more questions have been written; many changes have been made in the ones that the Academy saw.

-MORE-

800 North Capitol Street, N.W.
Suite 825, Mailstop 7583
Washington, D.C. 20002-4233
Phone: (202) 357-6938
Fax: (202) 357-6945

-2-

We are pleased that the Academy considers our plans for the review of VNT items to be "rigorous and thorough." By stretching out item development an additional two months, there will be more time for the contractor to make changes that are consistent with those reviews.

As part of this "rigorous and thorough" process, members of the Board will review and approve each reading and math item before it is included in the pilot tests. Based on the pilot test results, even more improvements will be made before the exams become available for operational use in March 2001, if the program is authorized by law.

In carrying out the job of developing the proposed Voluntary National Tests, NAGB is committed to preparing exams that are of the highest technical quality. The contractor and subcontractors who are writing the questions and assembling the exams have an enormous amount of talent and experience. The procedures for item review and quality control are extensive.

The independent evaluation by the National Academy of Sciences is helpful, particularly since we have received the interim report at a point when changes can be made. We welcome the recommendation of the National Academy report that the schedule for item development be adjusted to give more time for the best possible work to be done.

#



DRAFT

For Release:
June 12, 1998

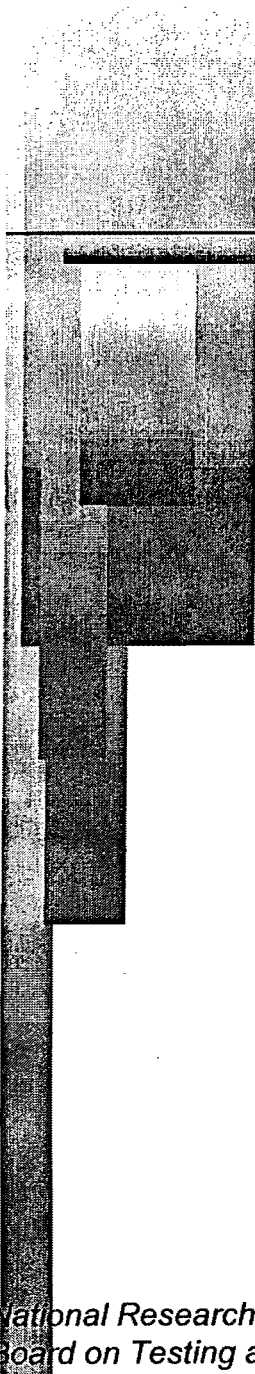
Contact: David Frank
(202) 401-3026

**Statement by U.S. Secretary of Education Richard W. Riley
on Voluntary National Testing**

The National Academy of Sciences study's conclusions clearly show that work on developing voluntary national tests must continue. Last fall the President and Congress agreed to have this study conducted, and agreed that the results should inform Congressional action on national testing. Now we have the results, and they show conclusively that if we want to know how well students are performing compared to rigorous and widely accepted national standards, then we simply cannot rely on existing tests. Work on the voluntary national tests in 4th grade reading and 8th grade math is proceeding under the authority of the independent, bipartisan National Assessment Governing Board (NAGB).

We must do everything we can to help our children master the basics all across America. These tests will help provide the needed information to accomplish that. It is clear that NAGB's work must continue without delay.

###



Equivalency and Linkage of Educational Tests

*Interim Report
Briefing Notes
June 10-12, 1998*

Committee on Equivalency and Linkage of Educational Tests

- *Paul W. Holland (chair): UC, Berkeley. Statistics, psychometrics, testing*
- *Robert C. Calfee: Stanford. Educational psychology, assessment*
- *John T. Guthrie: U. Maryland. Education (reading)*
- *Richard M. Jaeger: UNC/Greensboro. Educational measurement*
- *Patricia Ann Kenney: U. Pittsburgh. Education (mathematics)*
- *Vonda L. Kiplinger: Colorado Dept of Education. Assessment*
- *Daniel M. Koretz: RAND. Assessment*
- *Frederick C. Mosteller (NAS): Harvard. Statistics*
- *Peter J. Pashley: LSAC. Psychometrics*
- *Doris Redfield: Richmond. Educational consultant*
- *William F. Tate: UW/Madison. Mathematics education*
- *David Thissen: UNC, Chapel Hill. Psychometrics*
- *Ewart A.C. Thomas: Stanford. Psychology and statistics*
- *Laurens L. Wise: HUMRRO. Psychometrics*
- *Robert L. Linn (ex officio): U. Colorado. Educational measurement*

Staff

- *Michael J. Feuer, Ph.D., Study Director*
- *Meryl Bertenthal, Senior Research Associate*
- *F. Cadelle Hemphill, Senior Research Associate*
- *Bert F. Green, Ph.D., Senior Technical Consultant*

- *Lisa Alston, Administrative Assistant*
- *Nancy Kober, Editorial Consultant*

Committee Charge

Public Law 105-78:

National Academy of Sciences to

“conduct a feasibility study to determine if an equivalency scale can be developed that would allow test scores from commercially available standardized tests and state assessments to be compared to each other and the National Assessment of Educational Progress...”

Conclusions (1)

- *comparing the full array of currently administered commercial and state achievement tests to one another, through the development of a single equivalency or linking scale, is not feasible;*

Conclusions (2)

- *reporting individual student scores from the full array of state and commercial achievement tests on the NAEP scale, and transforming individual scores on these various tests and assessments into the NAEP achievement levels, is not feasible.*

Context

- *Voluntary National Testing*
- *State and local educational decision-making*
- *Prior investments in assessment development*
- *Recurrent interest in “linkage”: unity and diversity*
- *Science of linkage*

Definitions

■ *Feasibility*

- *practicality*
- *validity*

■ *Linking and equating*

Key Assumptions and Caveat

- *The development of a method to link and compare existing tests is seen as a possible substitute for the development of new national tests.*
- *We focus on “the full array” of commercial and state tests*
- *We take no position on the VNTs*

Findings

- *diversity:*
 - *content emphasis, format, reliability, administration, uses*
- *dynamic testing landscape:*
 - *state and local goals*
 - *NAEP*
- *effects on behavior of teachers and kids*
 - *high stakes and low stakes*
- *population differences*
- *NAEP achievement levels*

Still to come...

- *issues involved in linking subsets of tests*
 - *technical requirements*
 - *policy implications*
 - *research agenda*

- *further elaboration of*
 - *testing landscape in the US*
 - *statistics of linkage*
 - *prior studies*

Clinton Presidential Records Digital Records Marker

This is not a presidential record. This is used as an administrative marker by the William J. Clinton Presidential Library Staff.

This marker identifies the place of a publication.

Publications have not been scanned in their entirety for the purpose of digitization. To see the full publication please search online or visit the Clinton Presidential Library's Research Room.

PREPUBLICATION COPY
(Uncorrected Proofs)

EQUIVALENCY AND LINKAGE OF EDUCATIONAL TESTS

Interim Report

ADVANCE COPY

Not For Public Release Before
Friday, June 12, 1998
12:00 p.m. EDT

PLEASE CITE AS A REPORT OF THE
NATIONAL RESEARCH COUNCIL

NATIONAL RESEARCH COUNCIL

CLINTON LIBRARY PHOTOCOPY