Transitioning to the *Next Generation Science Standards* in Maryland: The Case for Standalone Field Testing

Version 6.0

1 December 2016

Psychometric Services

Pearson

Table of Contents

Transitioning to the *Next Generation Science Standards* in Maryland: The Case for Standalone Field Testing

Version 6.0

1 December 2016

1. Overview

The ever-changing world of the 21st century demands increased proficiency in science, technology, engineering, and mathematics (STEM) for all. To address this need, the National Research Council (NRC), along with others, worked over several years to develop a framework for K–12 science education (National Research Council, 2012) and an accompanying set of standards, namely the *Next Generation Science Standards* (NGSS). The purpose for developing the NGSS at this time was stated as follows:

States have previously used the National Science Education Standards from the National Research Council (NRC) and Benchmarks for Science Literacy from the American Association for the Advancement of Science (AAAS) to guide the development of their current state science standards. While these two documents have proven to be both high quality and durable, they are around 15 years old. Needless to say, major advances have since taken place in the world of science and in our understanding of how students learn science effectively. The time is right to take a fresh look and develop Next Generation Science Standards. (NGSS Lead States, 2013)

In keeping with the aim of preparing their students in science literacy and STEM education in the best possible way, the state of Maryland is transitioning from their current set of science standards, Maryland State Curriculum: Science (*MSC- Science*) to the more rigorous *Next Generation Science Standards.* Implementing NGSS will provide Maryland students with the essential knowledge and understanding of science and engineering necessary to be college- and career-ready, and to discuss and engage in science-related issues as informed citizens.

The transition to NGSS has necessarily involved many different steps. These include communicating with different stakeholders regarding the adoption and implementation of NGSS; identifying and providing professional development around the new standards; making the shift in instructional practices, materials, and curriculum to the new standards; reviewing and updating data needs such as tracking science achievement; and aligning educational policy with the NGSS in areas such high school graduation requirements.

An additional step in Maryland's move to NGSS is designing and implementing a new science assessment, to be called the *Maryland Integrated Science Assessment* (MISA), to replace their current *Maryland School Assessment - Science* (MSA Science). Accomplishing this goal would involve developing, field testing, and reviewing items for MISA that are aligned to the NGSS and meet new test and item specifications.

In order to do this, Maryland is requesting from the US Department of Education (USDoE) an ESEA waiver for the standalone field testing of the *Maryland Integrated Science Assessment* (MISA) for Grades 5 and 8 with no fault to students for 2016-2017. In discussions regarding this waiver request, the USDoE has proposed to Maryland, that rather than implementing a standalone field test for MISA, the state should instead administer MSA Science with the embedded field testing of MISA items and continue to use the same MSA Science scale for reporting student results. We argue that this suggested approach is inappropriate for developing the MISA system due to

- important differences in both the Maryland Science Curriculum Standards and NGSS;
- differences in the measurement targets, test designs, items, and test administration conditions that are used on MSA Science and that which will be used on MISA;
- the need for the field test design to meet the requirements of the MISA development from a content development perspective;
- the negative impact on the validity of MSA Science test scores in the presence of an embedded MISA field test; and
- the needs of Maryland students with respect to "opportunity to learn" issues with testing in transitioning to the NGSS-based system.

The following sections of this paper expand on these points.

2. A Comparison between the Current and Proposed Maryland Science Assessment Systems

2.1. Maryland's Current Science Standards Compared with the Next Generation Science Standards

As described in the *2014-2015 MSA Science Annual Technical Report* (Pearson, 2015), the Maryland State Curriculum (MSC) serves two purposes. First, it defines what students should know and be able to do, starting with broad measurable statements and then focusing on more specific indicator statements. Second, it serves as the tool to align the Maryland Content Standards and the Maryland Assessment Program.

The MSC is available and widely disseminated to Maryland educational stakeholders, including teachers, central office staff, students, parents and other stakeholders. In order to ensure that MSDE is in accordance with the federal law that requires states to align their tests to their content standards, the MSC served as the guiding document for test development and design (Maryland State Department of Education, n.d.). The following table provides a high-level overview of the MSC.

Table 1. Overview of the *Maryland State Curriculum Standards for Science*

Standard	Description
Standard 1.0 **Skills and Processes**	Students will demonstrate the thinking and acting inherent in the practice of science.
Standard 2.0 **Earth/Space Science**	Students will use scientific skills and processes to explain the chemical and physical interactions (i.e., natural forces and cycles, transfer of energy) of the environment, Earth, and the universe that occur over time.
Standard 3.0 **Life Science**	The students will use scientific skills and processes to explain the dynamic nature of living things, their interactions, and the results from the interactions that occur over time.
Standard 4.0 **Chemistry**	Students will use scientific skills and processes to explain the composition, structure, and interactions of matter in order to support the predictability of structure and energy transformations.
Standard 5.0 **Physics**	Students will use scientific skills and processes to explain the interactions of matter and energy and the energy transformations that occur
Standard 6.0 **Environmental Science**	Students will use scientific skills and processes to explain the interactions of environmental factors (living and non-living) and analyze their impact from a local to a global perspective.

Essentially, the MSC provides for two kinds of standards, namely, *skills and process* (Standard 1) vs. *knowledge in specific content domains* (Standard 2 through 6). The skills and processes addressed in Standard 1 include constructing knowledge by gathering and analysing data and designing and carrying out scientific investigations; applying evidence and reasoning to develop explanations; communicating scientific information; understanding and using such aspects of technology as design constraints, systems thinking, and models; and the history of science.

As stated above, the NGSS were based on the NRC's A Framework for K-12 Science Education (National Research Council, 2012). (Table 2). As described by the NRC, the framework

…structures science learning around three dimensions: the *practices* through which scientists and engineers do their work; the key *crosscutting concepts* that link the science disciplines; and the *core ideas* of the disciplines of life sciences, physical sciences, earth and space sciences, and engineering and technology. (National Research Council, 2014, p. 1)

An overview of this framework is shown in Table 2 below.

Table 2. Overview of *A Framework for K-12 Science Education*

Dimension	Description
Dimension 1: Practices	Practices describe behaviors that scientists engage in as they investigate and build models and theories about the natural world and the key set of engineering practices that engineers use as they design and build models and systems.
Dimension 2: Crosscutting Concepts	Crosscutting concepts describe concepts that bridge disciplinary boundaries, having explanatory value throughout much of science and engineering. These crosscutting concepts have application across all domains of science. They are a way of linking the different domains of science.
Dimension 3: Core Disciplinary Ideas	Disciplinary core ideas focus science curriculum, instruction, and assessments on the most important aspects of science. Disciplinary ideas are grouped in four major domains: physical sciences; the life sciences; the earth and space sciences; and engineering, technology and applications of science.

The NRC further argues that these standards

…should be interwoven in every aspect of science education, most critically, curriculum, instruction, and assessment. The framework emphasizes the importance of the connections among the disciplinary core ideas, such as using understandings about chemical interactions from physical science to explain biological phenomena. (National Research Council, 2014, p. 1)

The NRC uses the term *practices* instead of a term like "skills" to emphasize that engaging in scientific investigation requires not only skill but also knowledge that is specific to each practice. Part of the NRC's intent is to better explain and extend what is meant by "inquiry" in science and the range of cognitive, social, and physical practices that it requires. These practices include developing and using models, planning and carrying out investigations, analysing and interpreting data, engaging in arguments from evidence, and communicating evidence.

The *crosscutting concepts* provide an organizational schema for interrelating knowledge from various science fields into a coherent and scientifically-based view of the world. They include: patterns; cause and effect; scale, proportion and quantity; systems and system models; energy and matter: flows, cycles, and conservation; structure and function; stability and change.

The *disciplinary core ideas* represent the most important aspects of the domains of the physical sciences; the life sciences; the earth and space sciences; and engineering, technology and applications of science. They were chosen as to have broad importance across multiple sciences or be key organizing concepts of a single discipline; be key to understanding more complex ideas and problems; and be teachable and learnable over multiple grades at increasing levels of depth and sophistication.

In comparing the MSA- for Science and NGSS standards, it is clear that both make use of “process/practice-related” and “content/domain-related” standards. However, there are important differences.

The NGSS can be thought of as “three-dimensional standards” in that in addition to the process/practice-related and the content/domain-related standards there are the cross-cutting concepts that link key ideas across content domains and grade levels. All of the NGSS *performance expectations* – what students should be able to do in order to demonstrate that they have met the standard – are created by incorporating all three of the dimensions. This cross-disciplinary dimension is missing from the MSA-for Science.

Additionally, the NGSS provides explicit connections to the Common Core State Standards in Mathematics and English Language Arts for each of its performance expectations. This allows, for example, a NGSS performance expectation in physical sciences to be tied to a CCSS Mathematics standard needed as a prerequisite skill. These kinds of cross-subject linkages are not called out in the Maryland State Curriculum: Science.

Another feature of the NGSS that is different from Maryland State Curriculum: Science is the highlighting of engineering, technology and the applications of science as a separate core discipline. This discipline was added to the NGSS

…because of the interconnections between science and engineering and because there is some evidence that engaging in engineering design can help to leverage student learning in science. …The goal is to strengthen science education by helping students understand the similarities and differences between science and engineering by making the connections between the two fields explicit and by providing all students with an introduction to engineering. (National Research Council, 2014)

Finally, the NGSS performance expectations are worded in such a way as to require far more challenging ways for students to demonstrate their understandings and skills. Rather than being asked to "recognize," "identify," or "realize," the NGSS requires students to apply higher-order thinking skills in contexts that make use of more than one of the NGSS dimensions: The performance expectations are far more rigorous than the indicators and objectives found in the Maryland State Curriculum: Science.

These differences will require major changes in how Maryland assesses science in transitioning from the MSA Science to MISA, and are described next.

2.2. The MSA Science Assessment Compared with the MISA Assessment

Aligning to NGSS performance expectations leads to different kinds of the measurement targets, test designs, items, and test administration conditions for the MISA versus what is currently used for MSA Science.

2.3. The MISA assessment targets are more complex than those for MSA Science

The assessment targets for MSA Science are the six single standards of the Maryland State Curriculum: Science. Although there are both process-related and content-related standards that are being assessed, each item is designed to measure a single assessment limit. In contrast to this "one-dimensional" approach, MISA assessment targets will be focused on gathering evidence around the NGSS performance expectations, and will use items incorporating at least two of the three NGSS dimensions in measuring them.

2.4. The test design and items for MISA are substantially different from that of MSA Science

The MSA Science test design[1] consists of two core forms of items that are administered to all students, in both paper and online modes. The test consists of selected response (SR) and brief constructed response (BCR) items and most of the items are bundled together in what are called *Lab Sets*, i.e., that is a stimulus with a set of two to three related items. However, there are individual items on the test that are separate from any Lab Set.

The MISA is quite different in that it uses a matrixed test design (Childs & Jaciw, 2003) with additional item types. Besides selected-response items, there will also be composite items (e.g., SR with justification), technology enhanced items, and constructed response items. The technology enhanced items include graphing, drag-and-drop, hotspot, simulations, and other types. The constructed response items will be more extensive, and replace the BCR items that were used on MSA Science. The MISA design will also move to using exclusively, phenomenon-based sets of items with approximately six items per set. Items will no longer be allowed to be standalone.

2.5. The items on MISA will be far more rigorous than the items on MSA Science

The more complex nature of assessment targets for MISA will necessitate in turn items that are adequate for measuring those targets.

As a part of the development process for MISA, an analysis of the MSA Science item bank was completed to check for alignment of its items to the NGSS. This was done in order to determine if MSA Science items could be adequately used to measure NGSS performance expectations as operational or field test items for MISA. Science assessment specialists reviewed each item and then compared the item to the NGSS standards and supporting concepts to determine if the item was partially or wholly aligned to them. The analysis showed that these items did not align.

The reasons for the poor alignment of MSA Science items to NGSS have been alluded to earlier. The NGSS standards are multi-dimensional and ask students to apply their knowledge, demonstrate how to reason and/or argue through scientific concepts and processes, and construct and develop models to demonstrate their understanding. The MSA-Science standards are much more specific: They were designed to assess student knowledge and required very little application and use of scientific principles. The items written to these standards were often standalone, and asked students to identify or recall bits of knowledge.

---

[1] The number of forms listed is for the operational portion of MSA Science not field test forms.

3. A Standalone MISA Field Test Design is More Appropriate than an Embedded Field Test Design

The previous section detailed the significant differences between the current MSA Science program and the future MISA program, and the necessity for extensive development and field testing. Here we discuss why an embedded field test of MISA item sets within the MSA Science is inappropriate and a standalone MISA field test design is more appropriate with respect to addressing MSDE's development requirements, dealing with psychometric concerns related to context effects and student motivation, and providing Maryland students with the "opportunity to learn" in the new Next Generation Science system.

3.1. A Standalone Field Test Meets Maryland's MISA Development Needs Better than an Embedded Field Test

Currently, the MISA item sets have undergone review for content alignment and appropriateness. Those item sets that have been approved by MSDE will be census field tested in March 2017 with Maryland students to obtain data on item difficulty, discrimination, and so forth[2]. These data will inform test developers on the performance characteristics of the item and item sets, and in building the final operational forms. In order to address Maryland's development needs, the field test will need to provide enough items or items sets to allow the creation of multiple test forms in subsequent assessment administrations.

When using an embedded field test on an operational administration, there is a limitation on the number of items or item sets that can administered. If too many items are added to the operational test then this will fatigue students due to over-extended testing time. This limitation may be acceptable when a large item bank exists and the field test is being used essentially to "refresh" the test. However, in the case of MISA, we have shown that such an item bank does not exist and there is a need to develop a large number of NGSS items. An embedded design will not have enough places to field test the number of phenomena and associated item needed to populate the MISA item bank. Since *all* of item sets on the standalone field test will be field test sets, this design will provide a larger number of items for initial operational testing and build up the item bank for use in future MISA administrations. This is important as a large initial item bank can reduce the number of places needed for embedded field test items in subsequent administrations. This in turn can reduce student testing time and partially address the concern about taking time away from instruction.

[2] The schedule for this and other MISA development activities is in the Appendix.

3.2. The Negative Impact of Embedded MISA Field Testing on MSA Science Test Scores

The test scores on MSA Science are intended to measure the construct of science achievement with respect to content from the Maryland State Curriculum Standards for Science, while MISA is designed to measure content aligned to the NGSS in a very different way. This shows up in the differences between the MSA Science and MISA items.

Due to operational testing constraints, it is most likely that MISA field test items will need to be embedded in MSA Science at the end of a testing period in a separate section, which would be obvious to test takers. The embedded design would result in differential motivation, with more on the operational MSA Science items and less on MISA field test items. This would likely produce context effects that occur when "item parameters are influenced by the sequencing of the items or by the characteristics of other items in the test," and this is particularly important when field testing and selecting items, or when calibrating items for an item pool (Yen, 1980, p. 297). A standalone field test design would provide uniform (though not necessary more) motivation for all MISA item sets and not be subject to the same kinds of context effects as an embedded design.

It should be noted that while a standalone MISA field test would provide better data for item parameter estimates, these estimates would only be used as part of the process of examining the performance characteristics of item sets in support of creating the final MISA operational test forms; they will not be used for producing student score reports in 2017. Since the science curriculum has transitioned to NGSS, and the 2017 assessment will have no stakes for either students or teachers. If teachers and students know that there will be no student reports, test taking motivation will likely be lower, but since these data will not to be used for operational calibration, this will provide less of a construct-irrelevant problem than it normally would.

The item parameter estimates that will be used in creating the MISA scale will be based on data from the first operational MISA administration in May 2018. This should reduce as much as possible any concerns regarding student motivation, and provide the crucial impact data that will be needed by panelists for the standard setting studies planned for June 2018. The new MISA scale will be used to start a new trend line for science achievement in Maryland going forward.

3.3. A Standalone Field Test Better Prepares Maryland Students for Next Generation Science

Content development and psychometric concerns aside, a standalone MISA field test supports Maryland students' *opportunity to learn* in the transition to new Next Generation Science system. The *Standards* define opportunity to learn as the "extent to which individuals have had exposure to instruction or knowledge that affords them the opportunity to learn the content and skills targeted by the test… (American Educational Research Association, American Psychological Association, National Council on Measurement in Education, 2014, p. 56).

A standalone field test is a much better way of preparing Maryland students to engage in assessments that will be quite different from the ones they have had in the past. Although they are now being taught from a curriculum that is aligned to the NGSS rather than the MSA Science, students still need the chance to adjust to a much longer test administration, to try out new item types, be exposed to more complex, phenomenon-based item sets, and practice dealing with a test that entails more rigorous questions and demands more complex kinds of reasoning.

MSDE's current plan for the standalone field test is for it to follow the operational design of MISA with respect to the number of test periods, item types, and so forth. Students will have the opportunity to take a test that is aligned to the NGSS curriculum that they are using rather than previous MSA-Science curriculum. With a standalone field test, they will be able to answer item sets in their proper context as part of a test with "like" components, rather than as isolated sets of items that do not relate to the content or rigor items around them if they had taken an embedded field test. Finally, they – and their teachers – will be able to better understand what is demanded by the NGSS performance expectations and what they will need to do to succeed on MISA.

4. Summary and Conclusion

In this paper we have presented Maryland's case in using a standalone design for field testing MISA item sets, rather than the USDoE's suggestion of embedding MISA items on the MSA Science test.

We based our argument on the important differences between the Maryland Science Curriculum Standards and NGSS, and in the measurement targets, test designs, items, and test administration conditions that are used on MSA Science and that which will be used on MISA. We discussed the impact of those differences on the need for a field test design to meet the requirements of MISA content development; deal with the psychometric issues of construct irrelevant variance, context effects, and student motivation; and address the important opportunity-to-learn issues for Maryland.

Our conclusion based on this line of reasoning is that a standalone field test design, rather than an embedded design, best meets the needs of MSDE, Maryland students, and their teachers in transitioning to an NGSS-based assessment system.

For these reasons, we request that the US Department of Education (USDoE) grant Maryland an ESEA waiver for the standalone field testing of the *Maryland Integrated Science Assessment* for Grades 5 and 8 with no fault to students for 2016-2017.

Appendix: Schedule for Future MISA Development Activities

Schedule for Future MISA Development Activities

Dates	2017 MISA Standalone Field Test Administration	2018 Operational MISA Administration
Oct 2016 - Feb 2017	MSDE review and approval of field test forms	
Dec 2016		Begin item writer training for operational test
Jan - Feb 2017		Professional development meeting for MD educators for future item writing
Mar 2017	Field test window administration; Field test rangefinding	
Jun - Jul 2017		Operational item committee review
Jul - Aug 2017	Field test scoring window	MD Educator item writing for 2019 administration
Aug - Sep 2017	Field test data analysis	
Sep 2017		Operational test construction; Accommodated paper form finalized
Oct 2017		Preliminary test maps baselined
Oct 2017 - Feb 2018		MSDE review and approval of operational forms
Mar 2018		Operational test administration window
Apr - May 2018		Operational scoring window
Jun 2018		Operational rangefinding; Standard setting
Jul - Aug 2018		Embedded field test scoring
Nov 2018		State/district/student reports delivered

References

American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Childs, R. A., & Jaciw, A. P. (2003). Matrix sampling of items in large-scale assessment. *Practical Assessment, Research & Evaluation, 8(16)*, 1-12.

Maryland State Department of Education. (n.d.). *Standards: PK-8*. Retrieved from School Improvement in Maryland: http://mdk12.msde.maryland.gov/assessments/vsc/

Maryland State Department of Education. (n.d.). Teaching and learning: Science - Next Generation Science Standards for Today's Students and Tomorrow's Workforce. Retrieved from School Improvement in Maryland.

National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas.* Board on Science Education, Division of Behavioral and Social Sciences and Education, Committee on a Conceptual Framework for New K-12 Science Education Standards. Washington, DC: The National Academies Press.

National Research Council. (2014). *Developing assessments for the Next Generation Science Standards.* Board on Testing and Assessment and Board on Science Education, J.W. Pellegrino, M.R. Wilson, J.A. Koenig, and A.S. Beatty, Editors. Division of Behavioral and Social Sciences and Education, Committee on Developing Assessments of Science Proficiency in K-12. Washington, DC: The National Academies Press.

NGSS Lead States. (2013). *Development Overview*. Retrieved from Next Generation Science Standards: http://www.nextgenscience.org/development-overview

Pearson. (2015). 2014-2015 MSA Science Annual Technical Report. San Antonio, TX: Pearson.

Yen, W. M. (1980). The extent, causes, and importance of context effects on item parameters for two latent trait models. *Journal of Educational Measurement, 17*, 297-311.