

REIMAGINING STANDARDIZED TESTING IN THE US: A VIEW FROM HIGH-PERFORMING SYSTEMS

BACKGROUND PAPER FOR 2020-2021
LEGISLATIVE INTERNATIONAL EDUCATION
LEGISLATIVE STUDY GROUP



July 2021

For at least the past decade, there has been widespread frustration with standardized testing in the United States. Parents, students and teachers have protested that U.S. schools test students too much. They say standardized tests, generally designed solely for state and federal accountability purposes, exert too much influence over instruction. The questions on tests measure rote learning and superficial knowledge rather than depth and application of knowledge. While the tests are designed to provide useful information to policymakers, they are often of little use to teachers. Standardized testing has been used to measure learning but never used to promote teaching and learning.

Because of this history of dissatisfaction, there was little surprise when standardized testing was scaled back to an unprecedented degree due to the disruptions of the coronavirus pandemic. All states suspended statewide accountability testing in the spring of 2020. Not only state accountability tests were paused: the Advanced Placement tests went to an abbreviated, online, free-response format; the College Board and ACT cancelled testing sessions; and colleges and universities stopped requiring the use of the SAT and ACT in admissions decisions. The Governing Board of the National Assessment of Educational Progress (NAEP) even postponed the administration of the 2021 reading and mathematics tests.

But, as civil rights groups have long pointed out, standardized testing has a vital purpose in ensuring we have data we need to monitor equity in student learning – and canceling it without “building it back better” risks removing a necessary lens on whether students are achieving equitably. The information that well-designed state standardized tests could provide is particularly critical now, since the school closures and virtual learning have almost certainly resulted in more inequitable outcomes for students. It is for that reason that the Biden Administration announced that state testing must resume for students in the 2021-22 school year. But the situation remains fluid: states have been given flexibility in the timing of their assessments, how they are administered, and what exactly they measure.

So, how might state standardized testing be redesigned to measure deeper learning, to monitor inequities and enable positive interventions for struggling learners, and to provide useful information to teachers, parents, and students that supports stronger forms of teaching and learning for all?

Testing in high-performing education systems around the globe may offer a way forward. These systems, which have consistently succeeded in maintaining high and much more equitable performance at an efficient and sustainable cost, have long differed from the United States in how, when, and for what purposes they test students. Their experiences might be informative to U.S. state policymakers as they consider how to change standardized testing requirements to support much higher and more equitable levels of student learning.

Standardized Testing in High-Performing Systems

Standardized tests are used across the world to provide a consistent and comparable measure of student achievement across schools and other jurisdictions. That said, standardized

testing in high-performing systems differs from the U.S. approach in a number of important ways, including purpose and timing, what is tested, how it is tested, and how it is graded.

Purpose and Timing:

High-performing systems use standardized tests so that students can demonstrate that they are prepared to move on to the next level of their education. These exams, sometimes called graduation exams, are given to all students at key transition points, such as the end of 9th grade and the end of 12th grade. Students only take these exams at one or two points during compulsory school, and they are not used for school or district accountability. The accountability is on the students. These systems do analyze the results of standardized graduation exams to inform policy, but they also supplement them with additional low-stakes exams given at a limited number of grade levels — most often once in elementary and once in secondary school — to samples of students. Sample testing allows the jurisdiction to collect the data they need to monitor system progress and identify issues in a way that does not overly burden teachers or students.

Jurisdictions	Number of grade levels
Poland, Taiwan, Finland	2 grade levels
Japan, Singapore, Canada*	3 grade levels
China, Estonia, Korea, Hong Kong	4 grade levels
US	7 grade levels

*the number of grades varies by province

Content:

Because the key purpose of standardized tests in high-performing systems is to ensure that students are prepared for the next level of school, students are tested in the full range of subjects included in their national curriculum, including history, literature, sciences, and foreign languages. The tests typically cover a broader range of content than do state accountability tests in the U.S. that focus on only math and reading at many grade levels.

Format:

Standardized tests in high-performing countries are primarily constructed response and almost always include essay questions. In some cases, classroom work is also part of the exams. These formats allow the system to test deeper conceptual understanding of the subject areas as well as higher order critical thinking and problem-solving skills. These tests might include a limited number of multiple-choice questions, but that is not the primary format, unlike in the U.S.

Grading:

High-performing countries often use teachers to grade standardized exams. In some jurisdictions, certain teachers are trained specifically for this task; in others, it is considered to be a core part of the role of all teachers. In the latter, teachers learn how to grade these exams as part of their preparation for teaching and ongoing professional learning, which provides an opportunity for them to develop a shared understanding of the criteria for the exams and what standard-setting work looks like. They “grade” sample work and identify anchor examples of each grade. It is a key way to build deep understanding of the curriculum. There are policies in place to address concerns about the reliability of teacher grading, which include not allowing teachers to grade their own students’ exams, using additional teachers to confirm scoring on a percent of each set of exams, and doing external audits of grading.

Examples of What Subjects High-Performers Test

Estonia administers sample-based monitoring exams at the end of grade 3 in Estonian and mathematics and the end of grade 6 in Estonian, mathematics and an additional rotating subject each year. There is a graduation exam at the end of 9th grade in Estonian, mathematics and a choice of an additional subject from among a set of 14 foreign language, natural and social science options. A school-based cross-disciplinary project is also required. At 12th grade, students take a graduation exam in Estonian, mathematics and a foreign language and are required to complete a school-based inquiry project.

Finland administers a sample-based assessment to 9th graders in a rotating subject. Students in upper secondary school then take a Matriculation exam in mother tongue and a choice of at least 3 compulsory languages, math, humanities and natural science subjects and up to 2 optional subjects. There are 27 total subject tests.

Singapore has a graduation exam for primary school at 6th grade which tests students in English, mother tongue, science and mathematics. Students then take exams at 10th grade in English, Math, mother tongue, sciences and humanities, as well as computer applications for students in technical programs. At 12th grade, students take exams in 4 subject areas of their choice.

Implications for Policymakers

This approach to standardized testing means that:

The content of what is tested is clear to all teachers

and students: The purpose of graduation exams is to ensure that students have mastered the curriculum to a certain proficiency level and are prepared to move on. Past exams are made public so that the format and expectations of the tests are transparent to everyone: parents, teachers, students, and policymakers. With tests anchored in the national curriculum, the content is also transparent, as educators all know that curriculum. In general, there is little widespread pushback against testing, as its purpose is clear and it is generally perceived as useful for teachers and students.

These tests are better measures of the range of the critical thinking and applied skills we value for

students: Tests with open-ended questions and opportunities for students to explain the steps in their problem-solving and their reasoning give teachers much more insight into what students know and what they can do with that knowledge than traditional multiple-choice formats. They are tests of deeper learning. Many of these tests also include classroom work, which provides students different kinds of opportunities and time to show what they know.

Teachers have more time for classroom assessment: Standardized tests in the U.S. and other countries are only one kind of assessment. Other types include classroom assessment, including formative assessment to inform daily teaching and learning and summative assessments to ensure understanding of the course content. But because the U.S. spends so much time, money, and energy on annual standardized accountability testing, these other kinds of tests, more closely aligned to the work of teaching and learning, often fall by the wayside. In contrast, most assessment in high-performing systems is done by teachers to inform and monitor learning in the classroom. Teachers have more time to focus on classroom assessment, which allows for timely and personalized interventions with students. In addition, as standardized tests assess mastery of each subject in the curriculum, classrooms assessments and standardized tests reinforce one another.

Examples of Information is Available to Teachers, Students and Parents about Tests

Singapore's Ministry of Education develops a [syllabus](#) for each assessed course. The syllabus lays out the aims of the course and content to be covered, as well as the format and criteria for the assessment. Past years exam papers are [published](#) by a set of approved publishers and are available to teachers and students.

Alberta's Ministry of Education administers Diploma exams that are given to 12th graders in 10 subjects. The Ministry [releases](#) exemplars, examples of student writing that are graded using the grading criteria and practice questions for all exams in its website. There are also guides for students for each exam, explaining the goals, the format and the assessment criteria.

The cost of the kind of standardized exams is higher per administration than in the US,

but the overall cost of standardized testing is not higher: The kind of exams described here are often thought of as prohibitively expensive in the U.S. since they cannot be graded by scanning machine like multiple-choice tests can. But contrary to that perception, since test are not used annually, the overall cost is the same or less than the total for annual exams of the type that are used in the U.S.

Directions for the Future

High-performing systems are not standing still. While their approach to standardized testing has worked better in many ways than the approach in the U.S. to date, they, alongside the U.S., are rethinking their assessment systems broadly to better address the equity issues highlighted by the pandemic and to better prepare students for success as citizens and workers in a fast-changing economy. The answers to these challenges are not yet clear, but their emerging thinking can be instructive for U.S. educators. The issues that are front and center for them are:

How best to assess a broad range of cross-disciplinary competencies

In jurisdictions across the globe, schools are being asked to focus on cross-disciplinary competencies like problem-solving, communication, and digital skills in addition to core knowledge in disciplinary areas. Jurisdictions have increasingly incorporated competencies like problem-solving and critical thinking into their subject-based standardized testing, but questions have arisen about whether all competencies are best assessed this way and if all competencies even need to be assessed. For example, digital skills are a competency increasingly required for all students, but it is not clear if digital skills should be tested as a stand-alone set of skills, whether digital skills should be integrated into the testing of specific subject areas, or if digital skills are a competency that needs to be tested on a standardized assessment at all. Estonia, for example, has developed an assessment of digital skills but is just allowing teachers to use this as a tool in their classrooms.

How to report on the development of values and attributes that are key to success

Increasingly, systems around the world are recognizing the importance of values like ethical behavior and attributes like perseverance and leadership that are key in the development of students who are prepared for school, work, and life. They are grappling with how to monitor and report on the development of values and attributes like these that are impossible to quantify. Singapore, for example, has developed a framework for teachers to use to assess students' holistic development in co-curricular activities. The framework covers four domains: participation, leadership, service and achievement.

Reducing the stakes of high stakes standardized tests

While standardized testing serves the purpose of demonstrating whether students are ready to move on to a new level of schooling, there are concerns about using a single test as a measure of a student's capability. High-performing systems are allowing students to retake tests and making promotion decisions based on factors beyond test results, including teacher and parent recommendations. There are also efforts to break standardized assessment into small modules that students can take at different times as they are ready. Finland is using this approach in its standardized assessment of secondary vocational students.

Providing additional opportunities for students to show what they can do

While high-performing systems see a key role for standardized assessment, many systems are developing additional opportunities for students to demonstrate what they know and can do, in addition to a single high-stakes exam. These countries have typically admitted students to university based on test scores alone but are now broadening their criteria to include interviews or portfolios of work or records of outside-of-school-activities to students' records and to admissions applications, a practice that is more common in the U.S. than in much of the rest of the world. The goal is to develop broad, yet fair, profiles of students that are taken into account when making promotion and admissions decisions.

How to use technology to enhance testing

High-performing countries are considering how technology can improve the way they test students. They are experimenting with building authentic standardized assessment of real-world skills like designing science experiments, analyzing video content, interpreting foreign languages, and using technology-enabled scenarios as the topic of assessment tasks. The goal is to make the assessments more reflective of real-world applications of knowledge and skills and enable systems to assess a deeper range of skills and knowledge. That said, clarifying how to assess distinct skills and knowledge and how and if grading can be done by machines is a challenge.

Final Note

In the U.S., standardized testing has become a polarized debate between those who think we need annual accountability testing to ensure system quality and shine a light on equity issues and those who see it as a distraction from deep learning that exacerbates the inequities already present in the system. But high-performing systems show that it is possible to think about standardized tests as part of a broader assessment system that drives high performance for all students.

Appendix: Examples of Tests

Singapore:

Singapore has national exams for students at the end of 10th or 11th grade called O-level exams. Compulsory subjects include English, mathematics, Mother Tongue, Sciences and Humanities. Students also take several electives. The exam formats vary by subject. A syllabus lays out the content covered, the exam format and the assessment criteria.

The O-level science exam has 5 sections. It has a multiple choices section, three “structured and free response” sections with some required questions and some choices of questions focused on different areas of sciences (physics, chemistry and biology) and a practical test. According to the syllabus, “the practical test will require the students to suggest a modification or extension of an experiment, which does not need to be executed.” There are three questions on the practical test, one for each of the three areas of science. Examples of what might be asked in each area are included in the syllabus.

Practical exam questions in chemistry for the 2021 Science exam¹

Candidates may be asked to carry out exercises based on: quantitative experiments involving the use of a pipette, burette and indicator such as methyl orange or screened methyl orange; speeds of reactions; measurements of temperatures; problems of an investigatory nature; simple paper chromatography; filtration; tests for oxidizing and reducing agents as specified in the syllabus; and identification of ions and gases as specified in the syllabus.”

Singapore also offers a set of Technical exams at the 10 or 11th grade level for students wishing to pursue an applied curriculum. Compulsory exams are in English, Mother Tongue, Math, Computer Applications and Social Studies. Students also chose among a set of 20 elective applied subjects. Most of the exams for these subjects include a practical component. The exam is described in the course syllabus.

¹ Source: https://www.seab.gov.sg/docs/default-source/national-examinations/syllabus/olevel/2021syllabus/5078_y21_sy.pdf

The Design and Technology 2021 Exam²

The Technical Design and Technology exam includes two parts: a 20-week design project which counts for 70 percent of the exam grade and a one hour written exam which counts for 30 percent of the exam grade. The Design Project will comprise two components:

- The Design Journal is a real-time document that reflects the candidate's attempt at carrying out the design process. It should contain design sheets showing the use of:
 - a time-stages plan such as a Gantt chart to plan and to monitor the progress of the project
 - information and images, sketches (if necessary) and calculations to arrive at a design brief and design specifications, to generate ideas and to develop an idea into a working prototype leading to a proposed design solution. Use notes and annotations only if necessary. Candidates are advised not to re-work any design sheet. Mock-up(s) and the resulting prototype are to be submitted as part of the journal.
- The Presentation Board is to show the proposed design solution. Design features should be highlighted to illustrate the purpose of the proposal. Appropriate drawing skills should be used.

² Source: https://www.seab.gov.sg/docs/default-source/national-examinations/syllabus/nlevel/2021syllabus/7062_y21_sy.pdf

Finland

At the end of upper secondary school, students in Finland take a set of subject-based matriculation exams. Students are required to take at least four exams, one of which is Mother Tongue (Finnish, Swedish or Sami) and the other three chosen from among a second national language, a foreign language, mathematics or an exam in either humanities or natural science. There are 27 different tests. Each test is different; all offer a choice of questions to answer.

Sample questions from the 2018 Matriculation Exams ³

Social Studies: What does the budget of the European Union consist of and how does it contribute to the structural development of the various regions of Europe?

History: After several years of preparation, a Conference on Security and Co-operation in Europe (OSCE) was held in Helsinki from 30 July to 1 August 1975. Three text excerpts and a video excerpt about the Helsinki EEC Conference are included here.

- Compare the interpretations of the passages on the aims of the EEC meeting
- Consider how the video clip describes the atmosphere of the meeting and its achievements.

Finnish: Consider whether it is sometimes right to break the law.

Biology: Briefly explain how mutations in the genome affect the fitness of the individual and his or her offspring in the following cases:

- A mutation occurs in a bacterial cell that results in the development of antibiotic resistance.
- Radon gas in the breath causes a mutation that leads to the formation of a cancer cell and lung cancer in a 65-year-old woman.
- In a cyanobacterial cell, a mutation occurs in the regulatory region of a gene encoding a protein required for photosynthesis. The mutation prevents RNA polymerase from attaching to DNA.

³ Source: <https://yle.fi/aihe/artikkeli/2018/01/17/taalta-loydat-kaikki-kevaan-2018-yo-kokeet>

Hong Kong

Students graduating from secondary school take Hong Kong Secondary Diploma exams (HKSDSE). They take these exams in Chinese, English, mathematics Liberal Studies, and up to three elective subjects drawing from a set of 20 senior secondary subjects (such as Geography, History, Physics, Chemistry, ICT, Combined Sciences, Visual Arts, Tourism and Hospitality, and Design and Applied Technology); 52 Applied Learning subjects organized into six areas: Creative Studies; Media and Communication; Services; Applied Sciences, Engineering and Production; Applied Learning English; and Applied Learning Chinese; and six other languages (including French, German, Hindi, Japanese, Spanish and Urdu).

The format of the exam varies, but many exams include a “school-based assessment” component. Typically, teachers are given a set of “tasks” to choose from.

One task from the 2012 English language exam school-based assessment is doing a book review. The task includes a “description of activities leading to the assessment”; an “assessment activity” and a “post assessment activity”. For this task, here are:

The description of activities leading to the assessment ⁴

1. Tell students that they will be asked to give a brief book review on a good book they have read leading to assessment recently. Distribute the Book Review Questions and model the activity. Students ask you some of the questions from the handout and you provide the answers, e.g. “My book is called *The Lord of the Rings*, written by J. R. R. Tolkien. It is about ...”
2. Have students review their reading log, select a book they have read and be prepared to answer the Book Review Questions.
3. In the next class, students take turns asking questions and giving reviews to a partner. Students may use some personal notes (written on a 4x6 note card with points only) for reference, but they should try to look up and speak as naturally as possible.
4. Encourage students to extend their answers. Tell the questioners to pay attention to what has been said and not to ask a question that has already been addressed. Tell them that it is not necessary to ask all the questions on the list. They could ask other questions and extend their discussion based on the content of the book.
5. Allow a set amount of time, such as 3-5 minutes for each review. When time is up, tell students to switch roles.

⁴ Source: <https://www.hkeaa.edu.hk/DocLibrary/SBA/HKDSSE/SampleTask-ENG-PartA-050510.pdf>

6. After both partners have reported on and answered questions about their books, they can work with a new partner and repeat the same procedure in steps 3-4. After two or three rounds of sharing, encourage students not to refer to their personal notes or the Book Review Questions during their interaction.
7. Repeat the process a few more times if time allows.
8. Finish by asking several students to report back to the whole class about any books they have heard about in this activity that sounded particularly interesting.

The assessment activity

When the students are familiar with the procedure of book sharing, ask them to prepare to talk (Group interaction) about another book they have read for the SBA. Tell them that they will not know who their conversation partner will be until the time they are being assessed. During the assessment, the students will take turns asking questions and giving reviews as they did in the pre-assessment activity. The teacher-assessor may interact with the students to ask specific question(s) to clarify and/or extend their ideas after the students themselves have finished their conversation.

The post assessment activity

1. Ask students to do self- and peer-assessment after the assessment activity.
2. Give feedback on the overall performance of the class and give advice on how to make improvements by making use of the assessment criteria.