

Is It Time to Retire Long-Term Trend?

Louis M. Fabrizio, Ph.D.

North Carolina Department of Public Instruction

February 12, 2017

This is a reaction paper to Dr. Edward Haertel's white paper prepared for the National Assessment Governing Board (the Governing Board) on the "Future of NAEP Long Term Trend Assessments," dated December 9, 2016. The reaction paper will focus on arguments against keeping the National Assessment of Educational Progress (NAEP) Long-Term Trend (LTT) assessments. The thoughts expressed in the paper are solely the opinions of the author and not a reflection or representation of any official position of the North Carolina Department of Public Instruction or the North Carolina State Board of Education.

The comprehensiveness of Dr. Haertel's white paper is outstanding and he offers an engaging historical perspective covering the Long-Term Trend NAEP's roots and evolution, changes over time, challenges, and issues confronting the program since its beginning in the late 1960s.

As Dr. Haertel states:

The focus of this white paper is on the LTT assessments in reading and mathematics, as well as their relation to the main NAEP assessments in these same subject areas. It is intended as a starting point for a broad discussion of the LTT assessments, offering an overview of issues and options that might be explored in greater depth in future papers and symposia. Specifically, this white paper offers a brief history of the LTT and then addresses the following questions:

- What are some arguments for and against continuing the LTT component of NAEP in essentially its current form versus dropping it altogether?
- How might the LTT component instead be integrated (or blended) with main NAEP assessments?
- How might historical LTT data, main NAEP data, and bridge study data be integrated to make NAEP more useful for longitudinal research? (p.2)

Why have NAEP? It is the law, and the U.S. Congress and many others want to know how students are performing in the United States. The author in this paper will be focusing on the benefits of "main NAEP," which Haertel explains is what today refers to as the NAEP assessments from 1990 to the present that are based on assessment frameworks developed by the Governing Board and for which new scales and trend lines started. Results for main NAEP are available going back to 1990 for mathematics at the national level in grades 4, 8, and 12, and 1992 for reading at the national level in grades 4, 8, and 12, contrasted with the LTT (for which reporting results are available going back to 1973 in mathematics and 1971 in reading for 9-, 13-, and 17-year-old students).

National and state results. The main NAEP, in addition to reporting results at the national level, also reports results for all 50 states and the District of Columbia in grades 4 and 8. The state-by-state reporting of main NAEP allows each state to compare its performance with other states *as well as* the nation. Prior to the state-by-state reporting of main NAEP results, there was no way for states to benchmark their performance against other states (unless the states used the same assessments) and the nation. Even though each state and the District of Columbia are required to have assessment systems that can generate assessment results, it is main NAEP that provides that benchmarking capability for all states and it has done so for more than 25 years. This seems to the author to be a sufficient number of years to show trend analyses over time. This benchmarking capability, the author believes, is what state-level policymakers need to determine how well the students in their states are performing in grades 4 and 8 in reading and mathematics and what the U.S. Congress and other national policymakers need as well when looking at the national results.

Choices and recommendation? In essence, there are three choices for the Governing Board and the Commissioner of the National Center for Education Statistics (NCES) regarding the LTT: (1) continue with the LTT – it is in the federal law; (2) figure out a way to combine it in some way with main NAEP in the future; or (3) convince the U.S. Congress to remove the LTT requirement from the current legislation or future reauthorizing legislation. After reading Dr. Haertel’s paper and reviewing other reports, particularly Lawrence C. Stedman’s *The Long Term-Trend Assessment: A Review of Its Transformation, Use and Findings* (March 2009), this author believes that the last choice is the best. As a tweet, the author’s recommendation in 140 characters or less would be “Long-Term Trend... dump it!”

In the current political climate, if the Governing Board makes the recommendation to discontinue the LTT and focus all of its energies on main NAEP (and if the NCES Commissioner of Education Statistics agrees with the recommendation), the author does not believe that the U.S. Congress will make its ultimate decision based on many of the arguments included in Dr. Haertel’s paper. The author believes that the decision will be based on reasons that benefit the nation *and* the states and that is the perspective presented below.

Differences between the LTT and main NAEP. The author will not describe the differences between the two types of assessments in this paper, but NCES has a very concise comparison at https://nces.ed.gov/nationsreportcard/about/ltt_main_diff.aspx.

Who really cares about Long-Term Trend? The author contends that the stakeholder group most interested in the LTT is researchers. The author has never been in a meeting with a group of educators (outside of the Governing Board-related meetings) and heard anyone ask, “When are the results from the LTT being released? I can’t wait to see how the nation is doing compared to where the nation was in the early 1970s.” I am sure someone has asked that question but it is probably researchers who make the statement that they “can’t wait to get access to the LTT data to run some analyses.” This is not surprising because state-level educators and policymakers do not glean much information from the LTT results compared

with the benefits they receive from the national main NAEP results and the state-level main NAEP results (or LEA-level results for the 27 participating school districts in the Trial Urban District Assessments [TUDA]).

Benefits of main NAEP. Now, let us return to main NAEP. It is the state-by-state reporting of results at grades 4 and 8 that captures state-level policymakers' interest in how their state is doing in education. The U.S. Congress and others also benefit from main NAEP in that one can look at results at the national level in grades 4, 8, and 12 and compare student performance back to the early 1990s. Yes, that is about 20 years later than the LTT, but once one reviews all of the changes that occurred in the LTT during the time period of 1970 to 1990 and beyond, there are too many issues described by Haertel and Stedman to make the claim that one is truly comparing apples-to-apples with the reporting of results from the LTT during those 20 years and even beyond. Haertel states, "The assessment program envisioned by NAEP's founders looked very different from NAEP today." (p. 3) He goes on to describe numerous changes including, among others, how lists of objectives were initially developed, and how reporting would not be in terms of test scores but "in terms of estimated proportions of populations and subpopulations able to answer each exercise correctly." He then includes the following paragraph:

Almost from NAEP's very beginning, it had proven *necessary to modify* one after another of its original guiding principles. Among other *compromises*, inclusion of children with severe disabilities proved *prohibitively expensive*. Contractors submitted fewer very easy items and a greater proportion of multiple-choice questions than had been specified. Within a few years, *budgetary constraints* forced discontinuation of testing for out-of-school 17-year-olds and of young adults (Jones, 1996), although 16- to 25-year-olds were assessed under a separate grant in a 1985 "Young Adult Literacy Study." Testing in areas relying most heavily on performance assessment, including art and music, also *was judged too expensive to maintain*. The content areas of reading and literature were combined into one in 1979-80. (p. 3-4, emphasis added)

It is very clear in looking at the words in italics (*modify, compromise, prohibitively expensive, budgetary constraints, too expensive*) that many things have changed with the program and that will continue in the future. Haertel, later in his paper, makes the following statement:

Further significant changes loom as main NAEP continues the transition to a digital platform in 2017. These changes might be taken to imply that the LTT is increasingly irrelevant, or that the LTT is more important than ever. Regardless, the LTT assessment for 2016 has been twice postponed, first to 2020 and then to 2024. The LTT stands at a critical juncture. Its future is unclear. (p. 20)

Part of the reason why the author believes that the LTT should be discontinued, and that the current legislation that authorizes it should be amended to remove references to the LTT, is that the U.S. Congress also mandates (initially through the No Child Left Behind [NCLB] Act of

2001 and now through the Every Student Succeeds Act [ESSA] of 2015) that states and school districts must participate in main NAEP assessments in grades 4 and 8 and shall report NAEP results for the respective states on the state-level and school district-level report cards which must be made widely accessible on the respective state and local school district websites. This emphasis by the U.S. Congress on reporting main NAEP confirms to the author that it is main NAEP that warrants more attention than the LTT.

Haertel's arguments against maintaining LTT. Haertel makes the following arguments:

The principal arguments against maintaining the LTT in its current form are that it is expensive, that maintaining two trends is confusing, that performance on outdated content is no longer of interest to policymakers or other stakeholders, and that a range of changes in schooling, in assessment technology, and in society at large are rendering it irrelevant and possibly invalid. (p. 23)

Then Haertel addresses all four arguments. Regarding cost concerns, he states:

It is difficult to pursue this argument any further here, because that would require delving into the specifics of costs for the LTT and for competing NAEP priorities and then weighing these against perceived benefits, all within the constraints of historical commitments and statutory requirements. (p. 24)

The author supports Haertel's cost concern argument but believes that members of the Governing Board also support other priorities over the LTT and that is why the Governing Board has twice delayed the LTT administrations from the assessment schedule. Haertel then states that the argument that having two trend lines is confusing "seems weak" and that a related argument about redundancies between the LTT and main NAEP "also seems weak." (p. 26) However, the author agrees with Haertel on his statement that two trend lines being confusing is weak but the author still contends that the LTT should be discontinued in its current format and that doing bridge studies to determine if there are ways to combine the LTT with main NAEP should not be pursued by the Governing Board and NCES. In regard to Haertel's argument about outdated content not being a good argument against maintaining the LTT, the author disagrees. Thinking that LTT's "consistency" is going to be an arbitrator of changing curricular initiatives (like the Common Core State Standards [CCSS]) over the years seems like wishful thinking. If anything, the country seems to be moving away from anything that references the CCSS and the number of states participating in the assessment consortia is much lower than would have been predicted several years ago. Additionally, much of the confusion over the CCSS and the assessment consortia are the results of misinformation and misunderstandings. Finally, Haertel's reflection on the changes in schooling, technology, and society being used as an argument against maintaining the LTT will "require more careful consideration" (p. 25) is correct but the author believes that the value of LTT is diminished in light of the changes he references.

Stedman, in his paper, examines several pros and cons for keeping the LTT. He starts his “Recommendations for the Future” section with the following statements which align with the author’s arguments:

Overall, therefore, there is a compelling case for dropping the trend assessment and relying on the main assessment to generate trends. The dual-testing system is expensive and confusing. The main assessment already provides useful long-term trends. It provides greater coverage of the curriculum, more authentic testing, and is better grounded in contemporary pedagogy. (p.30)

The Governing Board and new directions. The Governing Board does an excellent job in carrying out its congressional mandates and, during the last decade, the Governing Board branched out and worked on the development of a new assessment called *Technology and Engineering Literacy* (TEL) which was administered for the first time to eighth graders in 2014. The expectation of the Governing Board is that it will be administered to 4th, 8th, and 12th graders in the future “if funding permits.” In response to the question, “How does this new assessment gauge levels of technology and engineering literacy?”, the answer in the frequently asked questions document is as follows:

An important and innovative design feature of the TEL assessment is its use of scenario-based tasks. These multimedia simulations use videos and interactive graphics to set up realistic situations. Then, students are asked a series of questions to demonstrate their knowledge and skills to solve problems within this practical context. For example, one TEL scenario-based task requires students to investigate why the well in a remote village is not working and how it can be fixed. In another, students are asked to troubleshoot and fix the habitat for a classroom iguana.

(<https://www.nagb.org/content/nagb/assets/documents/newsroom/naep-releases/naep-tel-webcast-may-17/naep-tel-faq-final.pdf>)

The author believes that including scenario-based tasks in assessments like the TEL assessment (and previously to a lesser extent in science assessments since 2009) is a great service to the states in showing the benefits of using such item types. It is through the research studies that the Governing Board commissions for these new endeavors that states become the recipients of the knowledge generated by them. Also, anecdotal feedback from students taking these scenario-based tasks has been very positive. Students actually enjoyed taking the assessment! Additional information on scenario-based assessments can be found at

<https://www.nagb.org/content/nagb/assets/documents/newsroom/naep-releases/naep-tel-webcast-may-17/naep-tel-scenario-based-one-pager.pdf> .

Another example of the Governing Board venturing in a new direction was whether it was possible to provide information from the main NAEP 12th grade assessments regarding academic preparedness of students for college, career, or the military. This resulted in over 30 studies undertaken to determine if the research findings would support reporting on the

academic preparedness for success in entry-level college credit-bearing courses, preparedness for success in entry-level job training programs, or preparedness for success in entering the military. The end result of the studies was the decision by the Governing Board to report only on academic preparedness for entry-level college credit-bearing courses but not on the job-training readiness or military readiness. Studies regarding the military were not pursued because the necessary approvals could not be secured for the studies. Regardless, the diligence on the part of the Governing Board to provide more relevance to the interpretation of main NAEP 12th grade test scores is noteworthy. The author believes that the Governing Board's decision to pursue these new initiatives provides more information of more value to policymakers than the continuance of expending resources on LTT that provides limited information to policymakers that cannot be determined by main NAEP (with the exception of the time span differences between LTT and main NAEP that has already been discussed).

What about merging the LTT with the main NAEP? While the author is only supporting the discontinuation of the LTT, the Governing Board also has the option to pursue research into the feasibility of merging the LTT with main NAEP. It seems to the author that there are so many issues with pursuing that possibility (age-based versus grade-based assessments, differing time limits for the assessments, paper and pencil versus computer-based assessments, to name a few) that there will be an increase in the number of bridge studies required to determine the feasibility. Bridge studies are usually required to confirm whether trends can be continued or how to make adjustments if there are significant differences in outcomes due to changes in the assessments or the administration procedures. Merging the LTT with main NAEP or continuing with the LTT in the future will only add to the number of new bridge studies required, especially in light of the Governing Board's and the Commissioner of Education Statistics' decision to move to all digital-based assessments starting in 2017, which the author supports and believes is the right thing to do. All of these bridge studies will take time and resources including additional students being tested in an environment where demands for fewer assessments are being voiced nationwide and the bridge studies will result in additional burdens on the participating schools and a loss of additional instructional time for students.

Future research. One of the benefits of discontinuing the LTT is that resources designated for the LTT could be better utilized in pursuing new ways to improve the current main NAEP and information generated from it. The author, however, does not support the pursuit of the research idea proposed by Haertel that could potentially be "used to project current trend lines [for main NAEP and LTT] backwards in time." (p. 33) Haertel even ends his paper with the statement, "The resulting data sets would be for research only, and would have to be treated with *considerable caution*." (p. 34, emphasis added) The author is not convinced that the backwards projections will be of any useful benefit to policymakers.

High on the author's list of research priorities that the Governing Board should consider are to determine how to improve the participation rates of private school students on the NAEP assessments. The legislation that authorizes NAEP includes language "to report on the

academic performance of students... in public *and private* (emphasis added) elementary schools and secondary schools.”

“In 2015, the school participation rates for private schools at both grades 4 and 8 did not meet the criteria so their results are not reportable.”

(https://www.nationsreportcard.gov/reading_math_2015/#reading/about%23footer?grade=4)

With the growing interest in discussions occurring nationwide on the issue of choice and programs like school vouchers for parents to use at schools of their choice (including private schools), it would be informative to have a more complete picture of student achievement nationwide to include data from private schools (beyond the results that sometimes occur for Catholic Schools). It appears that this has been an issue for a number of years. The author is aware of the NCES cautions of making comparisons between public and non-public schools (including private schools) and appreciates the information noted at

https://nces.ed.gov/nationsreportcard/reading/interpret_results.aspx that “Users are cautioned against interpreting NAEP results as implying causal relations. Inferences related to student group performance or to the effectiveness of public and nonpublic schools, for example, should take into consideration the many socioeconomic and educational factors that may also have an impact on performance.”

Conclusion and final caution. While the author is still convinced that the LTT should be discontinued and that attempts to merge the LTT with the main NAEP should not be pursued, there is one caution that needs to be mentioned if the Governing Board and the Commissioner of Education Statistics decide to seek a legislative solution to eliminate the LTT. Sometimes you get more changes than you asked for when you attempt to remove one requirement from a law. Let us hope that is not the case in convincing the U.S. Congress to remove the requirement for the LTT.

References

Haertel, E. H. (2016, December). *Future of NAEP Long-Term Trend Assessments* (Paper Commissioned for the National Assessment Governing Board).

Jones, L. V. (1996). A History of the National Assessment of Educational Progress and Some Questions About Its Future. *Educational Researcher*, 25(7), 15-22.

Stedman, L. C. (2009). *The NAEP Long-Term Trend Assessment: A Review of Its Transformation, Use, and Findings* (Paper Commissioned for the 20th Anniversary of the National Assessment Governing Board 1988–2008). Washington, DC: National Assessment Governing Board. (Downloaded from <https://www.nagb.org/content/nagb/assets/documents/who-we-are/20-anniversary/stedman-long-term-formatted.pdf>)