

NAEP 2011 Writing Assessment:
Issues in Developing a Framework
and Specifications

Arthur N. Applebee
University at Albany, SUNY

December 2005

Paper commissioned by the National Assessment Governing Board

Table of Contents

	Page
Introduction	3
Issue 1: What types of writing should be assessed, and how are they related to one another?	3
Issue 2: What writing tasks/ types should be assessed at each grade level?	5
Issue 3: How can the 12 th grade assessment be structured to measure preparedness for post-secondary endeavors, including college, workplace training, and entrance into the military?	5
Issue 4: Should the writing assessment be computerized?	7
Issue 5: What aspects of writing achievement should be measured?	9
Issue 6: What should students write about?	10
Issue 7: How should the Framework address the question of time?	11
Issue 8: What accommodations should be made for English language learners, students with disabilities, and low-achieving students?	13
Conclusion	14
References	14
Appendix: The Evolution of the NAEP Writing Framework	16
Framework References	18

Introduction

During 2006-2007, broadly representative committees will have the opportunity to revisit the Writing Framework for the 2011 National Assessment of Educational Progress (NAEP), which in its current version dates to 1989-90 with revisions, primarily to test specifications, in 1995-96 (National Assessment Governing Board, 2002). Much of the substance of the Framework, including its emphasis on three major purposes for writing and on the writing process, goes back even further, to the objectives for the 1983-84 assessment (NAEP, 1982). (See appendix for a summary of the NAEP objectives from 1969-2007.)

This paper is designed as a discussion-starter, attempting to frame a number of issues and debates in writing assessment that could be constructively revisited by the Framework Committees. Underlying all of these issues is a larger one: What information about how students in the United States write should NAEP provide to interested members of the general public, to policymakers, and to educators? Although a seemingly simple question, buried within it are a variety of difficult issues on which there is currently little consensus, including how to describe the domain of writing tasks; the relationships among component skills, content knowledge, and generalized writing “fluency”; and the relevance of computer-based applications to definitions of writing achievement as well as to assessment techniques.

Issue 1: What types of writing should be assessed, and how are they related to one another?

Recent research in writing has tended to emphasize the extent to which writing genres are socially situated and context specific. This is true whether one begins with Miller’s (1984) emphasis on genre as social action, or the systemic linguistics approach of the Australian genre theorists (Halliday & Martin, 1993; Cope & Kalantzis, 1993). These perspectives pose a challenge to the traditional emphasis on writing as a generic skill, taught primarily in English language arts or composition classes, and assessable through generic writing tasks detached from particular disciplinary or socially-constituted contexts. They suggest that what counts as effective argument and persuasive evidence varies greatly in moving from one context to another, so that what counts as “good writing” is itself socially constructed and context specific. As Halliday and Martin (1993) demonstrate, for example, science writing has many features such as reliance on technical vocabulary, use of the passive voice, and nominalization (use of verbs and adjectives as nouns) that English teachers would ordinarily find objectionable—though these features have evolved in science writing to serve particular communicative needs.

The current NAEP framework derives from the work of Kinneavy (1980), Britton (et al., 1975), and Moffett (1968) during the 1960s and 1970s, in interaction with perceptions of typical practice and school-based terminology for discussion of writing instruction. The domain of NAEP writing tasks is divided into three broad purposes for writing—

informative, persuasive, and narrative. The current NAEP framework encourages writing within each of these purposes involving a “variety of tasks” and “many different” audiences, triggered by a “variety” of stimulus materials. There is no consensus in theory or practice, however, about the proper way to partition the domain of writing tasks, and there has always been a perception of overlap among the categories: Doesn’t an author of an “informative” text implicitly intend to persuade a reader of the truth or accuracy of what is being said? Isn’t narrative an important technique for both informing and persuading? (“Narrative” has itself evolved out of concerns in earlier versions of the assessment with “personal,” “imaginative,” or “expressive” writing, in an attempt to capture the genres of literature as well as of personal reflection.)

The problems in terminology extend to state writing assessments, which have often turned to NAEP as a starting point in designing their own assessments. Texas, for example, requires writing for “various audiences and purposes,” in a variety of forms, including “business, personal, literary, and persuasive texts.” California instead treats these generalized purposes as part of “writing strategies,” and specifies a variety of specific genres to be assessed (for example, at grade 11, fictional, autobiographical or biographical narrative; responses to literature; reflective compositions; historical investigation reports; and job applications and resumes.)

There are other alternatives. College entrance exams from the College Board and ACT both assume that good writing is a generic skill, at least in academic contexts; the College Board, for example, advises that high scores will go to “essays that insightfully develop a point of view with appropriate reasons and examples and use language skillfully” (http://www.collegeboard.com/student/testing/sat/prep_one/sat_essay.html; retrieved 11/25, 2005).

Martin & Rothery (1980), from the Australian genre-theory perspective, point in another direction, with a list of schooled nonfiction genres: recount, report, procedure, explanation, persuasion, and discussion. Their listing, like others from the Australian group, introduces terminology unfamiliar to American readers, and also collapses their original insights about the situated nature of genre knowledge into a generic set of “school” genres that are not all that distant from Britton et al.’s (1975) and Moffett’s (1968) subcategories of informational or expository writing.

Considerations

Lacking a major recent theoretical formulation that is widely accepted as a way to resolve these problems in definition and categorization, the Framework Committees will need to consider how to provide a domain description that will make sense to teachers and the general public while also providing clear guidelines for prompt development.

Issue 2: What writing tasks/ types should be assessed at each grade level?

Tangled with the problem of specifying the domain of writing tasks is the distribution of tasks across grade levels. The current framework assumes that each of the broad purposes for writing is appropriate even for primary grade writers, with development taking the form of the ability to complete ever-more sophisticated or specialized tasks within those purposes. While informative writing tasks have been relatively uncontroversial across the grades, arguments have been raised against assessment of persuasive writing at the fourth grade level, and narrative (particularly story) writing at grade 12. At fourth grade, the arguments have been that persuasive writing is

- a) too difficult,
- b) developmentally inappropriate, or
- c) out of step with the curriculum.

At grade 12, the arguments have been that story writing is

- a) too easy,
- b) no longer relevant to the curriculum of most students, or
- c) not consistent with the types of writing expected in college and the workplace.

The current framework addresses this issue by placing more emphasis on persuasive writing in grade 12, and more on narrative writing in grade 4.

NAEP itself offers some evidence on these arguments, in that achievement has been somewhat higher on narrative tasks and somewhat lower on persuasive ones. There has been a narrowing of the range of task difficulty over time however; early assessments showed much greater between-task variation than is presently evident. This is the result of pilot-testing and task selection procedures that have eliminated tasks that were very easy or very hard at a given grade level. In fact the current Framework cautions against items that are either too hard or too difficult (NAEP, 2002). One result of this has been that it is no longer possible to comment on tasks that lower-achieving students can complete successfully, since these tasks are no longer included in the assessment.

Considerations

The Framework Committees should consider whether this narrowing of the range of task difficulty has been beneficial, or should be reconsidered. They will also need to consider how the types of tasks (or the proportionate emphasis on tasks of different types) should vary across the three grade levels in the assessment (4th, 8th, and 12th).

Issue 3: How can the 12th grade assessment be structured to measure preparedness for post-secondary endeavors, including college, workplace training, and entrance into the military?

In 2003 the National Assessment Governing Board established the National Commission on NAEP 12th Grade Assessment and Reporting to review the 12th grade NAEP assessment and to recommend improvements to NAGB. The Commission's report (2004) noted that the high school diploma is no longer a culminating degree for most

students; 88% of 8th graders report wanting to continue into higher education, and 70% of high school graduates actually do so within two years of graduation (p. 6). At the same time, 45-55% of entering freshmen are unprepared for college work, as reflected in placements in remedial coursework during their first year in college.

Lacking any other national standard for measuring preparedness, the Commission recommended that new NAEP frameworks for the 12th grade be oriented toward assessing preparedness for the challenges of college, workplace training, and the military. At the same time, the Commission noted that there is little consensus on what “preparedness” means, and that validating measures of preparedness is likely to require extensive follow up studies exploring how students at various achievement levels do in various post-high school contexts. The National Assessment Governing Board’s Assessment Development Committee has endorsed this emphasis on 12th grade preparedness, while noting that the issue is complex and the message that NAEP will send in this regard is very important.

The history of attempts to shape curriculum and assessment around preparedness for future life or work is not a happy one (Applebee, 1974). Past attempts to inventory necessary skills have tended to converge on simple skills that are easy to itemize (spelling, punctuation) rather than higher-level skills (e.g., thoughtful argument and use of evidence) that virtually everyone cites as essential goals of education. The result was usually a system of curriculum and assessment that focused on basic skills or on generic workplace tasks (e.g., business letter format) that easily degenerated into formulas with little real-world relevance.

The most extensive recent effort to relate high school achievement to preparedness both for college study and for the workplace is the American Diploma Project (2004). Drawing on studies of the skills needed in high-performance, high-growth jobs, as well as of the requirements for college level tasks, the American Diploma Project report emphasizes such higher-level skills as expressing ideas clearly and persuasively, and producing high quality writing resulting from careful planning, drafting, and meaningful revision (pp. 28-29). The report also includes extensive benchmarks meant to indicate the level of achievement appropriate for high school graduation. The 10 benchmarks for writing range cover a wide range, from planning, drafting, and revising; to selecting language appropriate for purpose, audience, and context; to writing well-structured academic essays and work-related texts; to using appropriate software programs. Benchmarks under other headings also refer to writing tasks, however, including benchmarks labeled as research, logic, informational text, media, and literature. Although the overall emphasis remains on higher-level accomplishments, the benchmarks show some of the problems of earlier attempts, with appropriate citation of print or electronic sources emerging as a benchmark at the same level of importance as writing an academic essay.

Considerations

The Writing Framework Committees will need to consider what modifications may be necessary to the assessment framework for grade 12 if it is to function effectively as a measure of preparedness for college, the workplace, and the military. Would this be best accomplished by introducing new kinds of tasks? By changing the balance among tasks included at grades 4 and 8? Or by changing the emphasis in the scoring system to provide more information on either more advanced skills (organization, argument, citing of evidence) or more basic skills (spelling, punctuation)?

Issue 4: Should the writing assessment be computerized?

Computer use is becoming widespread in American schools, and by the 2011 assessment it should be even more so. In 2003, for example, virtually all schools reported having computers with internet access, with no differences among schools serving demographically different populations. Student access to such computers for instructional use has also been increasing rapidly; there were 4.4 students per computer with internet access in 2003, compared with 12.3 in 1998 (<http://nces.ed.gov/surveys/frss/publications/2005015/2.asp>).

For writing instruction, the most important computer-based tool has been the word processor. Like the calculator in mathematics, word processing transforms the writing task, simplifying editing and revision and providing embedded tools for spelling and grammar checking. Although most assessments are still paper-and-pencil, computer-based assessment that allows the use of word processing is becoming more widespread. When the Test of English as a Foreign Language (TOEFL) exam was recently revised, for example, it was moved to an internet-based format that assesses reading, writing, and spoken language skills; and the Canadian province of Alberta has for a number of years made provision for optional use of word processors for Diploma exams in English and other subjects (Russell & Platt, 2002; http://www.education.gov.ab.ca/k_12/testing/diploma/dip_gib/examinationprogram.asp).

Computer-based writing assessment nonetheless raises some difficult issues of equity and access. Writing produced on a computer tends to be longer than writing produced by hand, and longer writing tends to be more highly evaluated than shorter selections, perhaps because of the inclusion of more evidence or elaboration (Bereiter & Scardamalia, 1987). The bias arguments run in both directions: Not having access to a computer may penalize those who are used to writing on a computer in school and at home; on the other hand, those who are not used to writing on a computer will either be handicapped by poor keyboarding skills, or if they compose by hand by the greater length of essays produced by their computer-using peers.

The research base on the effects of word processors on assessment results is slim and not particularly convincing; arguments that paper-and-pencil tests underestimate achievement

of students who are used to writing on word processors treat writing as though it were being evaluated against an external, fixed standard (e.g., Russell & Plati, 2002), when in fact writing rubrics reflect the circumstances of production. Rather than an overall increase in performance, a switch to a computerized assessment including word processing software is more likely to lead to changes in the benchmarks at each level in the scoring rubric to reflect the advantages accrued from the new format.

The most extensive study of the effects of computerizing a writing assessment is NAEP's 2002 study of writing online (Sandene et al., 2005). This special study compared performance on two NAEP writing tasks (one informative and one persuasive) at the 8th grade level, when given as part of the regular paper-and-pencil assessment or given in a special web- or laptop-based format that also included simple word processing tools. The detailed results show a number of topic-specific differences in performance across formats, but are generally encouraging. There were no equity-related differences in essay quality, though there was a 1% higher response rate for the paper-and-pencil version of one task. Males also wrote significantly longer responses on computer than on the paper-and-pencil version of one task, but their essays were not rated significantly higher.

Students with more hands-on computer skill (as measured by typing speed, error rate, and ability to use word processing tools) did better on both of the computer-based writing tasks; the correlation between their overall writing score and the measure of computer skill was .42; even after adjusting for paper-and-pencil writing achievement, computer skill still accounted for about 11% of the variation in computer-based measures of writing achievement. The "hands-on" computer familiarity measure, however, has a significant literacy component that may account for much of his relationship. Other measures of computer experience, including frequency of completing various kinds of writing assignments on a computer, were unrelated to computer-based writing achievement.

Overall, the authors of the NAEP writing online study conclude that aggregated scores from online assessment do not differ significantly from paper-and-pencil results, although results for individual students may do so.

Although school-level data have recently suggested that equity issues in computer access have been reduced, at the student level issues of access have not been completely resolved. In 2003, for example, there were fewer computers with internet access available in schools serving high proportions of minority students than in schools with the lowest proportion of minority students (5.1 students per computer versus 4.1 per computer). Data from the 2002 writing assessment suggest an even larger divide: Some 29% of White students reported using a computer for writing "A lot," compared with only 19% of Black and 18% of Hispanic students (NAEP Data Explorer, 2002 Writing Assessment).

Considerations

The Writing Framework Committees will need to consider how advances in computer use and availability are impacting writing instruction, and what this means for definitions of

what it means to write well. If equity issues can be resolved, a computer-base assessment has a number of advantages in measuring writing achievement and in providing accommodations to students who need them (Issues 5 and 8, below).

Issue 5: What aspects of writing achievement should be measured?

Just as there is no widely agreed definition of the domain of writing tasks, there are many competing approaches to measuring the various interrelated components of writing achievement. Over time, the primary rubric used to measure writing achievement in NAEP has evolved from a holistic rating to a prompt-specific primary trait rating (Lloyd-Jones, 1977) to the current set of purpose-related rubrics (one for each of the three purposes for writing) that can be seen as either generalized primary trait or focused holistic. Although NAEP reports have been organized around separate sections discussing informative, persuasive, and narrative writing, reporting has either remained at the level of individual writing prompts, or has been aggregated to a total writing score. There have been no separate subscales for types of writing in published reports or in the data available online (NAEP Data Explorer).

Other scoring systems have attempted to provide separate ratings for different features of a writing sample. The most widely used today is probably the 6-trait (or 6+1 trait) system disseminated by Northwest Regional Laboratory. This provides separate scores for ideas, organization, voice, word choice, sentence fluency, conventions, and (optionally) presentation. This system emerged out of the work of Paul Diederich and his colleagues at ETS (Diederich, French, & Carleton, 1961), and can be a useful tool in reminding teachers and students of the many dimensions of effective writing. As a measurement tool, however, it is not clear that the profiles that result yield psychometrically useful information. Diederich's (1974) suggestion was to use the traits for socializing raters to a common standard, and then to drop the traits and focus on total scores.

But there have been many attempts to measure other aspects of writing achievement, including syntactic complexity, ability to edit and revise, mastery of writing conventions (punctuation, capitalization, usage, spelling), organizational ability, and vocabulary level. Such features are arguably of interest in understanding writing achievement, but they have usually required time-consuming scoring procedures and been complicated by the fact that the results are task and content specific. Syntactic complexity is typically greater for an analytic or persuasive task than for a narrative task, for example, reflecting the typically embedded nature of clauses in argumentative discourse. Error rates in writing conventions similarly vary with task—with errors tending to increase as tasks become more difficult, presumably as the result of the deflection of cognitive and linguistic resources from one aspect of the task to another.

Many measures of interest that are tedious to derive by hand are very easy to derive by computer. There are now a range of text analytic software programs available that will report features such as number of words, variety in word choice, syntactic complexity, vocabulary level, and error rates. Many also calculate an overall quality score. If the 2011

writing assessment is computer based, it would allow the assessment of aspects of writing development that can currently only be examined in special studies on limited subsamples of papers.

There is of course another psychometrically efficient option to obtain measures of some of these features. Knowledge of written language conventions and vocabulary level, for example, can be tested quite efficiently in multiple-choice formats. Such measures are highly reliable and have good predictive validity (Godshalk, Swineford, & Coffman, 1966; Breland et al., 1987); however, they have long been resisted by the community of writing educators because of their impact on curriculum and instruction. Such short-answer formats divert the focus of instruction away from student experiences with writing extended text.

Thus another benefit of computer-based analyses of features of writing is the ability to derive these measures from samples of extended writing rather than from short-answer or multiple-choice formats. This could provide a richer portrait of writing achievement without sacrificing the emphasis on the creation of complete texts.

Considerations

The Writing Framework Committees will need to consider which aspects of writing development will be of most interest to educators and the general public, and how these aspects of development can best be measured.

Issue 6: What should students write about?

The current framework for the Writing Assessment emphasizes writing prompts that are accessible to all students. In practice, this results in an emphasis on common life experience, generic academic content (e.g., favorite books or music, fictional or historical figures, the value of space travel) and on writing that reflects public discourse in a democratic society (e.g., persuasive tasks about community or school issues). If content is provided, it is typically more illustrative than substantive—a brief “story starter,” a picture stimulus, or a brief framing of sides on a “controversial” issue. (Real controversies that have political volatility do not make it through the item-review process.) When reading and language difficulties of English language learners and low achieving students are taken into account, the push in item development is toward simple and “clean” writing prompts with a low vocabulary load.

At the same time, writing plays a role in virtually all of the other subject area assessments in NAEP. Both short and extended constructed responses comprise major sections of the current assessments in science, history, geography, civics, and reading, as well as the frameworks for new assessments in economics and foreign languages. Rubrics in these assessments bear little similarity to the rubrics in the writing assessment, however, often emphasizing listing of specific content rather than the construction of an argument or explanation.

This creates an artificial separation of writing from content knowledge. As Hillocks (2002) points out in his critique of state writing assessments, one of the biggest problems in many assessments is the lack of a substantive content base on which to base the writing. Without a content base, much of the writing that results is formulaic and shallow.

Considerations

Thus an issue for the Framework Committees to consider is whether it would be useful to increase the content load of student writing prompts, and if so, how this could be done within the current assessment framework or through extensions of it. One possibility, particularly if writing and other assessments become computerized, would be through the adoption of some common metrics across assessments.

Issue 7: How should the Framework address the question of time?

Time to write has been an issue for successive NAEP Writing Framework Committees, and has led both to changes in time allotments and to special studies (NAGB, 1996). From 1970 to 1979, NAEP writing assessments had items of variable length, from a few minutes for completing forms to nearly 30 minutes on some essay tasks. The move to a balanced incomplete block design (BIB spiraling) in the 1984 assessment reduced the maximum time to 15 minutes. Beginning with the 1992 assessment, this was increased to 25 minutes (with a subset of 50 minute writing tasks that was eliminated in the 2002 assessment).

Two issues usually dominate discussions of writing time: Do the results misrepresent overall writing achievement because students have too little time to write? And does the limited time allowed penalize some groups of students, particularly those whose classrooms have emphasized an extended process of writing and revision? (Conversely, will extended time frustrate lower achieving students and exacerbate achievement gaps?)

The issue of time has been driven by a tension between the constraints of assessment and the conventional wisdom on instruction. One of the accomplishments of the writing process movement in instruction was to remind teachers and students that writing takes place over time—that there are identifiable strategies for generating ideas, drafting, revising, editing, and sharing that shape and reshape a final written text. During the past 30 years of writing assessment, the proportion of teachers claiming to emphasize process-oriented approaches to writing instruction has risen sharply; by 1998 it was central to the instruction of 70% of fourth grade teachers surveyed, and used to supplement instruction by another 28%. Comparable figures were reported by 8th and 12th grade students in the 2002 assessment. (Background questions and grade levels at which they are asked vary from assessment to assessment so there is no single set of data on which to draw.)

Given the constraints of large-scale assessment, NAEP has always emphasized that the writing assessment focuses on first-draft writing (as do ETS and ACT in their college entrance examinations). Given the overall design of the assessment, when NAEP has

included 50-minute tasks the trade-off has been these tasks have not been scalable. (With a 50 minute prompt, each student completes only one task, so interrelationships among tasks cannot be determined.) In 1998 the results of these longer tasks do not seem to have even been reported.

Previous NAEP studies of the impact of additional time have yielded mixed results. One special study compared 11th-graders' performance on a persuasive writing task given in 16- or 50-minute time blocks but mixed together for scoring with identical rubrics. As common sense might suggest, the students who had more time for writing scored higher—though the gain was less than might have been expected: 45.4% produced adequate or better responses in 50 minutes, compared with 33.8% in the 16-minute format. The benefits of extra time were not equally distributed among students, however; the extra time made little difference to the weaker writers, increasing the performance gap between the two groups (Applebee, Langer, & Mullis, 1989). The 1992 assessment reported results for 50-minute as well as 25-minute prompts, with achievement noticeably higher on a 50-minute informative writing task than on the other, 25-minute informative tasks. But comparable increases in achievement did not occur on 50-minute narrative and persuasive tasks included in the assessment; the report concluded that the differences were likely to be topic-related rather than a function of the increased response time (Applebee et al., 1994; p. 27-30).

These results do not mean that time is not an important factor in quality of writing; simply that the effects of time within the constraints of NAEP writing prompts as they are currently designed are not as large as might be thought, and may be topic specific. It may be that for meaningful effects of time to emerge, the nature of the tasks would need to be radically reconstrued to incorporate, for example, significant content to be examined or reviewed, or significant feedback to be provided after an initial draft.

Related to the issue of time is whether to provide any special supports for students as they write, particularly supports related to how students use the time available to them. The current assessment format, for example, includes blank space which students are encouraged to use to plan their writing. Students also receive a booklet, "Ideas for Planning and Reviewing Your Writing," that suggests planning and revision strategies.

Considerations

If, as seems likely, the 25-minute format is the maximum that will be available for 2011, the Framework Committees should consider other ways to bolster the validity of results from on-demand, 25-minute tasks. In the past, one way to do this was to pair the main assessment with a study of classroom-based writing. Such a study makes it possible to address many of the questions raised about the 25-minute assignments: Do they misrepresent the writing instruction students are receiving? Does writing achievement look markedly better when examined from the perspective of the classroom rather than the timed examination? Are students from process-oriented classrooms being penalized by the first-draft format?

Of course there may be other formats for special studies that would address these issues more effectively, variations that might play off of innovative formats in state assessments. New York State, for example, provides substantive material for students to read and write about, using extended, three-hour time blocks. Kentucky pairs classroom-based writing with an assigned task, and also insists that some of the classroom-based writing come from subject areas other than English. Hillocks (2002) comments favorably on both of these assessments in his critical look at the quality of writing elicited by various approaches to writing assessment at the state level.

Issue 8: What accommodations should be made for English language learners, students with disabilities, and low-achieving students?

NAEP policy is to include as many students as possible in all of its assessments, without altering the construct being measured. In practice this is accomplished by careful item development procedures and, where necessary, by providing accommodations to students with disabilities and English Language Learners. Typical NAEP accommodations include more testing time, small group testing, and other appropriate accommodations depending on the NAEP subject being tested.

As noted earlier, recent Writing Framework Committees have been concerned to make all writing prompts as accessible as possible to all students. This has usually meant a lightening of the vocabulary load and of content provided through the prompt, so that these students would not be put off by problems in understanding before even beginning with their own writing.

The inevitable consequence of this accommodation in the current booklet-based testing format has been that there has been little room to experiment with alternative formats that have the possibility of providing a more substantive context for at least some of the writing tasks.

A computer-based assessment in 2011 would open up a variety of new possibilities, particularly if paired with writing analysis software that could make rapid initial judgments about writing proficiency of individual students. A simple “range finder” task, for example, might be used to place students in alternative formats adjusted to their general literacy levels. (New Zealand, for example, uses a very simple and quick initial task in its reading assessment, http://nemp.otago.ac.nz/read_speak/2000/_media/RR_Indicator.jpg). Or the response level on the first task administered to each student could be used (with computerized scoring) to select a second task of appropriate difficulty. This could serve to provide accommodations for students who need them, and also to provide greater challenges for higher-ability students.

Considerations

The Framework Committees should give serious consideration to the effects of accommodations for poor readers and English language learners on the overall content of

the assessment, and look for alternatives that might provide a richer array of assessment options for all students.

Conclusion

The NAEP Writing Framework has evolved significantly over the years, in the nature of the writing prompts, in the time available for each task, and in its emphasis on rhetorical features such as audience and purpose. By 2011, the current assessment framework will be 20 years old—more than time for an update that will reflect recent changes in scholarship and practice, and that will also return NAEP to its rightful position as a leader in assessment practice and assessment technology. Most of the issues outlined above have no easy answers, but the collective experience of the members of the Framework Committees offers an opportunity to make significant progress in addressing them.

References

- American Diploma Project. (2004). *Ready or not: Creating a high school diploma that counts*. Washington, DC: Achieve, Inc.
- Applebee, A.N. (1974). *Tradition and reform in the teaching of English: A history*. Urbana, IL: National Council of Teachers of English.
- Applebee, A. N., Langer, J. A., & Mullis, I. V. S. (1989). *Understanding direct writing assessments: Reflections on a South Carolina writing study*. Princeton, NJ: Educational Testing Service.
- Applebee, A. N., Langer, J.A., Mullis, I.V.S., Latham, A.S., & Gentile, C. A. (1994). *NAEP 1992 writing report card*. Washington, DC: U.S. Government Printing Office for the National Center for Education Statistics, Office of Educational Research and Improvement, U.S. Department of Education.
- Bereiter, C., & Scardamalia, M. (1982). From conversation to composition. In R. Glaser (Ed.), *Advances in instructional psychology*. (Vol. 2). Hillsdale, NJ: Erlbaum.
- Bereiter, C., & Scardamalia, M. (1987). *The psychogy of written composition*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Breland, H. M., Camp, R., Jones, R. J., Morris, M. M., & Rock, D. (1987). *Assessing writing skill*. New York: College Entrance Examination Board.
- Britton, J., Burgess, T., Martin, N., McLeod, A., & Rosen, H. (1975). *The development of writing abilities (11-18)*. London: Macmillan Education.
- Cope, B., & Kalantzis, M. (Eds.). (1993). *The powers of literacy: A genre approach to teaching writing*. Pittsburgh, PA: University of Pittsburgh Press.

- Diederich, P. B. (1974). *Measuring growth in English*. Urbana, IL: National Council of Teachers of English.
- Diederich, P. B., French, John W., & Carlton, Sydell T. (1961). *Factors in judgments of writing ability. Research Bulletin RB-61-15*. Princeton, NY: Educational Testing Service.
- Godshalk, F. I., Swineford, F., & Coffman, W. E. (1966). *The measurement of writing ability*. New York: College Entrance Examination Board.
- Halliday, M.A.K., & Martin, J.R. (1993). *Writing science: Literacy and discursive power*. Pittsburgh: University of Pittsburgh Press.
- Hillocks, G., Jr. (2002). *The testing trap: How state writing assessments control learning*. New York: Teachers College Press.
- Kinneavy, J. (1980). *A Theory of discourse: The aims of discourse*. New York: W.W. Norton & Co.
- Lloyd-Jones, R. (1977). Primary trait scoring. In C. R. Cooper & L. Odell (Eds.), *Evaluating writing: Describing, measuring, judging* (pp. 33-66). Urbana, IL: National Council of Teachers of English.
- Martin, J. R., & Rothery, JJ. (1980). *Writing Project Report No. 1*. Sydney: University of Sydney, Department of Linguistics.
- Miller, C. R. (1984). Genre as social action. *Quarterly Journal of Speech*, 70, 151-167.
- Moffett, J. (1968). *Teaching the universe of discourse*. Boston: Houghton Mifflin.
- National Assessment of Educational Progress (NAEP) (1982). *Writing objectives 1983-84 assessment*. Princeton, NJ: National Assessment of Educational Progress, Educational Testing Service.
- National Assessment Governing Board (2002). *Writing framework and specifications for the 1998 National Assessment of Educational Progress*. Washington, DC: U.S. Department of Education.
- National Commission on NAEP 12th Grade Assessment and Reporting. (March 4, 2004). *12th grade achievement in America: A new vision for NAEP. A report to the National Assessment Governing Board*. Washington, DC: National Assessment Governing Board.
- Persky, H., Daane, M.C., and Jin, Y. (2003). *The nation's report card: Writing 2002*. U.S. Government Printing Office for the Department of Education, National Center for Education Statistics.

Russell, M. & Plati, T. (2002). Does it matter with what I write? Comparing performance on paper, computer, and portable writing devices. *Current Issue in Education* (Online) 5 (4). Available: <http://cie.ed.asu.edu/volume5/number4>.

Sandene, B., et al. (2005). *Online assessment in mathematics and writing: Reports from the NAEP technology-based assessment project*. Washington, DC: National Center for Education Statistics.

Appendix: The Evolution of the NAEP Writing Framework

Cross-Sectional Writing Assessments

1969-70 Assessment

1. Write to communicate adequately in a social situation.
2. Write to communicate adequately in a business or vocational situation.
3. Write to communicate adequately in a scholastic situation.
4. Appreciate the value of writing.

1973-74 and 1978-79 Assessments

1. Demonstrates the ability in writing to reveal personal feelings and ideas (through free expression and through the use of conventional modes of discourse. [For 1978-79, reinterpreted as “ability to engage in writing for expressive purposes.”])
2. Demonstrates the ability to write a response to a wide range of societal demands and obligations. Ability is defined to include correctness in usage, punctuation, spelling and form or convention as appropriate to particular writing tasks (Social, Business/ Vocational, Scholastic). [For 1978-79, interpreted as explanatory or persuasive writing done for a particular audience.]
3. Indicates the importance attached to writing skills (recognizes the necessity of writing for a variety of needs, writes to fulfill those needs, and gets satisfaction, even enjoyment, from having written something well.

1983-84 and 1987-88 Assessments

1. Students use writing as a way of thinking and learning (for subject knowledge and self knowledge).
2. Students use writing to accomplish a variety of purposes (informative, persuasive, and literary). {Literary was variously interpreted as “imaginative” and as “personal /imaginative narrative” in reports on these assessments.}
3. Students manage the writing process (generate, draft, revise, edit).
4. Students control the forms of written language (organization and elaboration, conventions).

5. Students appreciate the value of writing (for interpersonal communication, for society, and for self).

1991-92, 1997-98, 2001-02, and 2006-07 Assessments

1. Students should write for a variety of purposes: narrative, informative, and persuasive.
2. Students should write on a variety of tasks and for many different audiences.
3. Students should write from a variety of stimulus materials, and within various time constraints.
4. Students should generate, draft, revise and edit ideas and forms of expression in their writing.
5. Students should display effective choices in the organization of their writing. They should include detail to illustrate and elaborate their ideas, and use appropriate conventions of written English.
6. Students should value writing as a communicative activity.

Long Term Trend Assessments

1969-79 through 1983-94

Writing prompts developed using the 1969-70 framework were re-administered to study long-term trend through 1983-84, though trend reports have reinterpreted prompts in light of the writing objectives in place at the time of reporting.

1983-84 through 1995-96

Writing prompts developed using the 1983-84 framework were re-administered to study long-term trend through 1996, again with reinterpretation of prompts in light of later revisions to the writing framework. Two assessments (1993-94 and 1995-96) were limited to long-term trend. The last writing long-term trend assessment administered and reported was for 1995-96. Although writing long-term trend data were collected in 1999, results were not reported due to instability of the score scale. NCES and NAGB determined that the writing long-term trend assessment be discontinued since too few prompts were administered to enable reporting of viable trend results.

Framework References

Norris, E. L. (Ed.) *Writing objectives*. Ann Arbor, MI: Committee on Assessing the Progress of Education, 1969.

National Assessment of Educational Progress (1972). *Writing objectives: Second assessment*. Denver, CO: Education Commission of the States.

National Assessment of Educational Progress (no date). *Supplement to the 1973-74 writing objectives*. Unpaginated insert to Writing objectives; Second assessment.

National Assessment of Educational Progress (1982). *Writing objectives: 1983-84 assessment*. Princeton, NJ: Educational Testing Service.

National Assessment Governing Board (2002). *Writing framework and specifications for the 1998 National Assessment of Educational Progress*. Washington, DC: U.S. Government Printing Office.