A Report Describing the Work and Recommendations of the Illinois Science Assessment Steering Committee for the Illinois State Board of Education

For consideration by the Board on April 17, 2019

Prepared by the
Division of Assessment and Accountability at the Illinois State Board of Education

&

Reviewed by the
Members of the Illinois Science Assessment Steering Committee

Introduction

History of the ISA and Context for the Work of the Committee

Beginning in school year 2015-16, Illinois has given the Illinois Science Assessment (ISA) to students in grades 5 and 8 and to students enrolled in a high school biology course who do not already have a full-year biology credit. The development of that ISA was predicated on making choices to meet compliance requirements as minimally as possible, in a context of limited resources. This has resulted in delays and data that provides only minimal information to parents, schools, and districts.

The U.S. Department of Education (ED) cited the State of Illinois on April 20, 2015, for failure to administer a science assessment in 2015, as required under the No Child Left Behind Act. Illinois was placed on "high risk" status with respect to the receipt of our federal Title I Part A dollars. The Illinois State Board of Education entered into an item-sharing agreement with the Office of the State Superintendent in the District of Columbia in order to bring Illinois into compliance.

ED did not approve this plan until fall of the 2015-2016 school year because the administration of an end-of-course biology assessment in high school would not fully meet the federal testing requirements. Approval was granted in light of the Illinois budget crisis; however, the lack of a state budget at the time significantly delayed ISBE's ability to enter into contracts with vendors to create and administer the ISA.

Work on the ISA began November 1, 2015, which pushed the start of the 2016 assessment window to May 1, 2016. Once administration was underway, ISBE began working on a solution for how to score the open-response items on the tests that would provide value to and build the capacity of the field. These efforts were also hindered by the lack of a state budget, and a final agreement for scoring was not executed until January 31, 2017. Recruitment of educators and science experts began immediately, and training was conducted in February of 2017. Scoring for the 2016 assessment began on February 23, 2017, and was finalized on July 25, 2017.

After scoring was complete, a process of data validation and quality review was conducted in collaboration with the field. In addition, a standard-setting and other psychometric work on the results was conducted by both ISBE and Southern Illinois University Carbondale (SIUC) to confirm the scoring and results. This work was ongoing through the end of 2017. Only after these steps were complete could data be aggregated and reported.

Data from the 2016 ISA administration were finally reported on January 17, 2018, a year and a half after they would normally have been released. Results from the 2017 ISA administration were reported one month later, on February 14, 2018, delayed approximately six months. The 2018 ISA results were released to schools on time in July 2018 and were reported officially along with all other 2018 assessment data as a part of the 2018 Illinois Report Card.

The process for administering, scoring, and reporting the ISA were improving, but the content of the test itself remained stagnant until 2018. Meetings were held in spring 2018 with various stakeholders to envision how to collaboratively redesign the test from the ground up. This work was postponed for the remainder of 2018 when agency resources were diverted in May to the preliminary and official implementation of the state's new accountability system and the issuance of summative designations in June and October of 2018. During July-September of 2018 SIUC coordinated multiple workshops to review and select NGSS aligned items from the ISAT and PSAE programs to augment the ISA for the 2019 testing. ISBE replaced roughly 50% of the 2018 items

with these workshop test questions. Work officially began in January 2019 with the convening of an Illinois Science Assessment Steering Committee (ISASC).

The ISASC was asked to assist in designing an assessment blueprint that was more fully aligned to the Illinois Learning Standards for Science (ILS-S), based on the Next Generation Science Standards (NGSS)[1]. A major difference between the NGSS and previous science standards is their "three-dimensional" nature. The standards represent and require of students thoughtful and deliberate integration of three distinct dimensions: Scientific and Engineering Practices (SEPs), Disciplinary Core Ideas (DCIs), and Crosscutting Concepts (CCCs). In this way, the NGSS emphasize that science is not just a series of isolated facts. Instead, learning is structured so that student experience science more as an interrelated world of inquiry and phenomena rather than a static set of science disciplines. The NGSS represent a fundamental shift in science education and require a different approach to teaching and assessing science than has been done in the past.

The work of the ISASC represented in this memo is incredibly timely, as Illinois received its feedback from the federal peer review process on January 3, 2019. As expected, the administration of the 2015-18 ISA did not meet federal assessment requirements, and ED has directed that "ISBE must have a new or substantially revised general science assessment in place to begin administering in the 2019-20 school year." The work of the ISASC has been structured to be directly responsive to the federal assessment peer review requirements, while still reflecting the beliefs and desires of the field as to what a high-quality science assessment should measure, what the student experience of the test should be like, and what claims or sub-claims the test should be able to support about student knowledge and skills. In our response to ED describing our plans for coming into compliance, ISBE expressed our intent to "request a waiver from the Department for the 2019-20 assessment in order to conduct a census field test of multiple forms worth of newly constructed items to fill the ISASC-recommended test blueprint."

Structure and Process of the Work

Invitations to join the committee were sent to all those who were invited to and participants of the early stakeholder meetings held in 2018 and to all individuals who had directly contacted either ISBE or our collaborators at SIUC expressing a desire to contribute to the work. Additionally, invitations were sent to the two teachers unions to nominate members. General public information about the committee was included in the Superintendent's Weekly Message, encouraging all who wished to participate to contact one of the three committee co-chairs from either ISBE or SIUC. Participation on the committee was capped at a total of 50 members, plus 3 co-chairs, 4 psychometricians and 2 SIUC experts, to ensure grade-band working groups of at least 10, but no more than 20, people. The intent was, if interest exceeded capacity, to prioritize requests to ensure diversity of perspective and even coverage of the grade bands and geographic regions, but the cap was never exceeded. Indeed, additional teachers and subject matter experts who represented key grade levels or student groups were recruited by existing committee members after the first meeting to ensure those perspectives were more robustly represented.

A series of four all-day meetings were held in Springfield on January 28, February 15, March 9, and March 29, 2019. Significant work was conducted by committee members in preparation for and after each of the meetings. The fourth meeting was added after the committee was unable to reach full consensus at the third (originally planned to be final) meeting. Attendance in Springfield was

[1] Next Generation Science Standards. An Overview for Principals. Accessed 4/1/2019 at: https://www.nextgenscience.org/sites/default/files/resource/files/NGSS%20Overview%20for%20Principals_0.pdf?

strongly encouraged, but virtual attendance was accommodated through online meeting software for those who required it. Care was taken to ensure that those who were participating online had opportunities to interact with the group in Springfield. Participants were compensated for their time and travel as applicable, as not all participants were able to attend all meetings or all portions of the meetings that they did attend.

The committee was tasked with answering three central questions, and a broad agenda was put forth for the work to arrive at those answers. The questions posed to the committee were:

1. Within the standards, what are the most critical elements to assess – with a large-scale accountability assessment and through other assessments? (In other words, what is most important for students to know and be able to do?)

2. What information should the results of the accountability assessment provide – to parents, teachers, administrators, and the state? (In other words, what claims are we trying to make about what students know and are able to do?)

3. What should the student experience of the test be like, and how does that experience relate to the classroom experience?

The committee co-chairs expressed a commitment in the very first meeting to an open, dialogic, and messy process to reach consensus. There was a general process envisioned for how the work might develop and build over time that would incorporate both whole-group and grade-band working groups and sessions, but each portion of the process was allowed to take as much time as needed to reach consensus. The report describes the answers developed by the committee to each of the three questions. The blueprint itself is the answer to the first question. The sections of this report on sub-claim reporting are the committee's response to the second question, while the section on item specifications and test structure is the committee's response to the third question.

The first meeting set the foundation for the work, providing members with a brief history of the old ISA program and an overview of the resources and expertise available to build the new assessment. There was a discussion of the vision, mission, and goals for the assessment and a discussion of the three core questions. Despite best intentions, the committee did not delve into issues specific to grade bands in the first meeting.

Participants were asked to prepare for the second meeting by reviewing readings suggested by committee members and watching presentations developed by the SIUC advisory staff on "Understanding Psychometrics" and "Item Interaction Types." A variety of exemplar science blueprints from other states was reviewed in the second meeting and a plus-minus-delta analysis was conducted as a whole group. The second half of the day was spent in grade-band groups attempting to generate a specific response to the first central question as it pertained to the standards in that grade band. During this meeting, committee members first introduced a 2015 publication by McNeill, Katsh-Singer, and Pelletier on assessing science practices[2] that serves as the foundation for one of the Science and Engineering Practices (SEP) groupings. Consensus was

[2] McNeill, Katherine L., Rebecca Katsh-Singer, and Pam Pelletier. "Assessing science practices: Moving your class along a continuum." *Science Scope* 39.4 (2015): 21. Accessed February 15, 2019 at: https://search.proquest.com/openview/d3359cde04378155c6d06302ce968e0d/1?pq-origsite=gscholar&cbl=36017

also reached at the high school level that the test would be a grade 11 test as opposed to a course or subject-based test.

Members were asked to prepare for the third meeting by reviewing a preliminary sample blueprint showing the Domain/SEP sets structure aligned to the Illinois Learning Standards for English language arts (ELA) and math that were developed in response to the committee's discussions in meeting 2. They were also asked to share their thoughts or reflections in writing. Members discussed the preliminary sample blueprint as a whole group during the third meeting and studied charts depicting the distribution of the performance expectations across the Domain/SEP matrix to inform recommendations about the percentage of items within each cell the test should cover. However, several members of the grade-band groups became interested if the distribution of the performance expectations across the McNeill SEP sets was similar and conducted that analysis on site during the meeting. The merits and potential consequences of the different SEP groupings were deeply debated when the three grade bands were brought back together, and the groups were unable to reach full consensus.

A fourth meeting was called at the conclusion of the third meeting to tackle two remaining tasks. The first task was to provide members of the high school grade-band group with additional information or simulations that would help them reach consensus with the elementary and middle school grade-band groups. The second task was to finalize the SEP grouping, approve names for each set, and define the meaning of the sets for the field, so that when sub-scores were given, there would be meaning to the results.

A second set of sample blueprints was developed in advance of the fourth meeting for committee members to respond to. Additionally, members were encouraged to articulate their reasons for supporting one grouping of SEPs over the other in writing and to provide additional supporting evidence for their perspectives, if possible. In response, one member solicited feedback from some of the original NGSS authors, while others shared input from the perspective of ED at the time the framework was being developed. This fourth meeting was successful in bringing the groups to consensus on the distribution of the items across the Domains and SEPs, with an equal distribution across the three Domains and the eight SEPs. However, after extensive debate, the committee could not come to consensus on a particular grouping of SEPs, in large part because there is no distribution of performance expectations across the Domain/SEP matrix that fully matches the consensus blueprint. In the end, the committee felt that it would be best to see how feasible it was to develop items representing the various combinations of Domains and SEPs as part of the phenomena-based item clusters and to further see how well the items survived bias and sensitivity review and field testing. The committee proposed to document both methods of grouping SEPs for the Board, but holding a final recommendation until there was potential empirical evidence supporting stronger validity, reliability, or quality of items using one grouping of SEPs over the other. This compromise was accepted.

Assessing the Illinois Learning Standards for Science

In 2014, Illinois adopted the Illinois Learning Standards for science, based on the NGSS. The standards envision science performance expectations as the ability to employ disciplinary core ideas (DCI), scientific and engineering practices (SEP), and crosscutting concepts (CCC) to solve scientific problems.

"The purpose of an assessment determines priorities, and the context of use imposes constraints on the design (Pellegrino 13)[3]." For this reason, in its *Operational Best Practices for Statewide Large-Scale Assessment Programs*[4], the Council of Chief State School Officers recommends that "an assessment cycle should begin with clearly defined test specification documents that enable the development of test items (21)." One of these is a test blueprint, which defines the coverage of curriculum or standards, often communicated as a number or percentage of test items per standard, benchmark, or domain.

Given the breadth, complexity, and three-dimensional nature of the standards, a primary challenge of developing a large-scale assessment is that it is not feasible to reliably assess all of the standards for a single grade or grade-band in a single assessment (Pellegrino 2). An aligned system of local formative assessments, conceptually aligned to large-scale summative accountability assessments, is needed to truly measure students' scientific competency. Test blueprints represent a series of intentional decisions and negotiations about what to assess on the summative assessment and how to do it. This reflects the expertise and values of practitioners about what students should know and be able to do in science and engineering. The blueprint presented represents the collective vision and understanding of ISASC. The committee was composed of 59 members, representing nationally recognized science and assessment experts, original NGSS contributors, statewide school and district administrators, curriculum and instruction leaders, teachers, higher education, ISBE, and not-for-profit organizations related to science and the NGSS.

The blueprint is designed to represent the full range of the Illinois Learning Standards for science and the broader testing process is intended to communicate to the field the importance of all science standards for all students. The blueprint achieves this representation by sampling the most important aspects of the performance expectations, across all domains at the relevant grade levels or grade band. Individual test forms will reflect the standards, and there is a plan for assessing the full range of standards through multiple forms within or across tested years. The blueprint will make decisions about sampling transparent and clear to teachers, students, and parents.

After much deliberation, the committee reached consensus that a series of interrelated blueprints better represented the three-dimensionality of the standards, emphasizing the cross-cutting focus on science disciplines and practices within the constraints of a single, summative test. Recognizing the role of the summative assessment to inform policy, the committee expressed a policy goal to improve science learning throughout the state for the greatest number of students in the most significant ways. The Committee recognized that to achieve that goal the test needs to provide administrators and teachers useful information about student science competency, but results alone are not enough. To truly be impactful, assessment results must be connected to high-quality local performance assessments through a system of statewide professional development and supports. That aligned and integrated system, similar in form and scope to that identified by Pellegrino (2), then drives professional learning and statewide systemic reform at all levels.

High-quality science instruction[5] provided every single year in every single grade -- kindergarten through graduation -- is necessary for students to graduate scientifically literate and ready for

---

[3] Pellegrino, Jim W. 21st Century Science Assessment: The Future Is Now. (SRI Education White Paper). Menlo Park, CA: SRI International. 2016.
[4] Eignor, Daniel R. *Operational best practices for statewide large-scale assessment programs*. Council of Chief State School Officers and the Association of Test Publishers, 2010.
[5] NSTA Position Statement Elementary Science Education. Accessed 4/1/2019 at: http://static.nsta.org/pdfs/PositionStatement_Elementary.pdf

careers and postsecondary education in science and engineering. The blueprint is intended to provide a score to individual students and parents that represents their performance on all dimensions of the range of NGSS standards from the elementary grade band, middle school grade band, and high school grade band. The blueprint is designed to provide -- at the school or district level, depending on the number of students tested -- sub-scores that represent student performance in the aggregate in six areas:

- The three Domains are:
  - Physical Science
  - Life Science
  - Earth and Space Science
- The eight Science and Engineering Practices are:
  - Asking Questions and Defining Problems
  - Developing and Using Models
  - Planning and Carrying Out Investigation
  - Analyzing and Interpreting Data
  - Using Mathematical and Computational Thinking
  - Constructing Explanations and Designing Solutions
  - Engaging in Argument from Evidence
  - Obtaining, Evaluating, and Communication of Information

The committee agreed that it would be ideal to report out on each SEP individually; however, the number of test questions needed to achieve a sufficient level of reliability makes this level of reporting impractical at this time. Over time, given sufficient development of the item bank, designs that would support reporting on SEPs individually will be revisited. The proposed SEP sets will be discussed in greater detail below in the section on aspects of the test to be determined by field testing.

Below are three ways of representing the same blueprint. Table 3 is a two-dimensional representation, and best reflects the consensus and understanding of the committee, however, there were concerns that the over specification of items in the cross tabbed cells might limit the item writing and forms construction process. Thus, to provide additional flexibility during the development process, the Table 3 was collapsed into Tables 1 and 2. It is important to note that Tables 1 and 2 are not separate blueprints, but rather different ways of representing the two-dimensional mapping of the blueprint to either content domains or sets of SEPs.

Table 1. Test Blueprint by Domain - Targeted Number of Dichotomous* (DCTMS) and Non-Dichotomous** (N-DCTMS) Items per Test

Reporting groups		**Grade 5: DCTMS**	**Grade 5: N-DCTMS**	**Grade 8: DCTMS**	**Grade 8: N-DCTMS**	**HS: DCTMS**	**HS: N-DCTMS**
Engineering Technology and Application of Science	**Physical Science**	13-16	3-6	13-16	3-6	15-21	3-6
	Life Science	13-16	3-6	13-16	3-6	15-21	3-6
	Earth and Space Science	13-16	3-6	13-16	3-6	15-21	3-6
Total Items in Test		**40-45**	**12-15**	**40-45**	**12-15**	**50-60**	**12-15**

Table 2. Test Blueprint by Science and Engineering Practice Set (SEP) - Targeted Number of Dichotomous* (DCTMS) and Non-Dichotomous* (N-DCTMS) Items per Test

Reporting groups		Grade 5: DCTMS	Grade 5: N-DCTMS	Grade 8: DCTMS	Grade 8: N-DCTMS	HS: DCTMS	HS: N-DCTMS
Cross Cutting Concepts	Total Items for SEP Set 1	14-18	4-6	14-18	4-6	17-23	4-6
	Total Items for SEP Set 2	14-18	4-6	14-18	4-6	17-23	4-6
	Total Items for SEP Set 3	11-13	3-4	11-13	3-4	12-18	3-4
Total Items in Test		**40-45**	**12-15**	**40-45**	**12-15**	**50-60**	**12-15**

Table 3. Two-dimensional Blueprint - Test Blueprint by Domain – Approximate Targeted Number of Dichotomous* (DCTMS) and Non-Dichotomous** (N-DCTMS) Items per Test

Reporting groups	Grade 5: DCTMS	Grade 5: N-DCTMS	Grade 8: DCTMS	Grade 8: N-DCTMS	HS: DCTMS	HS: N-DCTMS
SEP Set 1: PS	4-8	1-2	4-8	1-2	5-9	1-2
SEP Set 1: LS	4-8	1-2	4-8	1-2	5-9	1-2
SEP Set 1: ES	4-8	1-2	4-8	1-2	5-9	1-2
Total Items for SEP Set 1	**14-18**	**4-6**	**14-18**	**4-6**	**17-23**	**4-6**
SEP Set 2: PS	4-8	1-2	4-8	1-2	5-9	1-2
SEP Set 2: LS	4-8	1-2	4-8	1-2	5-9	1-2
SEP Set 2: ES	4-8	1-2	4-8	1-2	5-9	1-2
Total Items for SEP Set 2	**14-18**	**4-6**	**14-18**	**4-6**	**17-23**	**4-6**
SEP Set 3: PS	3-6	1-2	3-6	1-2	3-7	1-2
SEP Set 3: LS	3-6	1-2	3-6	1-2	3-7	1-2
SEP Set 3: ES	3-6	1-2	3-6	1-2	3-7	1-2
Total Items for SEP Set 3	**11-13**	**3-4**	**11-13**	**3-4**	**12-18**	**3-4**
Total Items for PS	**13-16**	**3-6**	**13-16**	**3-6**	**15-21**	**3-6**
Total Items for LS	**13-16**	**3-6**	**13-16**	**3-6**	**15-21**	**3-6**
Total Items for ES	**13-16**	**3-6**	**13-16**	**3-6**	**15-21**	**3-6**
Total Items in Test	**40-45**	**12-15**	**40-45**	**12-15**	**50-60**	**12-15**

Note: Tables 1 and 2 represent different mappings of the same set of multi-dimensional items. Their total rows count the same sets of items and should not be added together.

Item Specifications

As shown in Table 1, items on the test will be aligned to the *Engineering Technology, and Application of Science* (ETS), but this domain will not be reported as a fourth domain sub-score. Similarly, as shown on Table 2, items on the test measure will students' ability to apply *Crosscutting Concepts* (CCC), but information about CCCs will not be reported separate from the SEP set sub-scores.

The vast majority of items developed will be *two-dimensional* at minimum (measuring both a DCI and SEP), and as many items as possible will be *three-dimensional* (measuring a DCI, SEP, and CCC). Use of *one-dimensional* items will be limited and used only if and when absolutely necessary.

Phenomena, which are observable events that occur in the universe and that students can use science knowledge to explain or predict, play a central role in the Illinois Learning Standards for science, shifting from learning about a topic to figuring out why or how something happens[6]. All ISA items (multicomponent and stand-alone) posed to students will involve phenomena and/or problems. Information related to the phenomenon provided by the scenario (e.g., graphs, date tables) will be necessary to successfully answer the items with a cluster. Each specific item cluster will align to at least one Domain topic[7], but will cover multiple SEPs, CCCs, and individual performance expectations.

This is not a traditional large-scale assessment blueprint and design, where items are designed to measure a one-dimensional construct as precisely as possible. The new items still measure one construct, but it is a three-dimensional construct. Tasks targeting a specific standard or set of standards will individually reveal a key component of the scientific understanding associated with those performance expectation targets.

We are still several years from being able to release items from the test for analysis, but the kind of student performance and item analysis teachers are used to conducting on tests like the SAT or the old Illinois State Achievement Test will not be sufficient or appropriate. As such, professional development for educators and administrators that can help them understand the complexity of both the standards and the assessment results is critically important to ensure results from the new assessment are instructionally impactful. To support this, ISBE will release an annotated sample test for each grade that outlines the framework for item development as well as the mapping to each dimension for the sample items?

**Dichotomous* items have discrete right and wrong answers. These include, but are not limited to, item types, such as multiple-choice, true-false, select all that apply, fill in the blank, matching, etc. It is possible, albeit challenging, to develop dichotomous items that measure all three dimensions of the standards (DCI/SEP/CCC) and entirely possible to develop dichotomous items that measure at least two dimensions. Dichotomous items can be interactive. Dichotomous items are typically worth 1 raw score point.

***Non-dichotomous* items are items for which more than one answer may be valid. They are often, but not always, scored with a rubric as there are distinguishable differences in "correctness" and partial credit may be awarded. Non-dichotomous items can be and often are worth more than 1 raw score point. These include, but are not limited to, item type, such as extended response (e.g., long answer or essay), constructed response (e.g., short answer, simple diagram, develop a formula, etc.), and technology-enhanced items (e.g., graphing or charting results, interacting with a simulation, etc.). Non-dichotomous items are frequently differentiated by the length of time or type of technology needed to access or answer them.

---

[6] Using Phenomena in NGSS-Designed Lessons and Units, accessed 4/1/2019 at: https://www.nextgenscience.org/sites/default/files/Using%20Phenomena%20in%20NGSS.pdf

[7] Topic Arrangements of the Next Generation Science Standards, accessed 4/1/2019 at: https://www.nextgenscience.org/sites/default/files/AllTopic.pdf

No more than three non-dichotomous items will be *extended response items* in each Domain or SEP reporting set in this blueprint. Extended response items differ from other types of items in that they require a substantial written response. SEP 6-constructing explanations and designing solutions; SEP 7-engaging in argument from evidence; and SEP 8-obtaining, evaluating, and communication of information all require forms of scientific writing, but this is not an assessment of writing or literacy; thus, they are limited. Additionally, no extended response item will be worth more than 3 points each.

Proposed ISA Test Structure and Sub-claim Reporting to be Determined by Field Test Results

ISA Test Structure

Below is the estimated number of test units, the maximum time per unit, and approximate total test length. The committee's deliberations required a constant balancing act between ensuring that there are a sufficient number of items to report out with confidence on the areas of interest (Domains and SEPs) and keeping the overall length of the test manageable. A review of other states' NGSS-aligned accountability assessments shows a range from two[8] to four[9] hours in total length. Of primary consideration is keeping the length of the individual units short enough that they can be administered within a single period. This was also a concern with respect to students' ability to sit and focus for extended periods of time.

Table 4. Estimated Number of Test Units, Total Testing Time and Maximum Time per Unit

	Number of units	**Max. Time per unit**	**Total testing time**
Grade 5	2 - 4	60 minutes	Approx. 2 hours
Grade 8	2 - 4	60 minutes	Approx. 2 hours
Grade 11	3 - 5	90 minutes	Approx. 3 hours

There was a desire to keep the overall test length below the times approximated in Table 4: however, science is not tested every year so members of the committee present at the final meeting suggested that a longer amount of time is reasonable and necessary in order to cover multiple years' worth of standards. Additionally, there is a tension between the necessary logistics of scheduling multiple classes and grades for an assessment and a commitment to equity and validity by allowing each student as much time as they need to fully engage with the content and demonstrate their competency with science, as opposed to reading or writing.

Ultimately, all final details about testing times and unit composition will depend on the nature of the items and item parameters that emerge after field testing and that are needed to support the level of reporting desired by the committee. Writers will use text and cognitive complexity measures, as well as item parameters from previously used items, to estimate how long students will need to successfully complete an item, but only having students actually engage with the items will confirm whether these estimates were accurate or not. The test we had was too short to meet federal accountability assessment requirements. The length of the new ISA test will by default be

[8] Michigan Department of Education Science Assessment Update: Accessed 2/15/2019 at: https://www.michigan.gov/documents/mde/161024_MSS_Assessment_Update_Webcast_v.2_jl_539852_7.pdf

[9] New Jersey Student Learning Assessment for Science (NJSLA-S) Updates. Accessed 2/15/2019 at: http://measinc-nj-science.com/sites/default/files/2018-12/New%20Jersey%20Student%20Learning%20Assessment%20for%20Science%20NJSLA-S%20Updates.pdf

longer. Exactly how much longer will be driven by the blueprint and reporting constraints for validity and reliability.

Sub-claim Reporting

The committee quickly reached and sustained consensus that, if at all possible, the test should provide at either the school or district aggregate level, depending on the number of students tested, sub-claim information on the three Domains and especially the SEPs. A majority of committee members prioritized reporting on the SEPs over the Domains, as the SEPs were seen to be more reflective of the interdisciplinary nature of 21st-century scientific practice. There were concerns that reporting only by the Domains would encourage regressive curriculum design decisions. Given the topical nature of the DCI, reporting sub-claims reliably by the Domains is more easily achieved; thus, a solution that allowed reporting by both was pursued.

It was quickly determined that reporting on the eight SEPs individually was not feasible without employing some form of a matrix-sampled design, where not all students are assessed on the same content and/or not all content is assessed in the same year. Designs of this type were viewed as more or less problematic by the committee for a variety of reasons, and thus were not significantly pursued as a primary option past the second meeting. The committee is, however, open to revisiting the question of a matrix-sampled design once the item bank has been more fully developed.

Once matrix-sampled designs were removed from consideration, the committee did reach full consensus that it was acceptable to group the eight SEPs into sets, but only for the purposes of reporting. Two ways of organizing the SEPs were proposed and thoroughly considered. Sample blueprints built around the different SEP groupings were produced for the committee's evaluation and revision. However, the committee had significant concerns with both sets of groupings about imposing something onto the Illinois Learning Standards for science that was not grounded in the NGSS framework and intent.

The first grouping of SEPs came from the elementary grade-band group. This grouping aligns the SEPs in ways that are reflective of the other content area learning standards, with names that reflected this content alignment. This grouping of SEPs held intuitive logic for many committee members, but more than a few expressed significant concerns about the consequence of naming and reporting on something called "scientific literacy" and "scientific calculations," as there were concerns that this would lead administrators to teach more ELA and mathematics and less actual science, rather than promoting an increase in time spent on science instruction, which was the committee's shared goal. Additionally, this grouping of SEPs does lend itself to a slightly more balanced distribution of performance expectations across the Domain/SEP matrix; thus, members returned to it time and again for these reasons. Ultimately, the committee as a whole could not come to consensus to either fully accept or reject this way of grouping the SEPs into sets, but it did come to full agreement that naming the sets in ways that referenced other content areas was not desirable in any way.

Table 5. Science and Engineering Practice Sets Aligned to Content Areas

SEP Set 1	SEP Set 2	SEP Set 3
1) Asking Questions and Defining Problems	6) Constructing Explanations and Designing Solutions	4) Analyzing and Interpreting Data
2) Developing and Using Models	7) Engaging in Argument from Evidence	5) Using Mathematical and Computational Thinking

3) Planning and Carrying Out Investigation	8) Obtaining, Evaluating, and Communication of Information	

The second method of grouping the SEPs into sets comes from the McNeill publication shared by a committee member. Many of the committee members found the logic articulated by McNeill and her colleagues to be a compelling way to understand science practice. In particular, they felt that reporting out on categories of SEPs grouped in this way would lead to improvements in science instruction and learning, as opposed to the more content-aligned groupings. Additionally, there was perceived value in groupings that had a published research base. However, by that same logic, other committee members took issue with the fact that McNeill was imposing a way of thinking about the SEPs that was not envisioned in the original NGSS framework. Additionally, this particular grouping of SEPs is highly imbalanced, with a significant clustering of PEs in the sensemaking set. There remains a tension between constructing items that specifically pair the DCI/SEP/CCC that are reflected in the PEs, as opposed to combining the DCI/SEP/CCC in ways that are consistent with the NGSS framework, but not specifically identified in the Pes.

Table 6. Science and Engineering Practice Sets as Published by McNeill (2015)

Investigating	**Sensemaking**	**Critiquing**
1) Asking Questions and Defining Problems	2) Developing and Using Models	7) Engaging in Argument from Evidence
3) Planning and Carrying Out Investigation	4) Analyzing and Interpreting Data	8) Obtaining, Evaluating, and Communication of Information
5) Using Mathematical and Computational Thinking	6) Constructing Explanations and Designing Solutions	

The committee did come to full agreement that no one SEP or grouping of SEPs should be more important and, therefore, weighted more heavily than any other. As such, the committee did achieve consensus on the distribution of the items across the Domains and SEPs, with an equal distribution across the three Domains and the eight SEPs as reflected in Tables 1 and 2. However, after much debate, the Committee was not able to either fully accept or reject either method of grouping SEPs for sub-claim reporting.

The committee fully acknowledged, as does McNeill in her article, that there are many reasonable ways to group SEPs, and that different groupings help us understand science learning and practice in different ways. Thus, the committee understood the value of coming to consensus regarding particular SEP reporting groupings and developing a strong and clearly articulated logic for the group names and definitions of their meaning. All three pieces are needed for the results to have meaning to the field. Committee members were profoundly aware that the structure of the SEP reporting groups would promote and structure dialogue between administrators and their subject-matter expert teachers in both intended and unintended ways.

In the end, a compromise position was accepted to document both methods of grouping SEPs for the Board, but holding a final recommendation until there was potential empirical evidence

supporting stronger validity, reliability, or quality of items using one grouping of SEPs over the other based on the results of field testing.

Topics for Additional Consideration and Next Steps

The committee engaged in a number of discussions, raised important questions and considerations, and put forth proposals that were tabled for being beyond the scope of developing a blueprint for the new ISA. This does not mean these proposals are not without merit, simply that they would require legislative or other action beyond the power of the Board. Or they required resources, input, or expertise beyond the capacity of the committee. These topics for additional consideration are, in no particular order:

- **A proposal to add a goal regarding science competency and/or readiness to the Board's goals**

  The committee met at ISBE in rooms where the Board's mission, vision, and goals were posted prominently. It was noted on several occasions that science was not directly mentioned anywhere in those statements. Thus, the committee requests that the Board give consideration to including science directly in future iterations of its mission, vision and goals.

- **The development of a broader statewide system of professional learning, formative assessment, and improvement of learning**

  It was pointed out at various points throughout the committee's work that for the assessment and its results to truly be meaningful and impact the quality of scientific learning in the state, the Board would need to develop and support a set of policy initiatives, central among which needed to be sustained professional development for administrators and teachers in dialogue with each other. Thus, the committee requests that the Board, while monitoring the progress of the item development and field testing of the new ISA, gives consideration to the kinds of professional learning and engagement activities needed to best leverage the opportunities presented by the new assessment statewide.

  A significant message within this professional learning must be the importance of teaching all standards to all students, as was intended when they were adopted. The committee wanted an assessment that addressed the full scope of the standards, but acknowledge it might be a process over multiple years to fully recognize that vision. The committee worked hard to come to agreement on a design of the test that honored the design of the standards so it would drive practice—meaning three dimensional, phenomena based, and requiring application of the SEPs and CCCs.

  The committee spent considerable time contemplating reporting, with the idea that the nature of the reports determines how the assessment will be used. While the committee reached consensus on where it landed, which prioritized maximizing data to districts over providing specificity at the student level, there are implications of this recommendation that will need to be considered moving forward.

  The committee requests that the Board consider making a commitment to supporting professional learning developing other resources that help districts and schools understand and use the information provided by the sub-scores. These resources must provide clear explanations detailing how the information should be used by teachers, schools, and districts as well as clear explanations about how the scores relate to opportunity to learn.

- **A proposal to increase the graduation requirement for science from two years to three years and/or to specify some elements of the content required in those years**

  There was stronger interest and consensus regarding the desire to require three years of science in high school in order to graduate. One of the most significant challenges for the high school grade-band working group was that, currently, School Code only requires two years of science to graduate and does not specify either what content must be covered in those two years or what two years the classes must be taken in. This made it very challenging for the working group to come to consensus on what exactly should be assessed. The committee requests the future consideration of this change by the Board, while acknowledging this may require legislative or other action beyond the Board's direct influence.

- **A proposal to require a minimum number[10] of minutes of science instruction, either daily or weekly at some or all grade levels**

  This proposal, which is similar in nature to the proposal to change the high school graduation requirements, is designed to ensure that students are receiving a well-rounded education that prepares them for careers or postsecondary education in science-, engineering-, and technology-related fields. Even beyond the need to develop future scientists and engineers, the committee sees broad value to a society that is scientifically literate, curious, and capable of problem-solving. The committee requests the future consideration of this change by the Board, while acknowledging this may require legislative or other action beyond the Board's direct influence.

- **A proposal to assess science annually and/or to develop and require a science assessment in grade 2**

  There was strong consensus in the elementary grade-band working group that waiting until fifth grade was far too long to have an accountability assessment This results in science not being taught in any meaningful way until fifth grade in many schools and districts. Ultimately, the proposal was tabled for two reasons. First, the committee co-chairs were unable to recruit subject matter experts to sufficiently guide and support the work in the time the committee had to finalize its recommendations. Second, there remained questions from members in the middle or high school grade-band groups about whether assessment of science in second grade would have support from other key stakeholders or could be done validly and reliably to the standards required for an accountability assessment. ISBE's Division of Assessment and Accountability remains open to exploring the possibility of a lower-grade assessment, provided we can recruit sufficient subject matter experts and garner sufficient Board and broader statewide support.

[10] There was support for some unspecified amount of required minutes, but also a specific recommendation for a minimum of 60 minutes of science instruction either daily or weekly at all grade levels, but especially elementary, in support of the National Science Teacher's Association position statement for elementary and Science education. "**There must be adequate time in every school day to engage elementary students in high-quality science instruction that actively involves them in the processes of science**. NSTA does not find a research basis for recommending a specific number of minutes for teaching core content, including science. However, most states, districts, and schools currently prescribe a set number of minutes—either by day, week, or year. As a result of this practice, science receives far less instruction time than other core subjects (Horizon Research 2013). NSTA recommends that science be given equal priority as other core subjects, so schools should strive for at least 60 minutes of science instruction a day, including significant science investigations." (NSTA, 2018)

Both the elementary and middle school grade-band groups felt annual assessment might be an acceptable proposal (at least in grades 3-8), but acknowledged that this willingness to assess annually might stem from their passion for the subject, rather than represent a broad statewide consensus. However, the possibility that each year's test might be shorter in exchange for annual assessment was touted as a possible way to gain support for the idea. Again, ISBE's Division of Assessment and Accountability remains open to developing and implementing this suggestion over time at the direction and with the support of the Board.

- **The consideration of a computer-adaptive design for the ISA**

  The committee spent very little time considering a computer-adaptive design for the ISA. This is in part due to the significantly larger item bank required for computer-adaptive designs; the preference for item clusters anchored to real-world phenomena, which are not conducive to a computer-adaptive design; and a general preference for performance-based assessments of students' competency that are non-dichotomous and, therefore, typically must be human scored. The lack of and perceived expense of a sophisticated artificial intelligence scoring engine for non-dichotomous items meant that the committee had a general bias toward using item development and item scoring as a form of assessment literacy professional development for teachers that should be promoted over a computer-adaptive design. This does not preclude the committee or the Superintendent and the Board giving this option future consideration as the item bank and assessment scoring technology improves.

- **The consideration of cost and potential necessary future constraints**
  - The costs associated with the development of the assessment rarely the committee's deliberations. ISBE and Board members will need to give consideration to this issue in the context of the broader assessment system. It is entirely likely that the assessment envisioned by the committee will cost considerably more than the previous science assessment. When resource constraints are reached, the committee asks that the Board make transparent what tradeoffs are made, and asks that the Board keep in consideration the design principles articulated in this report.

The work of the committee truly begins with the Board's approval of this blueprint. The committee has already begun drafting a values and principles document that articulates the logic and reasoning behind the assessment's design in language accessible to educators. Many other test specification documents will need to be developed in support of the blueprint, and the item writing that needs to occur. Partners from SIUC and from other higher education and nonprofit organizations will work to develop a curriculum to train large numbers of item writers throughout the state. It is the hope of ISBE that many of the members of the committee expand their involvement with the assessment and become trainers or facilitators of these item-writing workshops. Additionally, working groups will need to be convened for bias and sensitivity review, native language speakers will need to be recruited to assist with development of transadapted versions of the assessment, and technology and accessibility specialists, along with special education experts and advocates, will need to be convened to ensure access to the assessment for students with disabilities.

Contributions to Equity

In conjunction with these next steps, the committee would like to conclude with a few reflections on the value of this work and its contributions to equity for students in Illinois. First, there is a deep commitment to equity embedded in the standards themselves and in a state's adoption of rigorous science standards. Science is what allows us to tackle entrenched problems in our world and in our society. Without well-reasoned, sustainable solutions to these problems, systemic inequities will persist. Second, by taking a phenomena-based performance approach in our assessment and in our instruction, we access a wider range of learning modes and provide opportunities for success to more students. There is substantial literature that supports the importance of providing opportunities for students who struggle in traditional oral and written learning modes, and science is rich in opportunity to engage with the whole child and develop these other ways of interacting with the world. Finally, it is the desire of the committee that the data from these assessments be used by the Board to ensure equitable distribution of resources in the areas of greatest need, so that all students can receive a well-rounded education and graduate ready for success in life.

Appendix A

Committee Membership

Name	Role	Title	Organization
Daniel Brown	Co-Chair	Assessment and Accountability Division Supervisor	Illinois State Board of Education
A. Rae Clementz	Co-Chair	Director of Assessment and Accountability	Illinois State Board of Education
Dr. Harvey Henson	Co-Chair	Interim Director & Assistant Professor of Geology and Curriculum & Instruction	STEM Education Research Center; Southern Illinois University Carbondale
Dr. Senetta Bancroft	Facilitation Staff	Assistant Professor of Science Education	Southern Illinois University Carbondale
Angela Box	Facilitation Staff	Assessment Coordinator & STEM Faculty	STEM Education Research Center; Southern Illinois University Carbondale
Dr. Lingguo Bu	Facilitation Staff	Associate Professor of Mathematics Education	Southern Illinois University Carbondale
Dr. Shu-Ren Chang	Facilitation Staff	Psychometrician	Illinois State Board of Education
Dr. Andy Metcalf	Facilitation Staff	Psychometrics Consultant	Independent Contractor SIUC
Dr. Yanyan Sheng	Facilitation Staff	Professor & Psychometrics Team Lead	Southern Illinois University Carbondale
Pam Stanko	Facilitation Staff	Assessment Consultant	Independent Contractor SIUC
Amy Alsop	Member	Union Professional Issues Director South	Illinois Federation of Teachers -AFT
Suzanne Asaturian	Member	HS Science Teacher	Carbondale High School
Norene Ault	Member	HS Science Teacher	Effingham Unit #40 Schools
Dr. Carol Baker	Member	Assistant Superintendent for Academics and CFO & NGSS Writer	Hinsdale Township High School District 86
Tara Bell	Member	Instructional Specialist/Title I Program Manager	Indian River School District #204
Dr. Alissa Berg	Member	Assoc. Director of Science Curriculum & Instruction	Academy for Urban School Leadership
Laura Bertermann	Member	HS Science Teacher	Grayslake North High School
Ranell Blue	Member	MS Science Teacher	Harvard District #50
Andre Botello	Member	6-8 Science Specialist	Chicago Public Schools

Amy Boyd-Telford	Member	HS Science Teacher	Salem High School
Marc Campbell	Member	Superintendent	New Holland-Middletown Elem SD #88
Juli Cannon	Member	Director of K-8 Science Instruction	KIPP Chicago Schools
Dr. Jason Crean	Member	Biology Instructor & NGSS PD Provider	Lyons Township High School / Saint Xavier University / ISTA
Jennifer Donatelli	Member	District Curriculum & Instruction Science Admin	Naperville School District #203
Holly Flanagan	Member	6-8 Science Teacher	Ladd C. C. School
Julie Fleming	Member	HS Science Teacher	Roxana CUSD #1
Kevin Frederick	Member	1St Grade Teacher	Booker T. Washington STEM Academy CUSD #4
Durenda Fuchs	Member	Special Education	Williamson County Education Services
Stacey Geluck	Member	5th Grade Science Teacher	Whiteside Middle School SD #115
Allison Grandberry	Member	PK-5 Science Specialist	Chicago Public Schools
Candace Hamilton	Member	8th Grade Science Teacher	Whiteside Middle School SD #115
Amie Hanneken	Member	HS Science Teacher	Roxana CUSD #1
Justin Harrison	Member	8th Grade Science Teacher	Vandalia Jr. High School
Dr. Danielle Hauser	Member	Director of Instructional Improvement	Township High School District 211
Julia Jeffreys	Member	HS Science Teacher	Roxana CUSD #1
Dr. Michael Lach	Member	Director, STEM Policy and Strategic Initiatives	University of Chicago
Dr. Rob Lang	Member	Director of Innovation, Teaching & Learning	Community High School District #99
Jeremy Larson	Member	Superintendent	Paris Union School District #95
Peter Leonard	Member	Director of Student Assessment	Chicago Public Schools
Jonathan Loucks	Member	HS Science Teacher	Mount Vernon Township High School
Deanna Mazanek	Member	HS Science Teacher	Grayslake Central High School
Hope McBrain	Member	Reading Teacher & Technology Integration Specialist	Hillsboro CUSD #3
Wesley Miller	Member	K-8 Science Instructional Support Leader	Chicago Public Schools
Mary Jane Morris	Member	IEA Teaching and Learning Director	Illinois Education Association

Nancy Kawecki Nega	Member	Retired MS Science Teacher & PD Coordinator & STEM Consultant	DuPage Regional Office of Education & ISTA
Aimee Park	Member	ISBE Science Foundational Service Provider	Lisle School District #202
Dr. James Pellegrino	Member	Distinguished Professor and Co-Director	Learning Sciences Research Institute; University of Illinois at Chicago
Kristin Rademaker	Member	President-Elect	Illinois Science Teachers Association
Gwendolyn Randolph	Member	Science Content Specialist	East St. Louis School District 189
Dr. Brian J. Reiser	Member	Professor, Learning Sciences	Northwestern University
Dr. Theresa Robinson	Member	Director of Secondary Education	Elmhurst College
Jenny Sarna	Member	Director, District Support	Achieve
Jeanine Sheppard	Member	ISBE Math/Science Content Specialist	Center for Education Initiatives at Illinois State University
Jennifer Smith	Member	8th Grade Science Teacher	Monticello Middle School
Terry Stroh	Member	Assistant Principal	Central School District #301
Sandy Villarreal	Member	Elementary Bilingual Teacher	Cowlishaw Elementary School
Caitlyn Walker	Member	Science Principal Consultant	Illinois State Board of Education
Dr. Donald Wink	Member	Professor of Chemistry	University of Illinois at Chicago
Anne Zahn	Member	District Administrator K-8 Science	Waukegan School District #60