A Rescue Plan for the NAEP Long-Term Trend Assessments: Thoughts on Edward Haertel's White Paper

Prepared for the National Assessment Governing Board

Andrew Kolstad, Ph.D.

February 2017

SUMMARY: In order to rescue the Long-Term Trend (LTT) assessments from failure, three activities will soon be necessary. First, administrative procedures for the LTT assessments should be changed by (a) adapting the paper-based items into a form compatible with administration on a digital platform—the way it is being done for the rest of the NAEP assessments, (b) shifting the testing window for all three age groups to correspond to the rest of the NAEP assessments, and (c) planning to field test replacement cognitive exercises in the year prior to each LTT assessment. Second, a bridge study should be conducted within the next three years to connect the old and new procedures for conducting the LTT assessment. Third, the framework and item specifications for the LTT reading and mathematics assessments should be rewritten to assist the public in understanding what these assessments were intended to measure.

Ed Haertel's white paper addresses two issues central to the future of NAEP's LTT assessments. The first issue is whether the LTT assessment component of NAEP should be continued in essentially its current form, or dropped altogether. I agree with Haertel's conclusion that the LTT assessments should be continued, but modifications that maintain trend lines will be necessary. Continuation is a policy decision that I will leave to policy experts, to the people who balance budget priorities, and to the National Assessment Governing Board. I'm more of a technical expert and will direct my comments to technical concerns.

The 2004 redesign of the LTT assessments came about because the prior design, which has been in place since the mid-1980s, had become unworkable. Dropping the science and writing assessments left holes in the assessment instruments, and students with disabilities were not being provided accommodations. The 2004 redesign and its associated bridge study corrected this situation, but left some problems unresolved. After two more decades, in 2024, the 2004 design for the LTT assessments again needs attention, because by that time the design will have become unworkable. If it is not redesigned, I believe that it cannot succeed.

The second issue Haertel raised is the operational and administrative modifications that are needed to ensure efficient conduct of the LTT assessments. I agree with Haertel's proposal to gain operational economies by integrating their mode of administration, their sampling activities, their field operations, and their scoring activities. I also agree that moving from paper to digitally based assessment formats will take advantage of the economies of digital administration. It is not simply a matter of how the students take the assessment, but a matter of the entire survey operations infrastructure. Resource savings would derive from not having to maintain separate production facilities for paper test forms (printing, distribution, data collection, converting to digital data files, and scoring the constructed responses). Such an integration plan will require a bridge study, in which the current and revised field operations are both conducted in the same year, so that the trend lines can be shown to be continuous.

My belief is that the current plan for the next administration in 2024 of the LTT assessments will be too expensive to succeed. This assessment can be cancelled or postponed again on

relatively short notice, but without some significant changes to reduce costs, this assessment survey program cannot continue. If it is to survive, it must be rescued (hence my paper's title).

I want to raise a third issue, which Ed Haertel's white paper addresses only indirectly. That issue is how to remedy the inadequacies of the materials that serve in place of a content framework and item specifications for the LTT assessments. While Haertel was able to develop an understanding of the content of the reading and mathematics assessments, he found the process surprisingly difficult. It is also difficult for the public and for NAEP contractors whose job it is to develop replacement cognitive exercises to get a firm grasp on what the LTT assessments are intended to measure. In my opinion, remediation of the framework is also essential, but can be addressed on parallel track with the other administrative changes (and could take longer to implement).

The technical challenges in merging the cognitive assessment content of the main NAEP and LTT assessments are insurmountable. Irreconcilable differences exist between the LTT and main NAEP assessments. One difference is that the two assessments use incompatible exercise booklet formats (with three 15-minute versus two 30-minute blocks of cognitive exercises). Pairing LTT and main NAEP exercise blocks in the same booklets would result in administrative difficulties due to mismatched timing. While this design feature could be changed, other differences virtually eliminate the value of making such a change.

The LTT assessments are commonly understood to measure the basic skills that were considered important two generations ago, but Haertel showed in his white paper that this vague understanding is inadequate. The content and cognitive processes assessed by LTT exercises fall within and below the range of typical curricular expectations for grades 4, 8, and 12, but more advanced topics and more complex processes at these grade levels are largely omitted.

Another important difference is that the LTT assessments' age-based populations contain substantial minorities who are enrolled below the modal grade level. At each age group, there is a mixture of a majority of students at the typical grade (4, 8, and 11) and another substantial proportion below the modal grade level (in grades 3, 7, and 10). Among NAEP's 9- and 13-year-old populations, 37 percent were enrolled in grade 3, not grade 4, and 39 percent in grade 7, not grade 8. Among NAEP's 17-year-olds, 23 percent are enrolled in grade 10, not grade 11, and only a small proportion are enrolled in grade 12 (see Table 1). The LTT cognitive exercise pool is appropriate for the inclusion of a substantial minority of students below the modal grade level and does not include the advanced content that is important to main NAEP.

A corresponding difference exists for main NAEP's grade-based populations. Among fourth graders, a substantial proportion are 10 years old. Among eighth graders, a substantial proportion are 14 years old. And among 12th graders, a substantial fraction are 18 years old when assessed. Items that are grade-appropriate, or include more advanced content for main NAEP are likely to be too hard for the below-modal-grade proportion of the LTT's population.

Haertel concluded, and I agree, that the technical challenges in merging the cognitive content of the main NAEP and LTT assessments cannot be surmounted. The cognitive measurements and target populations of the two NAEP assessment programs are too different to be integrated

into a common, dual-purpose assessment. Haertel recommended, and I agree, that two operational aspects of the LTT assessment should be substantially revised: (a) the assessment should shift from paper-based to digitally based assessment forms; and (b) the testing window for all three age groups should be time-shifted to correspond to that used for the rest of the NAEP assessments. These changes would make possible the pilot testing of replacement cognitive exercises the year prior to each LTT assessment, rather than four years in advance.

Without the need to include main NAEP assessment blocks or items of reading and mathematics in the LTT (or vice versa), there is no need to move the schedule for the LTT assessments into the same years that the main NAEP reading and mathematics assessments are administered.

The LTT assessments should stop using paper forms and switch to digitally based assessment instruments, just as the rest of NAEP is doing. Future resources can be conserved by redesigning these assessments to conform their administrative procedures to those of other NAEP assessments, while maintaining the defining content differences between reading and mathematics in main NAEP and in the LTT assessments.

I foresee an administrative obstacle to the current assessment schedule in which the next LTT is administered in 2024 on paper forms. The bulk of the transition to digital administration for the most of NAEP's assessments is scheduled to be completed by in 2019, with relatively small national-only assessments scheduled for 2020 and beyond. The NAEP project's capacity to handle paper-based assessment forms may have been lost several years before the next scheduled administration of the LTT assessments in 2024. It will be hard to justify the expense of maintaining or recovering NAEP's capacity to handle paper-based administration for the relatively small, national-only samples of the LTT.

The staff and contractors for the main NAEP assessments have gained substantial experience when switching from paper-based to digitally based assessment formats. The LTT assessments can take advantage of this experience. Since the vast majority of cognitive exercises in the LTT assessments are multiple-choice, the trans-adaptation of the items onto a digital platform should not be problematic. Instituting such changes will nevertheless require an initial investment in delivery software and converting the existing item pool into digital formats.

I do not foresee difficulties with moving to computer-based administration. However, it does take a certain amount of lead time to translate the paper-based items into a form suitable for administration on a digital platform and to develop the software needed to administer the items.

The LTT assessments should shift the time of year during which they are conducted in order to share organizational capacities with other NAEP assessments. Moving to a common testing window—the period from late January to early March—would provide for integrated sampling and data collection procedures. The savings would derive from not having separate sampling operations, school recruitment, staff training activities, field staff, scoring operations, and the management activities to oversee them over the longer period.

Haertel noticed that changing the time of year of administration will change the average age of the students being assessed and proposed that the date ranges that define age should be

changed as well, in order to maintain the same average age of the student populations. Table 1 presents the current date ranges that define ages 9, 13, and 17 for the purpose of determining whether students are part of NAEP's three age-based populations, along with the dates during which NAEP conducts the LTT assessments. Subtracting the midpoints of the two date ranges provides an estimate of the average age at the time of testing for the three age groups. The NAEP Data Explorer provides estimates of the proportion of 9-, 13-, and 17-year-old students in 2012 who were enrolled below the modal grade level. Each age group has substantial proportions of students below the modal grade level and small proportions (not shown in Table 1) who are enrolled above the modal grade level.

However, Haertel overlooked another consequence of changing the date ranges, and that is that the proportion of students below the modal grade level would also change. Schools have fixed cutoff dates for enrolling students when they start school. Changing NAEP's date ranges would shift more or fewer students across the local age-of-school-entry boundaries.

I did some investigation into what the effect might be. The Census Bureau publishes annual tables of single grade enrollments by single years of age based on the Current Population Survey October school supplement. I used these tables to estimate the proportion of 9-, 13-, and 17-year-old students who are enrolled below the modal grade level, which depends on whether the defining age range is changed by three months (as Haertel recommended), by one month, or not at all. These results are shown in Table 2.

Using the current age-defining date ranges, I was able to project the proportion of students who are below grade level for each age group and match within a couple of percentage points the actual proportions below grade level that were reported in the most recent LTT online data in Table 1. Haertel noticed that if the LTT assessment were to be administered during the main NAEP data collection window, the 13-year-old students would be three months older than they would have been during the fall testing window. Haertel proposed shifting the defining date range for these students later by three months in order to maintain the same average age. My projection (in Table 2) indicates that this change would result in a decrease of the proportion enrolled below the eighth grade from 41 percent to 30 percent. This decrease could have just as much impact, in the opposing direction, as the change in scores attributable to being in school for three months longer. Consequently, I recommend not changing the current date ranges that define age for the 13-year-olds.

Haertel also noticed that if the LTT assessment were to be administered during the main NAEP data collection window, the 17-year-old students would be three months younger than they would have been under the current testing window. In fact, the average age of "17-year-olds" at the new time of assessment would fall to from just over 17 to just under (16.9 years old). He proposed shifting the defining date range for these students earlier by three months in order to maintain the same average age. My projection indicates that this change would result in an increase in the proportion enrolled below the 11th grade from 24 percent to 30 percent. This increase could have just as much impact, in the opposing direction, as the change in scores attributable to attending school three months less.

However, a case could be made that the average age at assessment for a population of "17-year-olds" should be at least 17.0 years. Changing the defining date range by one month (rather than Haertel's suggested three months) would bring the average age above 17.0 years and raise the projected proportion enrolled below 11th grade from 24 to 26 percent. In my opinion, accepting a small increase in the proportion below modal grade might be a price worth paying to keep the average age at assessment over 17 years.

The development and pilot testing of replacement cognitive exercises also deserves attention. The Governing Board has a policy of releasing NAEP assessment items. One reason for the 2004 revision of the design for the LTT assessments was to make possible the release and replacement of old cognitive items, in order to inform the public about the nature of the NAEP LTT assessments. No cognitive exercises had been released from these assessments since the Educational Testing Service (ETS) took over the contract from Education Commission of the States (ECS) in 1983. Between 1984 and 1999, no replacement items were developed. Over these years, the craft of writing items to meet the LTT objectives was not passed along to the next generation of item writers, and the institutional memory at ECS about assessment objectives was lost. The content objectives have not been revisited in three decades, and the committees that set the objectives were disbanded long ago.

When in 2001-02, ETS began for the first time to develop replacement cognitive exercises for the 2004 assessment, the item development staff at ETS were unfamiliar with the objectives of LTT assessments. Lacking a content framework and item specifications, the staff created their own item specifications based on the existing item pool in place in 1999. The ETS item writing staff have tried to mimic very closely the features of items that are being released (and those in the item pool), rather than write new items to meet the old lists of objectives.

The existing main NAEP standing committees for reading and mathematics have subcommittees that review the items for the item development contractors as they are being created. However, these content experts are much more familiar with main NAEP frameworks and specifications than with the lists of objectives that are supposed to underlie the content of the LTT assessments. Because the materials that define the intended content of the LTT assessments are inadequate, this group of experts is not able to provide guidance on the adequacy of the coverage of the framework with the current set of items, and any holes that might need to be filled with replacement items.

New items were expected to be field-tested on the same age-based sample as the operational LTT assessment. This required such items to be ready four years in advance of the next administration. Allowing time for development means that staff needed to be working to develop replacement items at least five or six years ahead of their use. Integrating the sampling procedures with the main NAEP testing window will make it easier to integrate the field testing of replacement exercises with other NAEP assessments. Such a change would make possible a much shorter lead time for item development and a shorter interval between field testing and operational administration of the LTT assessments.

A bridge study is necessary to connect the old and the new administrative procedures. Haertel proposed conducting a bridge study in which the LTT assessment is administered twice, once in digital form and again with paper assessment forms. The two administrations would be linked either through a common population or through common items in the same year (or both). Comparison of the results of the two assessments would be able to demonstrate that the trend lines can (or cannot) be preserved. Continuity with past assessments would be ensured by linking to the paper version, and continuity with future administrations ensured by linking to the digital version.

Bridge studies like this are being conducted during the years 2015–2019 in other NAEP subjects, in order to maintain trend lines across modes of administration (ETS, 2015). By the time such a bridge study would need to be conducted for the LTT assessment, NAEP will have had a great deal of institutional experience with bridge studies that link digital and paper forms of the assessment.

The LTT bridge study needs to be conducted before the capacity to handle paper-based administration disappears from NAEP. This means that the currently scheduled 2024 LTT assessment, if it were to be administered digitally, would have to be preceded by a bridge study conducted several years earlier, before NAEP's paper-handling capacity is lost. I recommend that the Governing Board conduct the bridge study for the LTT assessment in 2020, and return to the four-year interval between administrations of the NAEP LTT assessments in 2024.

Each condition in a bridge study requires funding, and the combination of both would require more resources than simply re-administering the current paper-based design in its three testing windows. It is as if NAEP were to administer two LTT assessments in reading and mathematics at once (one on paper booklets and one on a digital platform).

Haertel proposed having a third experimental condition to distinguish the effects of moving from paper to digital administration from the effects of moving the time of year of the assessment. I think the extra resources required to have a third condition would not be worth the costs. Unless it would be possible to implement only one of the two administrative procedures, there is no need to know the separate effects of the two changes. But both of these changes are necessary to keep the costs of the LTT assessments under control. It would not be feasible to implement only one of them.

The objectives booklets that define content coverage for the LTT assessments are insufficient to explain what is intended to be measured. Haertel found it surprisingly difficult to clarify just what the LTT assessments measure. No explicit content frameworks exist for the LTT assessments. The topics that these assessments should cover are defined instead by booklets of objectives that provide much less detail than any current NAEP content framework. Main NAEP provides item specifications for the item writing staff, but no such item specifications exist for the LTT assessments.

The objectives booklets for reading and mathematics changed over time. During the first four cycles of reading and mathematics assessments, the lists of objectives changed with each administration. Because ECS thought that NAEP should provide a model of excellent cognitive

exercises to the field of education, about half of the cognitive exercises and reading passages were released each time. The pool of common items shrunk during NAEP's first decade, until the test booklets were frozen.

The trend lines that now constitute the LTT assessments emerged from the reading and mathematics scales developed by ETS when it took over the NAEP project in the early 1980s. The cognitive exercise pool during the frozen period are those items that survived screening on technical criteria, screening for bias, and screening for outdated or obsolete content. As a result, the item pool does not fully reflect the intended final list of objectives. It is quite likely that the existing item pool does not even fully cover the final set of NAEP objectives.

I have included an appendix containing references to published materials from the early days of NAEP. This includes all nine objectives booklets for reading and mathematics and released items from those early ECS years. Nearly all such materials are available from the Educational Resources Information Center, in printed form or downloadable as scanned documents.

A concerted effort is needed to clarify what the LTT assessments measure in reading and mathematics. A rewritten framework and item specifications would assist the pubic in understanding what it is that the LTT assessments were intended to measure. These materials would also be invaluable in developing cognitive exercises that not only replace released items, but also fill in gaps in intended content coverage that has long been missing.

In the decades since the transition from ECS to ETS, the standards for what constitutes an adequate content framework and item specifications have changed, and the NAEP authorization law assigned the responsibility for developing frameworks to the National Assessment Governing Board. The old objectives booklets are no longer sufficient guidance for developing replacement items. A rewritten framework and item specifications are needed to develop and field test cognitive exercises that can replace exercises as they are released.

I believe that the National Assessment Governing Board is the institution best suited to conduct the activities needed to retrofit an updated framework and item specifications onto the LTT reading and mathematics assessments. Currently, the Governing Board decides on the frameworks for main NAEP, and re-evaluates and makes changes in these frameworks from time to time. The law does not address the Board's responsibility with respect to the content of the LTT assessments, but it assigns duties that can be understood to apply to this aspect of the LTT as well as main NAEP assessments.

I read NAEP's authorizing legislation again, to see who was assigned the responsibility to oversee the content of the LTT assessments. I found that The National Assessment of Educational Progress Authorization Act (P.L. 107-279) does not prohibit the Governing Board from taking on this responsibility for the LTT assessments. Under 20 USC 9621, Section 302(e)(1) paragraph (C) the "Assessment Board" has been given the duty to "develop assessment objectives consistent with the requirements of this section and test specifications that produce an assessment that is valid and reliable." Paragraph (F) of this section implies that this duty pertains to the main NAEP assessments in grades 4, 8, and 12, but does not specifically mention this duty with respect to the LTT assessments for ages 9, 13, and 17. Paragraph (I) of the same section

assigns to the Board the duty to take action to improve the content of the NAEP assessments, not distinguishing between the main and the LTT assessments.

I found that the part of the law that *does* address the LTT assessments [20 USC 9622, Section 303(a)], assigns the responsibility for these assessments to the Commissioner of Education Statistics (with advice from the Assessment Board), but makes no mention of "assessment objectives" or "test specifications." However, NCES does not have experience with managing committees to originate or re-evaluate content frameworks. The National Assessment Governing Board does have considerable experience with such activities and should more appropriately take on this role.

From the existing objectives documents and item pools that are currently, or have been included in the item response theory (IRT) reading and mathematics scales, the Governing Board ought to be able to develop framework and specifications documents that would provide a blueprint for developing replacement reading and mathematics exercises for future administrations of the LTT assessments. The goal of such a project would be not to develop a new framework or to make such changes in the framework that would cause a break in trend, but to make explicit the frameworks and item specifications of this existing assessment survey program. Developing a retrofitted framework and item specifications that conforms as closely as possible to the intentions of the old objectives and the existing item pool may take several years, but in my opinion, this is an essential activity if the LTT assessments are to be preserved.

While creating a framework and item specifications for reading and mathematics are essential to the long-run health of the LTT assessments, such activities are not likely to be completed within three years. Since there already exists a supply of replacement items that can be used in a 2020 bridge study, completing the framework updating activity is not necessary prior to undertaking a LTT bridge study.

Summary of what needs to be done to rescue the LTT assessments. In order to rescue the LTT assessments from failure, three activities will soon be necessary:

1. Changing the administrative procedures for the LTT assessments by (a) moving the paper-based LTT cognitive exercises and reading passages onto a digital platform for administration the way it is being done for the rest of NAEP assessments, (b) shifting the testing window for all three age groups to correspond to the testing window for the rest of the NAEP assessments, and (c) planning to field test replacement cognitive exercises in the year prior to each LTT assessment.
2. Conducting a bridge study in 2020 to connect the old and new LTT lines.
3. Rewriting/retrofitting the framework and item specifications for the LTT reading and mathematics assessments to assist the pubic in understanding what it is that these assessments were intended to measure.

References

Educational Testing Service (2015). *NAEP's Transition to Digitally Based Assessment*. White paper prepared for the National Center for Education Statistics.

Table 1: Testing windows, age midpoints, and modal grade, by age group and various age definitions: NAEP 2012 Long-Term Trend assessments

	LTT assessment windows and LTT birthdate ranges		
	Age 9	Age 13	Age 17
Date range of assessment window	January 2, 2012 to March 9, 2012	October 10, 2011 to December 16, 2011	March 12, 2012 to May 11, 2012
Midpoint of assessment period	2/4/2012	11/12/2011	4/11/2012
Date range of birth year	January 1, 2002 to December 31, 2002	January 1, 1998 to December 31, 1998	October 1, 1994 to September 30, 1995
Midpoint of birth year	7/2/2002	7/2/1998	4/1/1995
Age at assessment	9.6	13.37	17.04
Percent below modal grade	37	39	23

SOURCE: U.S. Department of Education, National Assessment of Educational Progress (NAEP), 2012 Long-Term Trend Reading Assessments.

Table 2: Birth year midpoints, main NAEP assessment period, and projected age at assessment and modal grade, by age group and three birthdate ranges: NAEP 2012 Long-Term Trend assessments

		3-month change in birthdates	1-month change in birthdates	No change in birthdate range
9	Midpoint of birth year	—	—	7/2/2002
	Midpoint of assessment period	—	—	2/15/2012
	Average age at assessment	—	—	9.6
	Projected percent below modal grade	—	—	40
13	Midpoint of birth year	10/2/1998	8/2/1998	7/2/1998
	Midpoint of assessment period	2/15/2012	2/15/2012	2/15/2012
	Average age at assessment	13.4	13.5	13.6
	Projected percent below modal grade	30	35	41
17	Midpoint of birth year	12/31/1994	2/28/1995	4/1/1995
	Midpoint of assessment period	2/15/2012	2/15/2012	2/15/2012
	Average age at assessment	17.1	17.0	16.9
	Projected percent below modal grade	30	26	24

SOURCE: U.S. Census Bureau, Current Population Survey, October supplement 2011 (Table 2: Single years of enrollment by single years of age).

Appendix

EARLY NAEP PUBLICATIONS

Some early NAEP materials describing the content of what became the LTT mathematics and reading assessments are available from the Educational Resources Information Center (ERIC) in microfiche or hard copy, or downloadable as scanned documents. ED numbers are provided, enabling the user to consult any of approximately 600 libraries and ERIC clearinghouses and depositories around the United States, including the ED library in FB6, or to download electronic versions of the documents.

The term "technical report" as used in the early days of NAEP does not correspond to current usage in NAEP. They were published in several volumes. It appears that some volumes contained further tabulations of results and others may have contained sets of released items. The "exercise volumes" are fat documents that contain detailed pages of tabulations (text of an item, followed by performance on that item by various population subgroups).

ETS maintains a computerized database containing information about every item used by NAEP during its 30-year history—all of the descriptive, processing, and usage information on every NAEP item since 1969. More than a decade ago, ETS created a unified database that incorporated the item database received from Education Commission of the States (ECS) with its own database of items.

Early Methodological Publications

SY-ED-70 *The National Assessment Approach to Exercise Development*. A technical booklet outlining the early days of developing exercises, 1970.
ED 067 402 .....

12-IP-55 *The National Assessment Approach to Objectives and Exercise Development*. A policy paper describing methods for developing objectives and test specifications, writing and testing items, structuring test booklets, and scoring items. By Barbara Ward, 1980.
[NCES has a copy].....

432 *Stability of the National Assessment Scoring Methods*, by Nancy W. Burton. An article published in the Summer 1980 issue of *Journal of Educational Measurement*
.....

SY-OI-36 *A Guide to National Assessment Objectives and Items*. An 8-page folder explaining what NAEP has produced in the way of objectives and exercises or items in the various learning areas. .....

Early Mathematics Publications

The earliest mathematics assessments included in the ETS trend lines (1978-79) were administered using paced audio tapes to assure uniform assessment conditions. A bridge study without the paced tape was conducted in 1985-86 for the mathematics/science trend assessments. Since the impact on science was large, the paced tape was retained until the science assessment was dropped after 1999. In the early years, half the assessment items were released for use by the public after each administration.

Technical Materials:

04-MA-20 *Mathematics Technical Report: Exercise Volume*. 1,412 pp. (marginal legibility of original document). Appendix A discusses the mathematics objectives measured by the exercises; Appendix B outlines the 15 mathematics content areas covered by the assessment; Appendix C lists the released mathematics exercises; and Appendix D provides information about the unreleased items. 1977.
ED 138 468 .....

09-MA-40 *Procedural Handbook: 1977-78 Mathematics Assessment*. A description of NAEP's procedures for objectives redevelopment through data collection and analysis to reporting the results, 1980.
ED 186 280 .....

13-MCS-40 *Procedural Handbook: 1981-82 Mathematics and Citizenship/Social Studies Assessments*. A description of NAEP's procedures for development of objectives and exercises, sampling, data collection, scoring, analysis, reporting, etc., 1983, 125 pages
.....

Johnson, Eugene. (1988). Chapter 10.2, Mathematics Data Analysis: Scaling of the Trend Data. Pp. 236–240 in Albert E. Beaton (ed.), *Expanding the New Design: The NAEP 1985–86 Technical Report* (No. 17-TR-20). Princeton, NJ: Educational Testing Service.
This chapter of the second ETS-authored NAEP technical report describes the procedures used to create a unidimensional IRT scale for what became the LTT mathematics scale.
ED 355 248 .....

Objectives:

Mathematics Objectives. A 41-page booklet describing the objectives used for exercises administered in the first mathematics assessment, 1970.
ED 063 140 .....

09-MA-10 *Mathematics Objectives, Second Assessment*. A 56-page booklet describing the objectives upon which the second assessment exercises were based, the procedures used for developing them and the persons involved in the process, 1978.
ED 156 439 .....

13-MA-10 *Mathematics Objectives, 1981-82 Assessment*. A 48-page booklet describing the evolution of math objectives for three different assessments, and the persons involved in formulating them, 1981.
ED 211 352 .....

17-M-10 *Math Objectives, 1985-86 Assessment*. A 26-page booklet offering background on previous math assessments and setting forth the development process and framework for the 1985-86 assessment, 1986
ED 273 682 [NCES has a copy] .....

Mathematics Objectives, 1990 Assessment. A booklet describing the development process and framework for the 1990 assessment, 1988
ED 309 030 [NCES has a copy] .....

Released Exercises:

Selected Supplemental Mathematics Exercises. Information on the contents, 1977.
ED 183 388 .....

09-MA-25 *The Second Assessment of Mathematics, 1977-78, Released Exercise Set*.
A 362-page loose-leaf set with national results for attitudinal items and national and selected group results for cognitive items. Contains 252 exercises. Suitable for reproduction. Objectives booklet included, 1979. (marginal legibility of original document)
ED 187 543 .....

13-MA-25 *The Third Assessment of Mathematics, 1981-82, Released Exercise Set.* A 287-page set with national results for attitudinal items, and national and modal grade results for cognitive items. Loose-leaf format suitable for reproduction. Objectives booklet included, 1983
.....

Early Reading Publications

The three earliest reading assessments (1970-71, 1974-75, and 1979-80) were administered using paced audio tapes to assure uniform assessment conditions. The paced tape method was dropped after 1983-84 for the reading/writing LTT, but not for mathematics/science trend assessments. In the early years, half the assessment items were released for use by the public after each administration.

Technical Materials:

11-RL-40 *Procedural Handbook: 1979-80 Reading and Literature Assessment*. A thorough description of the methods used in this assessment: redevelopment of objectives, formulation of exercises, sampling, data collection, scoring, analysis and reporting, 1981.
ED 210 300 .....

Mislevy, Robert, and Sheehan, Kathleen. (1987). Chapter 10.4, Trend Analysis. Pp. 361-390 in Albert E. Beaton (ed.), *Implementing the New Design: The NAEP 1983–84 Technical Report* (No. 15-TR-20). Princeton, NJ: Educational Testing Service. This chapter of the first ETS-authored NAEP technical report describes the procedures used to create a unidimensional IRT scale for what became the LTT reading scale.
ED 288 887 .....

Objectives:

02-R-10 *Reading Objectives*. A 34-page booklet describing the objectives used for exercises administered in the first reading assessment, 1970.
ED 041 010 [NCES has a copy] .....

06-R-10 *Reading Objectives, Second Assessment*. A 21-page booklet describing the objectives used for exercises administered in the second reading assessment, 1974.
ED 089 238 .....

11-RL-10 *Reading and Literature Objectives, 1979-80 Assessment*. A 28-page booklet describing the integrated objectives used for exercises administered in the third reading and second literature assessment, 1980.
ED 185 503 .....

15-RL-10 *Reading Objectives, 1983-84*. This booklet describes the objectives used for the fourth reading assessment, 1985
ED 243 086 .....

17/19-R-10 *Reading Objectives, 1986 and 1988 Assessments.* This booklet presents the reading objectives for the fifth and sixth assessments, 1987.
ED 287 876 [NCES has a copy] .....

Reading Objectives, 1990 Assessment. A booklet describing the development process and framework for the 1990 reading assessment, 1989
ED 307 598 [NCES has a copy] .....

Released Exercises:

02-R-20 *Reading: Released Exercises*, 1973, 424 pp.
ED 079 684 .....

02-R-25 *The First Assessment of Reading, 1970-71 Assessment, Released Exercise Set*, 1979. 341 pp.
ED 191 017 .....

11-RL-25 *Reading/Literature Released Exercise Set, 1979-80 Assessment*, April 1981. A loose-leaf set of 82 released exercises with documentation, national results and scoring guides. Suitable for reproduction, objectives booklet included. 351 pp.
ED 205 588 .....

11-RL-26 *Reading/Literature Released Exercise Set, 1979-80 Assessment, Supplement*, April 1981. Provides sample written responses to open-ended questions. 471 pp.
ED 205 589 .....