
In-Progress Reflections No. 1 on *Current and Critical Issues in the Curriculum and Learning*:

Student Learning Assessment and the Curriculum: Issues and Implications for Policy, Design and Implementation



| | |
|-----------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------|
| Title | Student Learning Assessment and the Curriculum: issues and implications for policy, design and implementation. |
| Series | Current and Critical Issues in the Curriculum and Learning |
| In-Progress Reflection | October 2015, No.1 IBE/2015/WP/CD/01 |
| IBE Director | Dr. Mmantsetsa Marope |
| Coordination and Production Team at the UNESCO IBE | Massimo Amadio, Renato Operti, Lili Ji, Émeline Brylinski |
| Author | Joshua A. Muskin Non-Resident Senior Fellow, The Brookings Institution IBE-UNESCO Consultant |
| Keywords | Education – Learning Assessment – Curriculum – Competence – 21st Century Skills – Relevance – Teaching – Education Planning – Teacher Training. |

Note from the IBE team:

The IBE has launched the series In-Progress Reflections on *Current and Critical Issues in the Curriculum and Learning* to open a communal space for a global conversation, collective production and discussion on those issues of high concern for Member States. It intends to support country efforts in mainstreaming challenging issues within the processes of curriculum renewal and development across different levels, settings and provisions of the education system.

Initially, the focus areas of the In-Progress Reflections series encompass, among others,: (i) Early Childhood Care and Education (ECCE) as a foundation of holistic child development and learning; (ii) Reading and writing in early grades to support the development of essential competencies; (iii) Youth Culture and competencies for Youth in the early 21st century (covering formal, non-formal and informal education); (iv) ICT curricula and inclusive pedagogy contributing to relevant and effective learning outcomes; (v) STEM (Science, Technology, Engineering and Mathematics) curricula to foster sustainable development; (vi) Curriculum for Global Citizenship Education (peace, human rights, sustainable development, values, ethics, multiculturalism, etc.); (vii) Assessment to enhance and support learning opportunities and (viii) Inclusive education as an over guiding principle of education systems.

The series of reflections covers a wide array of knowledge products, among them: discussion papers, policy briefs, frameworks, guidelines, prototypes, resource packs, learning tools and multimedia resources. These materials are discussed, refined, used and disseminated engaging education and curriculum agencies / institutes, and in particular curriculum developers and specialists, development experts, policy makers, teacher trainers, supervisors, principals, teachers, researchers and other educational stakeholders. Also, they serve as reference materials for the IBE menu of capacity-development training on curriculum, learning and quality education – namely masters, diplomas, certificates and workshops –, to forge policy and technical dialogue involving a diversity of stakeholders and to support sustainable country field work.

Through blogs and e-forums, we encourage the audience to actively interact and bring in diverse perspectives. Effectively, the online space for reflection allows us to stay connected, facilitates exchange between experts from different regions of the world, and truly fosters continuous reflection on the issues concerned. The blog is structured to gather diverse resources, which include tools and documents (as previously mentioned) under specific themes so as to provide a complex and rich set of materials targeted to the specific needs of Member States. The In-Progress Reflections will capture relevant visions, views and comments shared by the audience,

and serve as a key resource to support Member States' efforts in mainstreaming relevant findings and effective practices in national policies, curriculum frameworks and developments and in professional practices.

Student Learning Assessment and the Curriculum: Issues and Implications for Policy, Design and Implementation

Abstract: The role of assessment in education has grown greatly over the past few decades, a trend that has two major manifestations. One is the rapid increase in the number of countries and other jurisdictions either participating in international surveys (tests) of learning or initiating their own system-wide assessments; or both. The other is the ever-rising importance of assessment to hold systems and their key actors (notably teachers) accountable for education outcomes. The recent renewal by the world's nations and lead international organizations at Incheon, Republic of Korea of their commitment to an education "of quality" for all by 2030 and the upcoming global commitment to the new Sustainable Development Goals will now 'raise the bar' for education in terms both of equity and of how to perceive "quality," which now requires a much more relevant lens. Measuring progress towards these goals will begin with the assessment of learning, to determine both whether students are acquiring the required knowledge and competencies and whether a system is providing students with the appropriate education to acquire these outcomes. While assessment will be vital to this process, there is a severe double risk that systems and their partners will continue to rely excessively on tests to drive its reforms. First, most major tests do not reach all students and focus on just a few subjects – primarily Reading and Mathematics, and sometimes Science –, with the common result of a narrowing of the curriculum and of other distractions to the education process. Similarly, with rare exceptions, such tests neglect the broader range of personal competencies, such as the acquisition of new knowledge using a variety of methods, and the practical application of the basic knowledge and techniques students learn in school. The second risk is the continued failure to coordinate assessment with the other major functions of the education system – perhaps most notably, the curriculum, operating instead in relative (if not total) isolation. For assessment to be of high quality and relevant, and for it to inform real improvements to the overall education system and its outcomes, it must be in full and functional harmony with a system's curriculum, teacher training and support, texts and materials, planning, budgeting and all other departments. The present report explores the ways in which assessment is vital to education and posits means by which it can connect effectively to the other key education functions to drive a national system forward to 2030.

I. An overview of student learning assessment in the post-2015 EFA¹ and SDG² context

As Peter Drucker, the Austrian-born American management guru³, is widely credited to have famously asserted, what gets measured gets done⁴. While intended to explain what happens in the management field, Drucker's observation pertains widely across sectors, including education; and more particularly, to the curriculum. The curriculum embodies the knowledge and skills with which a nation or other jurisdiction or institution *intends* to equip its children and youth so that they are ready to assume adulthood in an engaged, productive and fulfilling way. In order for the official (intended) curriculum to generate this outcome, it is important that it also be the curriculum that schools (and other education institutions) actually deliver and that students learn. Unfortunately, though, in many instances the official curriculum is compromised, with the content that actually appears in classroom instruction resembling only a dim or narrow facsimile of what the education system, and indeed all of society, expect and require that students learn.

Many reasons can explain this gap, which in many settings is sometimes more of a chasm. As explained by Benavot⁵, a frequent cause is that the textbooks that a system provides to its teachers do not correspond to the curriculum. This may be a matter of a clash of content, or knowledge and techniques, but it can also be a matter of education philosophy, or approach. For example, a curriculum may feature competency-based methods; but if the textbooks remain information-heavy, presenting facts to consume rather than problem-solving and other analytic and operational tasks to practice and master, instruction and learning will likely remain largely rote. A curriculum may sabotage itself in a similar regard. For example, while espousing a broad range of competencies and knowledge for students, it may be so overburdened with content – both with the number of subjects and the amount of material in each – that teachers feel they must virtually 'sprint' through the lessons to cover all the material⁶. Developing competencies takes much longer and is more complex than delivering information. A third explanation is found in the capacity of teachers, which itself manifests a few different dimensions. Most obviously, do they possess the necessary skills and knowledge to deliver the curriculum? This pertains equally to pedagogic acumen and to their mastery of the subject content. Capacity also refers, however, to factors such as suitable facilities, access to appropriate learning materials, adequate time, and encouragement and support. When

¹ - Education for All, the global initiative for achieving universal basic education begun in Jomtien, Thailand in 1990, renewed in Dakar, Senegal in 2000 and embraced again in Incheon, Republic of Korea in 2015.

² - Sustainable Development Goals, the global initiative for a comprehensive, equitable and sustainable development program at a global level, is scheduled to be finalized and formalized in September 2015.

³ - See more on Drucker at <http://www.druckerinstitute.com/peter-druckers-life-and-legacy/> (26 May 2015).

⁴ - A search on the Internet reveals that, in reality, this statement does not appear in any of Drucker's major writings, though the attribution is frequent and the basic sense of the phrase is reflective of much of what he has written; and, perhaps, it is something he actually uttered in one form or another on one or many occasions. See <http://athinkingperson.com/2012/12/02/who-said-what-gets-measured-gets-managed/> (26 May 2015).

⁵ - Benavot (2012)

⁶ - Those students who can keep up, an 'old guard' of educators may contend, are "brilliant" and therefore are effectively certain to advance through formal schooling. The fact of many dropouts, the same logic asserts, serves simply to demonstrate the intense rigor and high quality of the academic program. These are conclusions that few systems and societies can afford to believe any more. The requirements of the economy and of society overall increasingly demand higher levels of literacy, numeracy and other knowledge and cognitive skills of all citizens; so a brutal triage is rejected. In addition, it is widely understood that there are many factors beyond a student's intellectual capabilities that can undermine her/his school performance and progress through the grades.

these other factors do not correspond to the content and requirements of a curriculum, even the most capable and knowledgeable teacher will be handicapped in delivering the curriculum.

The potential “obstacle” to the full and successful implementation of a curriculum that is the focus of the present paper is student learning assessment. Where these two functions are out of sync – an all too common phenomenon –, the efficacy of both is in peril. The curriculum risks being hijacked by ‘renegade,’ or contradictory, assessment standards and strategies, and assessment risks either irrelevance or, unfortunately more common, providing misleading and disruptive signals concerning achievement and accountability to the overall system and its diverse constituencies (students, parents and society) and partners. Channelling Drucker again, it would seem only minor hyperbole, if exaggeration at all, to say that system-level examinations and assessments *hold all other education initiatives hostage*. Irrespective ministerial policy statements concerning pedagogic methods, irrespective the curriculum, and irrespective what teachers learn in professional development sessions, if the content and skills on which students will be examined and the manner in which they will be expected to exhibit that knowledge do not conform to the curriculum and the official pedagogy, most teachers will typically “teach to the test.” Indeed, if a teacher were not to do this, colleagues, parents, students and even many local authorities would judge her/him to be irresponsible, since it is the test that determines students’ opportunity to progress in her/his formal education and, therefore, to attain the social, economic and personal rewards that come with more education. While many will argue that teaching to the test is not necessarily the best strategy for teaching *to succeed* on a test⁷, this is unfortunately the solution that many embrace, whether they hail from the Global South or North. In a country such as the United States where, according to Kamenetz, an average student will take over the 12 years of her/his primary and secondary school career a full 113 standardized tests⁸, the implications of a test-driven instructional strategy on learning and on the qualities of a system’s graduates must be seen as prodigious.

The on-going Education for All 2015 and Sustainable Development Goals processes will hopefully have a significant impact on what countries teach and how they teach it over the next decade and a half, as it should influence greatly what international education donors and other institutions promote, fund, support technically and research. Following the logic above, these global processes also involve the elaboration and prioritization of key indicators to reflect the learning that undergirds these goals along with the tools and methods by which nations and their partners can measure the degree of attainment of these learning (and other education) indicators. The period leading up to 2015 (the end of the Dakar EFA mandate) has seen a global emphasis on assessing the basic competencies of Reading and Mathematics, and in some instances of Science⁹. This has been true of low, middle and high-income countries alike. Moving beyond 2015, however, the SDGs will require of education systems to place supreme emphasis as well on other aspects of the curriculum and, especially, on the range of cognitive and non-cognitive competencies that equip students and give

⁷ - See Muskin (forthcoming, 2015).

⁸ - Kamenetz (2015) and <http://www.americanradioworks.org/the-test/> (5 June 2015).

⁹ - This trend is documented in the 2015 Global Monitoring Report, which shows that over "the past 25 years, the number of national assessments has grown sharply, increasing from 12 in 1990 to 101 in 2013 (UNESCO, 2015b)."

them confidence to use these skills for full, productive and fulsome economic, social, environmental and personal engagement¹⁰.

The balance of the present report aims to explore what precisely this prognostication means for assessment. What of these new curricular challenges and objectives can and should systems and their many stakeholders and partners be assessing? How can the results of such assessments constructively inform curricular developers and the many persons who are responsible for the various elements required to deliver the curriculum successfully?

II. An overview of the political and operational foundations of learning assessment

Before answering these questions, it is important first to review what assessment means operationally and politically within an education framework. Defined most simply, the assessment of learning is the means by which an institution of person determines the degree to which a person has acquired a pre-determined body or knowledge and or skill. In practice, however, there is nothing at all simple about assessment. Rather, assessment comprises a variety of dimensions, all of which have a significant impact on what is taught and how, although to varying degrees. One way of parsing assessment is across the following aspects:

- the **Purpose** of assessment;
- the **Level** of assessment;
- the **Substantive Scope** of assessment;
- the **Nature** of assessment; and
- the **Impacts** of assessment.

The **Purpose of assessment** divides most basically across two vectors; one is to serve summative and formative aims, and the other distinguishes between these two pedagogic functions and accountability. Beginning with the pedagogic, *summative assessment* is essentially a process to determine what students have learned of a specific set of knowledge and skills at the end of an instructional period. This can happen upon completing a section, a chapter, a semester, a year or some other block of time or content. It can serve mainly to generate a score, or often just a portion of a cumulative score, or some other signal of a degree of attainment against the curriculum or it can permit to certify or accredit a student, whether to confirm her/his successful completion of a course of study (e.g., a diploma) or to qualify and, in many instances, select her/him for some further learning opportunity (e.g., college entrance) or a social, economic or other function. In contrast, *formative assessment* typically refers to methods to perceive which aspects of a specific body of content a student or group of students is and is not mastering with the aim of taking deliberate

¹⁰ - The two major international assessment regimes, IEA (the International Association for the Evaluation of Educational Achievement) and the Program of International Student Assessment (PISA), operated by OECD (Organization for Economic Cooperation and Development), have already pioneered large-scale assessments in this direction. IEA has administered since 2009 the International Civic and Citizenship Education Study (ICCS) and conducted the first International Computer and Information Literacy Study (ICLS) in 2013. OECD has operated PISA since 2000, which test focuses on students' problem-solving skills and the practical application of their reading, mathematics and science knowledge.

action to remedy any gaps. These ‘checks’ might occur in the course of instruction, posing questions, checking exercises, answering questions and the like. They might also occur at planned pauses in the learning process created specifically to identify what students are and or not comprehending, such as quizzes, homework assignments or other activities designed to indicate to a teacher and her/his students what aspects require further effort, whether for the whole class or select students from the class.

Usually, summative and formative assessments are kept separate. However, summative assessments can have important formative applications, even if systems do not (or even rarely) take advantage of these. Within a course, a teacher who finds in a summative assessment that a majority of her/his students have not mastered content that is vital to their learning the next block may fruitfully choose to review the preceding lesson to consolidate the needed learning before proceeding. Similarly, a few systems, for example, Palestine, Bhutan and Australia¹¹ have introduced mechanisms to use the results of examinations or other large assessments to provide teachers with guidance and materials with which to strengthen their instruction for the new cohort of students and to alert teachers receiving the previously tested students to what content they have and have not yet fully mastered.

Assessment for *accountability* aims broadly to determine whether the investments a system, society or even partners makes in education yield the anticipated and required outcomes in terms of learning. Relying normally on summative assessments, the idea is to hold different actors across the system responsible for results, maybe rewarding them when positive and, more often, criticizing or even penalizing them when not. It is in this manner that failure to attain at least minimal standards on an assessment may result in a minister’s resigning, if a whole system underperforms, or in the firing or re-assignment of teachers and directors or the closing of a school or its transfer to new management¹² if the failures are localized. It is the prospect of such draconian measures at the local level that often compels teachers to teach to the test rather than to teach to learn (a distinction that is explored further later.)

The *Level of assessment* refers at the same time to the locus of management of a process and the data it produces and to the coverage of the content tested and its ‘proximity’ to the official curriculum. The management of assessment resides at three basic levels. The first is that of the whole system, which uses assessments of a variety of natures both (i) to verify, or certify, that students have achieved satisfactorily the curricular standards and (ii) to determine the degree to which different education policies, strategies and inputs are yielding the desired results. In the first case, the type of assessment is usually an examination that all students take at the end of a cycle; or at least those who wish to be officially certified and to be eligible for selection for the opportunities available to recipients of a formal credential. In the latter case, assessment often happens at non-transition points in students’ academic careers and can be either sample-based or census-based. These tests also often cover only a narrow range of subjects; usually Reading and Mathematics, and sometimes Science.

¹¹ - See Muskin (forthcoming, 2015).

¹² - See for example Rogers, W.T. (2014); p. 4.

The second level sits above these system-wide assessments. These are regional and international surveys, or tests, of knowledge and skill. At the international level are the two main IEA¹³ tests, PIRLS (the Progress in International Reading Literacy Study) and TIMSS (the Trends in International Mathematics and Science Study) and PISA. PIRLS and TIMSS have operated since 1960 and now reach students in grades four (both tests), eight and twelve (TIMSS only) in 57 countries (in 2011) and 48 (for grade 8, in 2007). PISA launched in 2000 and was administered to 15 year-olds (irrespective the grade level) in 65 countries in 2009. There are also a few regional assessments, all of which also focus on the “big two” subjects of Mathematics and Reading. Two of these two operate in Africa – SACMEQ (the Southern and Eastern Africa Consortium for Monitoring Educational Quality) and PASEC (the Program for the analysis of the education systems of CONFEMEN¹⁴) – and the third in Latin America – LLECE (Latin American Laboratory for Assessment of the Quality of Education). The IEA and regional assessments aim to test students on a range of knowledge and skills that are meant to reflect the curricula of the wide range of countries that participate. (The task of capturing many curricula within one assessment is likely more challenging for the IEA instruments than for the regional tests since PIRLS and TIMSS reach across many more languages and contexts.) PISA, in contrast, tests knowledge and skills that they claim surpass curricular standards to, instead, determine how well a country is educating its youth for the global economy. Particularly in Africa, the regional assessments have represented for many countries the only true assessment other than end-of-cycle examinations that systems conduct. Both sets of extra-systemic assessments permit countries to compare the levels of their students to those of other countries against a common set of content and skills. While such comparisons are meant to provide systems inspiration and, especially, substantive analysis to help guide them towards strategic improvements, and sometimes do so, the resulting rankings, or “league tables,” often yield categorically unhelpful and even sometimes pernicious consequences in many countries and other jurisdictions¹⁵.

The third level of assessment is that of the classroom. While the other two feature summative purposes (though, as explained above, summative can also double as formative), classroom level assessments comprise both summative and formative strategies and aims. The design of such assessments can involve teachers alone but may also engage groups of teachers in creating instruments to assess a common subject or elements of learning that cross different subjects within a curriculum, such as expository writing, integrated applications and personal competencies, such as collaboration. Teachers may also find assessment materials – individual test items, instruments and broader protocols – in the public sphere, whether provided by the central education authority, by the private sector or by some civil society entity. In most settings, assessment at this level is teacher-driven, but many systems also promote strongly self-assessment and peer assessment. Bowing to the formative assessment purpose that is basically reserved for teachers, it is worth pointing to the vital aspect of “feedback,” by which teachers (or peers) use the results of assessment to help students perceive precisely the gaps in their understanding and/or performance in order to clarify misperceptions or to take concrete steps towards remediation. As reported by Hill¹⁶, “There is

¹³ - IEA refers to the International Association for the Evaluation of Educational Achievement

¹⁴ - CONFEMEN stands for the *Conférence des ministres de l'éducation des états et des gouvernements de la francophonie*; see www.confemen.org (22 May 2015).

¹⁵ - See, for example, the open letter to the Director of PISA published in *The Guardian* on 6 May 2014 and signed by over 80 academics and education officials; <http://www.theguardian.com/education/2014/may/06/oecd-pisa-tests-damaging-education-academics> (23 May 2015).

¹⁶ - Hill, Peter, in Barber & Rizvi (2013); p. 65.

compelling evidence from meta-analyses of hundreds of studies to indicate that formative assessment, when used to provide feedback on a daily basis to both teacher and students, is one of the most powerful interventions ever recorded in educational research literature. But it is rarely practised.”

The *Substantive Scope of assessment* pertains here to what knowledge and skills a particular instrument or approach measures. Scope concerns equally the domains of content a student is expected to have acquired and retained and the manners in which s/he is asked to demonstrate what they have learned. As indicated above, large-scale assessments, both systemic and extra-systemic, tend to focus on just two or sometimes three topics: Reading, Mathematics and Science. Examinations, to the contrary, will reach across the curriculum to assess all subjects, or at least the academic ones and those that can be measured using paper and pencil methods. Classroom-level assessments also reach across the full curriculum, covering all topics, and permit appraisal *and feedback* with greater depth, precision and timeliness. There is also diversity in the ways assessments require students to exhibit what they have learned. At one extreme is the rote presentation of information or procedures: What date?; What calculation result?; Which definition?; Which branch of government?; and so on. Such questions are well-suited to mechanically graded multiple choice and true-false “bubble” tests.

At the other extreme are open-ended questions that measure students’ answers based not just, or even not necessarily, on the factual ‘correctness’ of their answers but also, and perhaps even more so, on the process by which they get there. Such assessments appraise students’ mastery of the knowledge and skills from the curriculum as well as their critical thinking, reasoning, communication, inferential and many of the other competencies they need in order to turn the curriculum into meaningful and effective practice during their studies and beyond. These latter competencies defy bubble test formats, but they also defy simple capture as content that a system can add to the curriculum and textbooks. Rather, one may argue, they require space within the curriculum in which teachers (with training and guidance) can accompany their students to achieve “deeper learning¹⁷” of the curricular content by providing situations in and through which to cultivate these skills.

Spanning the middle space between the two extremes are assessments of students’ abilities to use problem-solving, reasoning and application as relate to the curriculum. These include such things as word problems for Mathematics, inferring and describing emotions or intentions from a written story and proposing a sequence of steps for researching a scientific hypothesis; for example, respectively, (i) rather than simply calculating the area of a circle from the length of the radius, such strategies might ask a student to determine the size of a pizza, (ii) instead of identifying facts from a text, a student may be directed to select from a set of possible motivations for a hero’s decision, and (iii) instead of calculating the length of the side of a triangle based on the size of an angle and one other side, a test may require a student to describe a trigonometric strategy for measuring the height of a cell phone tower. Operationally, this space may be suited equally to single answer ‘bubble’ tests and open-ended questions.

¹⁷ - See Pellegrino (2012).

Substantively, the approach is reflective of the high vulnerability of assessment to cultural clues that might be more familiar to certain groups of students than to others, even within the same country. Using the preceding examples to illustrate this risk, pizza is not a universally popular or even recognized food, what is considered heroic can differ from one context to the next, and cell towers are not ubiquitous in all settings. Test developers are certainly aware of this challenge and have strategies for managing it. Still, just a casual review of sample questions that the major assessment regimes do reveal would suggest that they can overlook important nuances that might be significant for many students. Test-takers might be able to surmise the intention of a question despite an unfamiliar clue, but if this requires even a little more time on a strictly timed test, which is highly possible, there is a built-in disadvantage; and why bother with the contextual cue at all if it is as likely to confound as to illuminate a question¹⁸?

What are missing from almost all standardized tests are the ‘translational’ competencies referred to above. As indicated, these are the behaviours, talents and traits that students require, and for which employers, civic leaders and others around the globe clamour, in order to use their ‘book knowledge’ from the curriculum to be effectively productive, engaged, and fulfilled participants in ‘real life.’ Just to scratch the surface of the long list of such attributes, these include confidence, perseverance, ambition, curiosity, teamwork, independence, learning, communication, empathy, listening, leadership, ‘followership,’ judgment, patience, and creativity. Also to repeat, these are precisely the competencies which so many governments have codified in their national education plans and reforms *but which so few have incorporated strategically, prominently and coherently* across their curriculum, teacher training (pre- or in-service), textbooks, inspection and, perhaps most essentially (remember the ‘hostage’ situation), assessment standards, instruments and practices.

This lacuna may be explained by a very simple reason: all of this is hard to do, and maybe most especially assessing such competencies. As Levin says,

... the specific non-cognitive or personality attributes required for successful adulthood are more diffuse and more contested and have not yielded to the straightforward measurement methods used for standardized tests. There is simply no global agreement on what is of consequence beyond student achievement and how it should be measured. For these reasons, and perhaps others, discussions of world-class education and educational systems have been limited to student achievement¹⁹.

Yet, is it really necessary to judge these competencies across all students in a uniform, standardized, even fully objective way? Certainly standardization is important for the measurement of knowledge and skills that are basically learned, used, performed and represented similarly across all contexts. But such is not necessarily or always the case for personal competencies, which are instead often learned, used, performed and represented in ways and combinations that are largely unique for each person. Referring to Gardner’s different natures of intelligence²⁰, one can also think of how confidence might manifest differently in students (and adults) whose strengths lie respectively in verbal-linguistic curiosity, logical-mathematical creativity, spatial-visual communication, bodily-

¹⁸ - See Muskin (forthcoming, 2015).

¹⁹ - Levin, Hank, in People for Education (June 2013); p. 10.

²⁰ - Gardner, Howard (2006). See also:

http://www.niu.edu/facdev/resources/guide/learning/howard_gardner_theory_multiple_intelligences.pdf (24 May 2015).

kinaesthetic. Such differences might serve as just a hint of the full range of other permutations that might result from combining intelligences and competencies. How does a system standardize this?

- A. By identifying a statistical middle range of indicators for each.
- B. By creating different tests for different 'types' of student.
- C. By establishing ideal standards based on expert research.
- D. You don't.

The 'correct' answer is "D," or at least that is the argument of this author. Given the vastness of variety in the nature and manifestations of these competencies, it would seem that their assessment is best suited to the classroom. On the one hand, a true and rich appraisal requires both a variety of methods to be accurate, valid and complete and, partly as a consequence of the first, demand considerable time and variety in the circumstances, or tasks, by which a student is judged. On the other hand, it would seem that only first-hand knowledge of a student's exercise of her/his competencies would truly furnish information that would permit a teacher to provide meaningful reflection, guidance and opportunities to help each student improve.

In the same vein, such rich information, even if subjective, should constitute highly useful information for a system to guide its efforts to strengthen instruction for these competencies. With results aggregated across a system's classrooms, decision-makers, curriculum designers, teacher training authorities, materials developers and others can all get ideas on how students are evolving across these skills and what teachers are, and are not, doing to foster these. Thus informed, they can create and disseminate policies, content, strategies, documents, technical training and support and even assessment tools and strategies to support greater teaching and learning in these domains²¹.

These viewpoints do not aim to dismiss entirely the opportunity to assess personal competencies using standardized, 'objective' indicators and methods; nor do they discount entirely the utility of such measures. Indeed, teachers can benefit greatly from standardized indicators, measures and tools; but these should probably not be tools that limit the assessment of performance to narrowly delimited, tightly time-bound tasks. A guide with clear and precisely defined rubrics and criteria for assessing creativity across a student's performance, for example, can plumb widely and deeply her/his related attributes in ways that a 'bubble' test or even a time-limited, precise task simply cannot²². Instead, the main point is to underscore the limitations of standardized approaches to

²¹ - This approach is reflective of the plan proposed by the State of Massachusetts, USA, to develop an index of creative and innovative education (Massachusetts Department of Elementary and Secondary Education, 2012).

²² - Robert Sternberg (2003:xvii), a leading education researcher in the development and assessment of creativity, has created standardized tests around tasks such as writing "stories with diverse titles like *The Octopus's Sneakers*" and doing "art work for topics such as *Earth from an Insect's Point of View*." While these challenges certainly offer students the chance to 'let their creativity flag fly,' there is no guarantee that a student will find the necessary inspiration to exhibit fully, or even partially, her/his creativity on one, or even a selection of these assignments. Perhaps the pressure of a test setting evokes serious 'writer's block' or 'artist's block,' or some other factor stifles her/him at that instant in that setting. Or, perhaps such tasks simply do not correspond to the interests of a student or to the aspects in which s/he is most creative? A highly ingenious math student is not necessarily an artist or visually adept. The fact of being a gifted guitarist and composer does not translate necessarily into inspired, inventive writing. Does the inability to complete one of these tasks mean a student is uncreative? And perhaps more meaningful, what decisions does a system make about promoting creativity from the information on standardized tests? Does it rally resources around creative writing and the arts, or does it equip teachers with the knowledge, tools and opportunities to assess and cultivate creativity in all domains as correspond most closely to the passions and talents of each and every student? Sternberg himself provides two major findings from his work that suggest the latter of the two options, the classroom option; though it is possible that he would not necessarily concur with this conclusion: "First, creativity tends to be fairly but not completely domain-specific. Second, [creativity] tends to be rather but not totally distinct from psychometrically measured intelligence (*ibid.*)."

testing both to capture sufficiently rich and reliable information and to be able to use this information to make truly meaningful decisions. Instead, there is a vitally important role here for classroom-level assessment. This may be asking a lot of teachers, and a lot more than many experts and critics are prone to believe that most teachers can handle. The view here, though, is that most teachers can handle this *if the system operates sufficiently to create, provide and support the conditions, inputs, context and competencies for teachers to do so.*

The *Nature of assessment* relates closely to what is being assessed, the intended purposes of the results and who is doing the assessment, as outlined above. Classic assessment for accountability, certification, selection or other summative reasons requires that all students answer the same range of questions in writing to situate them objectively *vis-à-vis* a common set of standards, or concrete learning objectives (criterion-referenced). Especially for selection purposes, the aim may be to situate, or rank, each student's performance against that of her/his counterparts (norm-referenced). If conducted as large-scale learning surveys, assessments commonly favour multiple choice and true-false questions which students answer by filling in bubble sheets. Certain large-scale examinations do include essays or other full-text responses, such as for the international baccalaureate. While the quality of responses and assessment is often higher for text-based answers, the cost is also significantly greater²³. In addition, for many large-scale assessments, especially sample-based ones, a detailed analysis of responses is of little importance, with systems seeking instead to get a broader overview of the knowledge and skills of its student population. Indeed, as Green explains, such tests routinely aim *not* to measure what each student knows but, rather, are contrived to differentiate between students in order to get a sufficient spread of performance²⁴, aspiring to the famous (or infamous) bell curve²⁵.

While such information might provide an acceptable general signal of the quality of student learning and the efficacy of the education system – for example, fourth grade students are reading at a second grade level or, with a bit more exactitude, Grade 7 students are strong in long division but weak in fractions –, it fails woefully in at least two vital ways. One, as indicated above, large-scale assessments (other than examinations), neglect much of the curriculum, both marginalizing or excluding entirely certain subjects and even narrowing the curricular content of tested subjects. (There is only so much one can squeeze into a test of a couple of hours.) Stecher and Chun provide a vivid illustration of how an emphasis on assessing Reading and Mathematics in Washington State, USA, resulted in a significant and sizable reallocation of time assigned by teachers to the other subject areas of the curriculum: over half of the Grade 4 teachers surveyed reported increasing the

²³ - Kamenetz (2015) reports in her podcast interview on American Radioworks a difference of \$30 and \$75 per student for, respectively, machine-graded and human-graded tests in the United States.

²⁴ - Green (2015). Extrapolating from Green, consider a case in which all students have mastered perfectly the curriculum, obtaining 100% on a criterion-referenced assessment. A standardized assessment must still aim to create a spread across performance levels; so in such an instance it must presumably "raise the bar" beyond the curriculum to be able to create a clear set of "winners" and "losers." Green ponders how ethical such an approach is. One may similarly wonder simply how effective such an approach and outcome is for teachers, for a school system and for all of society. How do curriculum designers, text and materials developers, teacher training institutions and services, planners or even broader society react to such information that essentially defies the curriculum, and therefore the official vision of a learned society, in order to generate an acceptably wide range of performance? What does accountability look like in such a scenario, for institutions and its actors?

²⁵ - For more on the Bell Curve, see Fendler and Muzzafar (2008).

amount of instructional time in Reading, Writing and Mathematics while sacrificing time to study Social Studies, Science, Arts and Health and Fitness²⁶. In Botswana, agents of the Ministry of Education's Curriculum unit report that the official, intended curriculum is narrowed in many classrooms by teachers who opt to concentrate solely or especially on the "assessment syllabus" provided early in the school year by the Botswana Examinations Council²⁷. As stated above, such a structural separation of these two domains which, in actuality, should be joined at the hips, leaves both weakened and at risk of diminishing the whole education endeavour.

The second failure concerns the generation of precise, deliberate and strategic information that would help systems, schools and educators to improve the level of student learning and, therefore, their performance on future assessments. Such an outcome is clearly a priority for systems, yet for some reason they tend to treat assessments primarily as an alarm bell, triggering media headlines by signalling general success or failure, rather than as a richly refined diagnostic tool that can point to precise areas of learning lacunae and, therefore, to precise solutions.

A pair of broad approaches might serve to treat assessments in a more diagnostic way. One is to provide to schools, teachers and the education agents who train and support them a thorough detailed analysis of the answers students provide on specific items from a test, allowing them to adjust their lesson design and delivery and their own assessments, to concentrate on any weaknesses; using summative assessments for formative purposes. This is what Bhutan has attempted to do with the help of Educational Initiatives, an Indian firm, producing an Annual Status of Student Learning that provides each school a full report of what test items its students got right and wrong and providing explanations for each of why most students got a question wrong²⁸. In South Africa, a pilot program is attempting to do the same, allowing teachers of the tested students to strengthen their instruction for the next cohort while informing those teachers who receive the tested students of their new students' strengths and weaknesses so that they also can adjust their teaching²⁹.

Of course, for teachers to be able to use such information effectively to improve their instruction, a few key conditions must be met. To begin, the analysis must be clear and comprehensible, it must be sufficiently detailed and relevant and it must be timely. On the 'demand' side of the equation, teachers must not only be sufficiently trained and able to understand the data and use it to influence their teaching, but they must have adequate time and authority to make adjustments to their instruction, including such aspects as pacing, sequencing, the examples and exercises they use, and supplementation, among others. They must also have access to suitable resources and instructional facilities to adjust their teaching.

Such detailed diagnosis provides whole systems with information they can use to furnish teachers also with supplementary instructional resources. This is what Palestine has done since its "shocking" TIMSS results in 2007³⁰. In addition, systems may choose develop and make available to teachers

²⁶ - Stecher and Chun (2001); p. 13.

²⁷ - Personal communication, 23 April, 2015; and Muskin (forthcoming, 2015).

²⁸ - Educational Initiatives (2010); and Muskin (forthcoming, 2015).

²⁹ - Kanjee and Sayed (2013); Kanjee, personal communication, 19 February 2015 ; and Muskin (forthcoming, 2015).

³⁰ - Matar, Mohammed, personal correspondence, 17 January 2015; Muskin (forthcoming, 2015).

sample assessment items and instruments focused especially on aspects of learning on which students struggled especially on a previous formal assessment, as in Australia³¹ and New Zealand³².

As indicated above, the nature of assessment is limited neither to pencil-and-paper tests nor to standardized learning objectives. Particularly as relate to the appraisal of students' reasoning, critical thinking and application skills and, perhaps especially, of their personal, or cognitive and non-cognitive, competencies, many other types of assessment are possible and, it is argued here, even recommended. As outlined by the Ministry of Education of the Province of Ontario, Canada in its official guide for the *Assessment, Evaluation and Reporting in Ontario Schools – Grades 1 to 12*, such alternatives might include "... formal and informal observations, discussions, learning conversations, questioning, conferences, homework, tasks done in groups, demonstrations, projects, portfolios, developmental continua, performances, peer and self-assessments, self-reflections, essays, and tests³³." While it may be important that there be clear standards even for these sorts of assessments, a system (and its partners) may wish to consider seriously whether it is really important that it have scientifically pure and comparable data on all dimensions of learning. Extrapolating from the argument from above, the more comparable the data on aspects of learning that tend to be more idiosyncratic, such as how a student exhibits curiosity, creativity or confidence (or even how a teacher nurtures these attributes in her/his students), the less precise the measure is likely to be and the less useful the results will probably be to the teacher in strengthening the development of these skills³⁴.

Finally, it is also important to mention the nascent but rapidly growing use of ICT (information and communication technologies) to conduct assessment. Developers and educators claim many advantages to the use of technology. Prominent among these are the ability to improve the logistics, decrease fraud and accelerate the analysis and reporting of results for large-scale assessments. At the classroom level, ICT allows teachers and students alike to adapt tests to the particular needs, abilities and pace of a class and, even, of individual students (adaptive assessments) and to generate virtually instant feedback, not just indicating what a student got right or wrong but also providing specific guidance and materials for remediating areas of weakness³⁵. Increasingly, such solutions do not require Internet but can use mobile connectivity and devices³⁶.

The main *Impacts of assessment* fall across three broad domains of decision. One concerns policy, program and planning for which the *education system* is responsible. At this level, assessment results might trigger reactions ranging from system-wide 'fixes' or more substantial adjustments or reforms – for example, as relate to the curriculum, professional development, budgeting and the equitable allocation of resources – to more targeted initiatives as might pertain to particular institutions, subject areas or assessment strategies, among others. The second domain concerns

³¹ - Cornell-Farrow, Sarah (2014); Muskin (forthcoming, 2015).

³² - Brown, Gavin T.L. (2014).

³³ - Ministry of Education, Ontario (2010); p. 28.

³⁴ - This observation pertains as well to the comment on the work of Sternberg, in footnote #21.

³⁵ - Redecker (2013:11) characterizes ICT-based adaptive assessment as "Generation 2,... tailoring the difficulty or contents of the next [test item] presented or an aspect of the timing of the next item on the basis of [individual] examinees' responses," which method she dates to the early 2000s. See also the New Zealand Assessment Tools for Teaching and Learning (asTTle), <http://e-asttle.tki.org.nz/> (4 June 2015).

³⁶ - See Hill & Barber (2014); Brown (n.d.); Muskin (forthcoming, 2015).

decisions about all individual students. The results of an assessment or combination of assessments under this guise will generally determine the educational path a student will follow, including such aspects as the level of study s/he will attain, the areas of study s/he may pursue if able to continue to a level of specialization, whether this will be in an academic or some training institution, the reputed quality of the institution in which s/he continues her/his education or training, and other aspects. (Certainly other factors also matter in many settings, such as socio-economic status, ethnicity and geography.) The third domain is the *classroom*, providing information that teachers (and other educators) might use (as described above) to adjust and strengthen their design, delivery, assessment and feedback of lessons and that students (and their parents) may use to revise and invigorate their learning strategies, conditions and outcomes; or perhaps to abandon school and pursue another kind of education.

The core argument of the present paper is that it is the very last of these purposes, to improve the student learning *process and outcomes*, which should guide ultimately *all assessment*, whether directly or indirectly. Unfortunately, this is often not the case. Instead, several international assessment experts asserted plainly, if ‘off the record,’ during the research for this paper that many assessments (at least large-scale, summative ones) often happen either as a matter of prestige, for internal political reasons, under pressure (whether explicit or implicit) from a funding partner, or to gauge a jurisdiction’s relative education attainment against other jurisdictions. Even when a system conducts assessment with the clear purpose of ‘taking the temperature’ of its education programs, there are scant examples worldwide of a deliberate follow-on deep diagnoses and efforts to use the results deliberately and strategically to strengthen the quality of the system’s education delivery and outcomes³⁷. (The reaction of Palestine to its shocking 2007 TIMSS results, referred to above, provides one promising exemplar.)

In many instances, even if a country wanted to use the results of a large-scale assessment to make precise, deliberate and strategic decisions to strengthen the quality of its programs, delivery and outcomes, the design and, perhaps especially, the management of many assessment programs either make this difficult or preclude the chance of such a reaction altogether. The design of items and instruments can be of such a high degree of scientific sophistication that it can be difficult to use individual items for reporting and remediation purposes³⁸. Furthermore, as Green implies, the actual test items may in the end have little to do with the key learning aims of the curriculum.

Of a more pedestrian nature, assessment institutions are often loathe, or refuse outright, to share their instruments and items for diagnostic (or any) purposes so as to be able to use them in subsequent versions of their tests. In the United States, teachers who oversee test-taking must even sometimes sign a legally binding document committing not to divulge any items from the instrument

³⁷ - Germany is widely reported to be the ‘poster child’ for a reaction to a large-scale assessment that led to the significant qualitative turnaround of an education system, following its “disastrous” PISA 2000 results. (see Figazzolo (n.d.); p. 14; and Neumann, Fischer & Kauertz (2010)). Missing from the story in the literature, though, seems to be what diagnosis Germany actually undertook to identify how precisely the system should react to the ‘fever’ signalled by PISA to improve its results on the next round of the test. What also seems noteworthy, though, is how, 15 years later, Germany still appears as the preferred example of a true ‘turnaround’ story from PISA; where are the others? Instead, places such as Singapore, a regular “high flyer” on the major international assessment rankings, is reportedly deciding to re-orient its national education programs away from a myopic focus on success on international tests towards other priorities, finding that “...the price of high academic performance can be staggering.” As reported by Tan (2010; p. 53), it is not, however, the financial cost that worries the country but rather the costs in terms of lost “creativity and thinking skills among students and members of the workforce.”

³⁸ - Guadelupe, Cesar (personal communication, 30 January 2015).

they are administering under risk of prosecution³⁹. The growing presence of private companies in the international and national assessment sphere might appear to be aggravating this situation, concerned not only about future versions of a test but also about related spin-off services and products which they sell and that are linked to an assessment. Finally, also handicapping systems in using assessment results to improve education quality is the factor of time. On the one hand, there is the simple fact that test results and analyses can arrive well past the time of the assessment's implementation. Even if the conditions from the time of the assessment have barely changed, it is not certain that any problems indicated by the results still prevail; especially when results are not reported until three, four or even more years after a test, as has been attributed to SACMEQ. On the other hand, for those large-scale assessments that occur only once every several years, a system may also have to wait a long time even to generate results to analyze. This may be especially problematic in the event that the system wishes to appraise the effectiveness of some innovation or other sort of intervention in between tests.

Ultimately, as asserted above, all assessment should lead *deliberately and strategically* to better teaching and learning. At the Policy level, results *and their deep diagnosis* should lead to clear and coherent decisions about what to teach, how best to teach it and how to mobilize and allocate resources (across sectors) to achieve optimal quality and equity across the system. At the Planning level, results should guide systems in any revisions to the curriculum and, especially, to the inputs that systems provide directly to teachers and other front-line educators in order to translate the curriculum and the broader education vision into classroom instruction and learning. This concerns such aspects as the design and distribution of textbooks and other pedagogic materials, school facilities, teacher training, monitoring and support, extra-curricular programs, social supports to students and their families and communities, and assessment standards, strategies and instruments. Finally, at the level of Practice, system-generated assessment results should essentially do what teachers themselves do with their formative assessment efforts. This is to provide relevant, detailed information on what their students – whether individually or as a group – do and do not understand so that teachers can take decisions on the best methods, materials, content, pacing, balance, revision, timing, management, evaluation, and so on to incorporate into their lessons to meet the learning needs of their class and of the individual students therein. Assessment is too costly in terms of finances and, perhaps especially, of time and effort to squander on reasons other than to improve the quality of teaching and learning.

³⁹ - Green (2015).

III. Implications of current priorities and practices in student learning assessment for curriculum

To repeat, the implications of assessment (and of Drucker's iconic admonition) for the curriculum should be apparent: if a system truly expects its whole curriculum to be taught, then the whole curriculum is what must be tested, and this should occur in ways that matter. This conclusion pertains equally to the content of the curriculum, to the way it is taught, and to the ability to employ the content to solve problems and perform effectively and with fulfilment in all spheres of social and personal life. It also concerns the long list of personal competencies that society expects and needs the education system to help children and youth to cultivate, also referred to above.

As a corollary, if a system, or society, truly only values and therefore only really tests parts of the curriculum, perhaps it might consider eliminating from the curriculum those parts it finds unworthy of assessing. On the one hand, an education official might perceive this suggestion as somewhat facetious; tests should not drive curriculum. On the other hand, given that in practice tests do in fact drive much, if not most, of the curriculum that is taught and how, it would seem incumbent upon curriculum developers and assessment units to work closely together to ensure that the aims of both are attained. Unfortunately, such collaboration is often not only absent but education structures can represent functional barriers to this. In an effort to maintain the methodological and political integrity of assessments, the responsible unit can be completely separate from the rest of the Ministry departments, even operating with partial or full autonomy. In this view, assessment or examinations units operate to fulfil a manner of audit function that must not be influenced by the object of its scrutiny. The interpretation of this autonomy as institutional isolation can result in assessment that is out of sync, or even at odds, with the rest of the ministry, with the curriculum department and its goals' being perhaps most vulnerable. This can, in turn, provoke confusion or conflict within the supporting education departments – for example, teacher training, inspection, planning, and materials development – as they confront contradictions between what is tested and what the system professes as its official learning program; i.e., the curriculum.

It is therefore worth considering what communication and cooperation pathways might be valuable and, even, vital to establish between an assessment unit and the curriculum (and other) departments of a ministry of education. Besides helping to ensure that what a system sets as its learning goals for students is what its assessment programs actually measure, the communication of assessment results back into and across the system will promote the many linkages required to improve the quality of teaching and learning, as elaborated in the previous section. Additionally, moving beyond any inference of facetiousness from the corollary above, an analysis of what from the curriculum truly warrants assessment may indeed lead to questions about which content and objectives are truly essential or even just interesting and useful and which have exceeded their 'shelf-life' and are no longer relevant, or perhaps never really were.

In conversing with curriculum and other education authorities and partners as well as with teachers across many countries, especially in low-income regions of the world, there is routine agreement that curricula are often bloated and unwieldy. This, many also agree, often results from an additive approach to curriculum development, simply introducing new learning objectives, content and

strategies (for example, competency-based methods or citizenship as a subject) to an academic program without eliminating or, at least streamlining, other elements. This addition can be purposeful, completed as a planned part of curricular reform, or it can be *de facto*, with new content added while continuing to use old textbooks and assessment criteria (for students and/or teachers) and furnishing teachers with no meaningful guidance by which to navigate inherent contradictions. Especially when combined with the global phenomenon of increased time devoted to assessment and the greater ‘stakes’ associated with outcomes, particularly for schools, the conflict between the content and delivery of a curriculum and the assessment of its attainment would seem to be a formula for poorer rather than better quality education.

Such a provocative conclusion seeks justification in at least two arguments. One is that a curriculum on ‘steroids’ – i.e., overly bulky – leaves little time for formative evaluation and feedback, shown above to constitute two vital elements of learning. As teachers ‘sprint’ through the syllabus to cover all the material that the system requires students to learn in a year, there is precious little opportunity to take stock of what they are and are not grasping. And when lacunae do inevitably arise, teachers often feel that they cannot pause to review content and techniques that their students will require to move on successfully to subsequent lessons. The requirement for schools to prepare for ‘high stakes’ standardized assessments, whether national or international, robs teachers and students of even more time to devote to the curriculum and to valuable formative assessment and feedback while also sometimes resulting in a virtual hijacking of the content and the instructional process.

This latter danger resides in the practice of teaching to the test. A teacher, and indeed a system, may truly believe that by equipping students to succeed on the test, students are learning better. This logic is not certain, however, and indeed the opposite may be true. For one, teaching to the test, especially in high accountability contexts such as the United States, often involves aiming for the ‘floor’ of learning, or the basic minimal requirements for meeting official standards, rather than for the ceiling, the roof or the sky beyond. When a system sets its bar for success at the mechanics of reading – letter recognition, phonemic awareness, syntax, vocabulary, expository prose, etc. –, that is where many, if not most, teachers will stop in their instruction. Indeed, this is where a system may require a teacher to stop, as reported anecdotally by different teachers in the United States⁴⁰.

Some assessment leaders suggest that teaching to the test is acceptable if the test is good. Perhaps a legitimate position in the ideal, the limitations of time, the ability to process, disseminate and use the results and of other factors would suggest that such an ideal is manifestly elusive⁴¹. In addition to narrowing the content a teacher covers, whether due to the limited standards of assessment or to the time devoted to lesson preparation and delivery, a focus on the “assessed syllabus” can both demotivate students and undermine their learning. Focusing on grammar and vocabulary instead of on reading and writing for information, intellectual stimulation and pleasure or on mathematics equations and formulae rather than on using mathematics to solve compelling problems and answer stimulating questions compromises learning at at least two levels. For one, it saps a student of the enthusiasm to study and learn, turning curiosity and ambition into the routine and rote. At a more cognitive level, it deprives students of the opportunity to deepen and consolidate their

⁴⁰ - See Muskin (forthcoming, 2015).

⁴¹ - See Kamenetz (2015), *The Test: Why Our Schools are Obsessed with Standardized Testing—But You Don’t Have to Be*; and listen to a podcast interview with the author at <http://www.americanradioworks.org/the-test/> (25 May 2015).

comprehension of a particular concept, content or technique by excluding opportunities to explore the topic from a variety of angles and to use it in relevant ways⁴².

This same reasoning pertains to the second argument pertaining to how the conflict between assessment and curriculum diminish learning. This is that a bulked up curriculum and an emphasis on assessment for accountability leaves little time and generates little impetus for the development of students' vital personal competencies, as described and discussed above; regardless the priority placed on these in the national education vision. Learning to read and write, to perform mathematics or to demonstrate mastery of any other subjects for a standardized assessment leaves little time or energy to foster students' broader cognitive and non-cognitive competencies, whether in extra-curricular activities, as is most common, or in academic endeavours. A history lesson on World War II that requires students to read and retain related facts and to grasp text-based interpretations is much more efficient in terms of covering the content than would be a group project that engages students in planning and conducting an exploration of the topic by consulting documents and conducting interviews on the experience and impacts of the war on their own community and in putting together and sharing a multimedia presentation on the topic. Besides missing the opportunity to help students discover and develop their personal talents and other 21st Century skills, they lose the chance to learn the content better in at least two ways. One, as concluded in the preceding paragraph, they forego chances to deepen and consolidate their learning by approaching the material instead in a largely rote, one-dimensional way. Two, the very competencies that a more robust approach to education would promote and hone for more productive performance in 'real' life are those attributes that equip and motivate students to succeed better in their academic endeavours: confidence, curiosity, perseverance, ambition, planning, self-assessment, and so on. So, while text-based learning may be more efficient, it might appear that it is far from more effective.

Taking more of a bird's eye view of the dynamic between curriculum and assessment, the essential question would seem to be what combination of measures and measurement strategies would enable a system to use assessment equally for accountability, certification/selection and pedagogic purposes to optimize education delivery and outcomes. In the preceding analysis, the goal has been to point in at least two, closely connected directions to find a suitable answer. The one is to establish as a concrete strategic priority the use of standardized assessments and their results to inform decisions and actions at all levels of an education system – Policy, Planning and Practice. Assessment for accountability and for certification and selection are certainly valuable objectives; but these purposes need not and should not replace or overwhelm the implementation of assessment to generate results and analysis that will help actors and institutions across the education system, *and most importantly teachers in the classroom*, to improve classroom instruction and learning.

The second direction also points towards prioritizing improved pedagogic outcomes while seeking along the way a constructive balance between the three main aims of assessment and among the various natures, or methods, of measuring learning. As argued, accountability and certification/selection are not at odds with pedagogic improvement. Rather, they have much of value to offer and, in turn, with strengthened pedagogic performance, the results of accountability and certification/selection assessments should improve, to the acclaim of the education system and

⁴² - See Pellegrino and Hilton (2012); Bloom, Engelhart, Furst, Hill and Krathwohl (1956); Gardener (2006).

the benefit of society and, of course, of its children and youth. Mostly implicit in the preceding analysis but no less significant is the notion that education systems might find much more benefit in placing greater emphasis in terms of both importance and time on classroom level assessment. For one, assessments that teachers create and conduct can be more precise, more nuanced, more thorough and multi-dimensional and more reliable than a solitary standardized test or battery of tests. This is because teachers can (i) link their assessment directly to where a class is in the curriculum at a particular point in time, (ii) employ a diverse set of assessment methods to perceive and measure the diverse ways by which students can demonstrate their understanding of a topic or technique, (iii) create conditions that are more conducive to students' having confidence and, therefore, to their being able to exhibit what they really do and do not know, (iv) appraise each student's test answers with the ability to interpret and identify where understanding may hide behind a wrong answer, (v) shape an assessment to the particular level and needs of individual students, and many other reasons. Technically, standardized tests can imitate many of these same advantages, but to do so would likely entail exorbitant costs in terms of budget and logistics, though the advent of ICT strategies for assessment might mitigate these obstacles.

It is important to add here, however, that this recommendation does not imply a total abandonment of conventional standardized assessments. To the contrary, these are and will remain quite important. Rather, the proposition is that there be a greater balance between external standardized instruments and those assessments, both summative and formative, that teachers lead. In addition, as argued earlier, there is great potential, and a growing number of examples, of standardized assessments that serve to provide teachers with improved and flexible items, instruments and strategies that they can use with their students to assess and to strengthen their learning.

The other key benefit to learning that the implementation of assessment in the classroom under the heightened control by the teacher of the measurement methods and results can yield is the ability to analyze and use the outcomes in precise, timely, targeted, and various ways. Teachers can make decisions about their lessons mid-delivery, in preparation for the next day's or days' classes or over a longer period in reaction to instantaneous indications from assessments in the moment, from graded assignments or from information they receive on the results of their students or of a sample of other students on external tests.

Of course, being able to formulate and conduct assessments and to analyze and use their results in these highly constructive ways require considerable capacity of teachers and other front-line educators. Such capacity is hardly guaranteed in many if not most settings, and especially in low- and middle-income country settings, as many experts in the field of assessment are quick to claim. Rather, they find that the capacities required for such a robust and informative approach to assessment surpasses greatly those of the vast majority of teachers. Even assuming that this is a legitimate assertion, should it justify the continued marginalization of teachers in the implementation of assessment and their routine deprivation of meaningful, timely information on the results of tests that teachers might use to strengthen their instruction and their students' learning? Or rather, should the acknowledgment of the low capacity in assessment lead to a thorough investigation of the causes for this situation in order to elaborate and implement a set of measures to elevate this capacity and thereby raise the quality of teaching and learning? Such an attitude might be particularly germane given that the matter of "capacity," as explicated in the

report's final section, does not pertain to teachers alone but also to many other factors that are controlled by the system.

IV. A framework for harmonizing student learning assessment with curriculum within the 2030 world education agenda

A review of what a system needs to assess and, therefore, what capacity is needed to conduct assessment and manage its results may appropriately begin with a clear statement of the curriculum that a system truly wants and a society truly needs its children and youth to learn *and eventually to be able to put to good use*. Channelling the core message of UNESCO's 1996 Delors Commission Report, what does society require that its school graduates learn to Know, to Do, to Be Together and to Be? Referencing the joint Declaration of the just completed World Education Forum 2015 in Incheon, Republic of Korea, the expectations for learning are even more steeped than before in the competencies students must possess for success and fulfilment across all spheres of life and livelihood:

"Quality education fosters creativity and knowledge, and ensures the acquisition of the foundational skills of literacy and numeracy as well as analytical, problem-solving and other high-level cognitive, interpersonal and social skills. It also develops the skills, values and attitudes that enable citizens to lead healthy and fulfilled lives, make informed decisions, and respond to local and global challenges through education for sustainable development (ESD) and global citizenship education (GCED)" (paragraph 9⁴³; italics added).

Undergirding all the commitments to a "lifelong and sector-wide perspective, addressing access and results, equity and quality for all – children, youth and adults - from early childhood care and education to higher education and adult learning, and in formal, non-formal and informal learning"⁴⁴, is a fundamental obligation to furnish to all an education that matters. What matters most in learning looking ahead will also certainly be among the pressing topics of the discussions, debates, negotiations and agreements that occur at the culminating United Nations Summit on the Sustainable Development Goals in New York City on 25 to 27 September 2015. There, the nations of the world will convene to adopt and commit to a series of goals, targets and broad actions that aim at the elimination of global poverty, conflict and inequality and the attainment of educated, healthy, secure, productive populations and of a sustainable and healthy planet.

How does a curriculum prepare children and youth to create and nurture such a world, and how do systems assess how successfully they are preparing its students to be effective advocates and stewards of this world? The first part of this formulation is the focus of a paper that the UNESCO International Bureau of Education prepared for the Incheon Conference, the intent of which is captured well in the title: "Repositioning and reconceptualising the curriculum for the effective realization of Sustainable Development Goal Four, for holistic development and sustainable ways of

⁴³ - The full text of the Incheon Declaration, "Education 2030: Towards inclusive and equitable quality education and lifelong learning for all," is available at <https://en.unesco.org/world-education-forum-2015/incheon-declaration> (4 June 2015).

⁴⁴ - From the UNESO World Education Forum 2015 web site: <http://www.unesco.org/new/en/education/themes/leading-the-international-agenda/education-for-all/world-education-forum-2015/> (26 May 2015).

living⁴⁵.” Addressing particularly in the present report the formal assessment aspect of this equation, it might seem most critical to consider three operational objectives:

- i. confirm that the desired and required knowledge and skills (ideally, the curriculum) have been learned not just as information for short-term retention and representation on a test but as a coherent body of knowledge and competencies that the student can summon and use to navigate effectively and with fulfilment her/his path in life while contributing, both actively and passively, to achievement of the SDGs (assessment OF learning, or accountability, but for improving teaching and learning and for contributing to the aims of society, not for shame or acclaim);
- ii. contribute constructively to the learning process, both providing vital feedback for teachers and students alike (assessment FOR learning) and serving as an additional instructional tool (assessment AS learning⁴⁶);
- iii. provide the ‘right’ strategies, conditions and incentives to motivate, equip and guide teachers and other educators to teach to the full curriculum and the entire range of learning objectives and competencies, motivating students to learn to know, to do, to be together and to be within a dynamic SDG world rather than narrowing the curriculum primarily to ‘teach to the test’ and thereby produce rote learners and consumers of information with limited other competencies.

Achieving these objectives will require significant reflection, investigation, negotiation and reform for assessment in most, if not virtually, all education systems and regimes. In addition, the trend in many settings seems unfortunately to be away from the sorts of decisions systems must consider to assess for an SDG world. The story behind Anya Kamenetz’s latest book on testing in the United States illustrates this point bluntly: while she set out to research and write about some of the most exciting pedagogic models in the country – for example, project-based learning, blended learning, and maker spaces –, she had to change course because what she found in reality was that such innovations are in reality rarely used, squeezed out, she asserts, by the pressures of standardized testing. Perhaps the nascent but seemingly growing grassroots movements opposing the tyranny of testing mark a reversal of this trend. One might also hope that the new global benchmarks and commitments to emerge from the World Education Forum in Incheon and the Sustainable Development Goals will stimulate intense scrutiny and significant reforms of both curricula and assessment standards and strategies by countries and international organizations alike.

Anticipating such a shift is at the root of what the Learning Metrics Task Force (LMTF) initiative undertook to do beginning in 2012⁴⁷. Managed by the UNESCO Institute of Statistics and the Brookings Institution, a US-based think tank, and advised and supported by many of the major bilateral and multilateral international education institutions and prominent civil society organizations and researchers, LMTF first conducted a massive global consultation⁴⁸ to identify a

⁴⁵ - The draft paper is available at:

http://www.ibe.unesco.org/fileadmin/user_upload/Publications/drafts/WEF_IBE_Position_Paper_eng.pdf (8 June 2015).

⁴⁶ - See Ministry of Education, Ontario (2010) and Black and Wiliam (2001).

⁴⁷ - Other international initiatives to focus on promoting better and more useful assessment of learning and quality of education include the new UIS Catalogue of Learning Assessments (see http://www.uis.unesco.org/nada/en/index.php/catalogue/learning_assessments), the World Bank’s Systems Approach for Better Education Results (SABER – see: <http://saber.worldbank.org/index.cfm>) and the imminent Assessment for Learning program of the Global Partnership for Education.

⁴⁸ - The consultation for phase one of LMTF involved over 500 persons including teachers, ministers and ministry officials, representatives of civil society, academics and researchers, from 57 countries on six continents.

core set of Learning Domains and Measurement Areas⁴⁹ towards which all education systems might educate their children and youth. With these lists, the challenge has turned towards identifying the indicators and strategies for each domain and area by which to measure the degree of attainment, assessing equally individual students, education institutions and systems. To accomplish this, LMTF is currently partnering with 15 countries and other jurisdictions on four continents⁵⁰ - dubbed “Learning Champions” – to develop, experiment with and validate a range of assessment methods and protocols that might serve as models or, at least, inspiration for other countries to borrow, adapt or emulate.

One of the early lessons of the Learning Champion initiative is that achieving this aim will require attention to a few key aspects, a conclusion that reflects the related experience of many other systems across multiple contexts over the years. Most fundamentally, these aspects entail at the very start a firm official and societal commitment to a consensual vision of an ‘educated school graduate’ and to a clearly articulated goal and set of strategic objectives by which to realize the vision. The hope now in 2015 is that each country and other jurisdiction will refer to the Incheon Declaration and the SDG’s to review and either confirm, revise or scuttle and start afresh in committing to a “21st Century” vision, goals and objectives for its education system that is fitted to its particular cultural, social, economic and environmental priorities. The second key aspect is to mobilize and coordinate all the elements of an education system to review and adjust their respective policies, programs and actions to promote the renewed or new vision, goals and objectives with a combination of distinct and joint efforts. This pertains equally to the pedagogic, material, political, administrative, technical and contextual dimensions of a system and includes more precisely such elements as classroom instruction, student well-being, school and system-level planning and management, curriculum and pedagogic strategies and materials, and teacher training, monitoring and support.

The third and final key aspect is that of assessment, recalling again Drucker’s reputed admonition. In this regard, countries and international actors will wish to reconsider precisely why they are performing assessments, whether they are assessing what is really important, for learning and for life, and whether the way they are assessing is truly contributing to improved teaching and learning that fosters the broader goals of the curriculum and the overall society. The first two reflections here will relate closely to the first of the two key elements identified above: the vision, goals and objectives. Is the system assessing what society has identified as the priority competencies for moving society towards its aspirations, be they economic, social, ethical, environmental or related to personal dignity and fulfilment? As commented above, these competencies must embrace equally and together (i) basic information and techniques, combined to constitute knowledge, (ii) the discernment and technical ability to apply this (and other) knowledge, *and to acquire new knowledge*, to solve problems and to function effectively in social, personal and economic roles, and (iii) the

⁴⁹ - The seven Learning Domains are 1) Physical well-being; (2) Social & emotional competencies; (3) Culture and the arts; (4) Literacy & communication; (5) Learning approaches & cognition; (6) Numeracy & mathematics; and (7) Science & technology. The seven Measurement Areas are (1) Access to & completion of learning opportunities; (2) Exposure to a breadth of learning opportunities across all seven domains; (3) Early childhood experiences that promote development & learning in multiple domains; (4) The ability to read; (5) The ability to use numbers and apply this knowledge to real-life situations; (6) An adaptable, flexible skill set to meet the demands of the 21st century; and (7) A “Learning for All” indicator which measures the percentage of all children who have completed a learning cycle and achieved adequate learning outcomes. See <http://www.brookings.edu/about/centers/universal-education/learning-metrics-task-force-2> (5 June 2015).

⁵⁰ - The list of Learning Champions comprises Bogotá (Colombia), Botswana, Buenos Aires (Argentina), Ethiopia, Kenya, Kyrgyz Republic, Nepal, Ontario (Canada), Pakistan, Palestine, Rwanda, Senegal, Sudan, Tunisia, and Zambia.

broad range of personal competencies – cognitive and non-cognitive – that provide an individual with the ambition, the attitudes and behaviours to muster effectively all her/his intellectual, affective, social, moral and even physical attributes to engage in the world, to produce and to continue to learn in ways that contribute to the SDGs and the more parochial goals of a community in whatever ways suit her conditions, context and own sense of purpose.

V. Moving Forward

Returning to the focus of this report, what questions should education systems be asking themselves about how to review and revise their assessment priorities, programs and strategies as they look towards 2030, the end of the new SDG and Education for All mandates. This challenge pertains especially to the third of the objectives above:

provide the ‘right’ strategies, conditions and incentives to motivate, equip and guide teachers and other educators to teach to the full curriculum and the entire range of learning objectives and competencies, motivating students to learn to know, to do, to be together and to be within a dynamic SDG world rather than narrowing the curriculum primarily to ‘teach to the test’ and thereby produce rote learners and consumers of information with limited other competencies.

It also harkens to the reference above to the capacity to design and conduct assessments and analyze and use the results in ways that don’t just signal problems but truly contribute to their diagnosis and remedy; when a thermometer shows that a patient is febrile, the doctor does not proceed directly to treatment without first completing a diagnosis.

In both the literature and conference discussions, the worry about capacity in assessment usually concerns two areas. One is the knowledge, skill and tools of the assessment, evaluation or examinations unit of a ministry. Does the assessment team have the technical ability and the technology (the hardware and software) to be able to construct reliable, valid and appropriate tests and to archive, analyze and disseminate the results in a coherent, comprehensible and truly useful manner? The other pertains to the knowledge and ability of teachers and other local educators to conduct effective student assessment and to use the results, and those of large-scale tests, to improve their teaching and their students’ learning? Finally, what can and maybe should systems and their domestic and international partners be doing to strengthen capacity in both of these areas?

These two areas are certainly good places to start in addressing a system’s capacity to assess student learning and to use the outcomes to improve teaching and performance. However, despite common custom, they are a flawed place to stop when aspiring to ensure the most effective design and productive impacts of assessment. Capacity does not simply matter for those persons and departments who are directly responsible for assessment and its outcomes; primarily, the examinations unit and schools and teachers. Rather, effective assessment – that is, assessment that both generates accurate, reliable and thorough information on the performance of students and, therefore, of the whole system *and that provides data and analysis that contribute directly to improving teaching and learning at the classroom level and across all schools* – depends on solid systemic capacity, articulating, strengthening and coordinating the interests and impacts of all ministerial units (and even of other sectors) as relate to measuring and improving student learning.

At the system level, a fruitful reflection on this challenge might engage decision-makers in exploring the following sorts of questions:

- How accurately and thoroughly do system-level examinations and other assessments reflect the curriculum and other major learning outcomes sought by the system and by the overall society? To what degree should these in fact mirror the curriculum? What are the effects of any gaps or differences between what is assessed and what is intended for learning upon what and how teachers teach and what and how students learn? Are these effects a problem, and if so what might be done to improve the situation?
- What contact, cooperation and collaboration exist between the system's examinations or assessment unit and other education departments with a 'stake' in the content, planning, administration and/or outcomes of large-scale testing and outcomes? How does the assessment unit interact with the curriculum experts to identify what content to cover and what skills to feature in requiring students to demonstrate their knowledge and abilities as relate to the content? What data and analysis does the assessment unit report to the curriculum, materials development and teacher training departments about the areas and nature of students' learning strengths and lacunae and to these and the planning department about the locations of egregious shortcomings? How do all departments consult and coordinate to construct convergent strategies for improving teaching and learning that are informed by test results?
- What vital data about teaching and learning and the full body of knowledge and competencies to which these processes contribute fundamentally do system-led assessments NOT provide? Which of these dimensions of learning would be possible and appropriate to integrate into these large-scale assessments? Conversely, which ones would not? What might the different education departments *and actors on the front lines of education delivery* – that is, teachers, directors, local education agents and even parents and students – contribute to revising the large-scale assessments to incorporate these other learning dimensions? What assessment responsibilities (and opportunities – see below) might a system assign to these front-line education units and actors, and perhaps especially to schools and teachers, to provide accurate and useful information on students' competencies that are either poorly assessed or for which there is too little time to measure within large-scale instruments? How might a system merge and balance its central and school-level assessment results to make decisions about students (certification, selection, progression), about the performance of the system (all dimensions) and about actions to take to strengthen classroom delivery and outcomes?

At the school and classroom level, productive reflection, decisions and actions might emerge as system authorities and technical agents join with teachers, school directors, local education agents and even parents and students to explore another, related set of questions, such as:

- Do teachers and other front-line education actors have the knowledge and abilities required to formulate and implement effective and relevant assessment tools and strategies to use with their students in the classroom, both for formative and summative purposes? Are they sufficiently skilled to use the results of these assessments as well as the data and analysis they may receive from system-level tests to make appropriate strategic decisions about their lesson planning and delivery and to guide best the learning of individual students? Do parents (and

even students)? What are the specific knowledge and abilities they require to perform these functions and how best might the system train and support them to gain and use these?

- Are the data and analyses generated by assessment, especially from large-scale assessments, delivered in a form and format that is comprehensible and useful to teachers, other local educators and parents and students? Is the information sufficiently precise, pointing clearly to specific content and skills, or does it represent broad averages with little to no indication of where students have answered correctly and on what aspects of a topic they have erred? Do the data and analyses report on information that pertains to the curriculum in an obvious manner? Are they distributed in a timely manner, allowing teachers, other educators, parents and students to react in ways that allow for suitable remediation, ensuring that students master at least the fundamental knowledge and skills that they will require as they proceed in their schooling and then into 'real' life? Do teachers and schools receive the information in time to make decisions to improve their instruction of the subsequent cohort and to work effectively with the tested group as it moves on to the next grade?
- Assuming a teacher and others do receive information on student assessment results in a clear, thorough, detailed, relevant form and in time to use it meaningfully, and assuming they have the necessary knowledge and skills to make strategic decisions based on the data, do the conditions exist that truly enable them to do this? Is there adequate space in the curriculum or time in the school day and calendar to adjust the pace of instruction for remediation? Can a teacher spend a second lesson on a topic for which the official syllabus anticipates just one if assessment shows the students have not grasped the related content? What would an inspector say if s/he found a teacher who is substituting content from the textbook with other materials to help students learn better a lesson that last year's examination results showed that her/his students had not grasped? Can teachers abridge the syllabus to find time to conduct proper assessments of her/his students, whether it be of learning, for learning or as learning?
- In the same vein, do teachers have the time, resources and authorization needed to conduct their own assessments, both formative and summative? Is there space for assessment and feedback built into the syllabus and school day? Does the system provide models or actual assessment items, tools, rubrics, guidelines and strategies that teachers might consult or use, including standardized elements and instruments? Are there adequate supplies for conducting assessment? Does the protocol for determining a student's accumulated learning to inform decisions about progression, certification and selection take into account the results of assessments performed by teachers? Does it allow for alternative, more profound assessment techniques, such as portfolio review, project appraisal and measures of behaviour and other personal competencies?

The capacity to create and administer learning assessments and to analyze and use the results to improve teaching and learning is fundamental to a system's success in ensuring that students leave school with the knowledge and competencies that society requires of them. As the world moves further into the 21st Century, the nature of these assets is becoming ever more diverse, complex and important, as reflected in the new SDGs and EFA commitments. Achieving this capacity and using assessment not just as a thermometer but to contribute to rich, precise and relevant diagnoses of

what students are and are not learning and how systems can function better to improve these results will require reflection and actions by all education stakeholders. These efforts will concern assessment but also the other dimensions that affect the quality and impact of education, including those controlled directly by the education system and those that are the responsibility of the broader society. The questions immediately above aim to help education system leaders and other actors to join with representatives of the system's various partners to assess assessment (of all sorts) especially for its relevance, the thoroughness of its coverage, and its true utility in making decisions concerning the improvement of quality at all levels of the system. Such a dialogue should affect not just what, how and why assessments happen and how they combine to tell a precise and strategic story about a system's and each of its units' and students' performance, but also how all of these dimensions act, alone and together, to improve education outcomes in the future to move closer to the vision a society sets for itself.

Bibliography

- Barber, M. and Rizvi, S. 2013. *Asking More: The path to efficacy*. Pearson.
<http://efficacy.pearson.com/wp-content/uploads/2013/11/Asking-More-The-Path-to-Efficacy-high-res1.pdf> (Accessed 5 June 2015.)
- Benavot, A. 2012. *Primary School Curricula on Reading and Mathematics in Developing Countries*. Technical Paper No. 8, Montreal, UNESCO Institute for Statistics.
- Black, P. and Wiliam, D. 1998. Inside the Black Box: Raising Standards Through Classroom Assessment.
<http://weaeducation.typepad.co.uk/files/blackbox-1.pdf> (Accessed 5 June 2015.)
- Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., and Krathwohl, D. R. 1956. *Taxonomy of Educational Objectives, Handbook I: The Cognitive Domain*. New York, David McKay Co Inc.
- Brown, G. T. L. 2014. AsTTle: a national testing system for formative assessment: how the national testing policy ended up helping schools and teachers. in Lai, M. and Kushner, S. (ed.) *A Developmental and Negotiated Approach to School Self-Evaluation (Advances in Program Evaluation, Volume 14)* Emerald Group Publishing Limited, pp.39 – 56.
<http://www.emeraldinsight.com/doi/abs/10.1108/S1474-7863%282013%290000014003>
(Accessed 5 June 2015.)
- Cornell-Farrow, S. 17 November 2014. *Concerning Learning: a Review of My School and its Potential Applications to Improving Educational Standards in Developing Countries*. Report prepared for the Office of The Honourable Julia Gillard, former Prime Minister of Australia.
- Educational Initiatives. January 2010. *Student Progress Tracking System*. EI Working Paper Series – Issue 4. <http://www.ei-india.com/wp-content/uploads/2012/07/Student-Progress-Tracking-System-Issue-4.pdf> (Accessed 5 June 2015.)
- Fendler, L. and Muzaffar, I. 2008. The history of the bell curve: sorting and the idea of normal. *Educational Theory*, vol. 58, no. 1; pp. 63-82.
<http://fendler.wiki.educ.msu.edu/file/view/2008+Bell+Curve.pdf> (Accessed 5 June 2015.)
- Gardner, H. 2006. *Multiple Intelligences: New horizons in theory and practice*. Basic Books.
- Green, D. 31 May 2015. *How to Cheat on State Standardized Tests and Not Get Caught*. Education Week: Teacher.
http://blogs.edweek.org/teachers/work_in_progress/2015/05/how_to_cheat_on_state_standard.html (Accessed 5 June 2015.)
- Hill, P. and Barber, M. November, 2014. Assessment Renaissance. Seminar Series Paper No. 239, Centre for Strategic Education. <http://www.cse.edu.au/content/assessment-renaissance>
(Accessed 5 June 2015.)
- Kamenetz, A. 2015. *The Test: Why Our Schools Are Obsessed With Standardized Testing – But You Don't Have To Be*. Philadelphia, PublicAffairs Books. (Listen also to the podcast, <http://www.americanradioworks.org/the-test/>) (Accessed 5 June 2015.)
- Kanjee, A. and Sayed, Y. 2013. Assessment policy in post-apartheid South Africa: challenges for improving education quality and learning. *Assessment in Education: Principles, Policy & Practice*, vol. 20, no. 4.

- Massachusetts Department of Elementary and Secondary Education. 2012. Massachusetts Commission to Develop an Index of Creative and Innovative Education in the Public Schools. <http://www.doe.mass.edu/research/reports/> (Accessed 9 January 2015.)
- Ministry of Education, Ontario. 2010. Growing Success: Assessment, evaluation and reporting in Ontario Schools – for Grades 1 – 12. <http://www.edu.gov.on.ca/eng/policyfunding/growsuccess.pdf> (Accessed 5 June 2015.)
- Muskin, J. Forthcoming. Using Student Learning Assessment Data to Improve Teaching and Learning (working title). Washington, DC , The Brookings Institution.
- Pellegrino, James W. and Hilton, Margaret L., Editors. 2012. Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century. National Academy of Sciences. <http://www.nap.edu/catalog/13398/education-for-life-and-work-developing-transferable-knowledge-and-skills> (Accessed 5 June 2015.)
- People for Education. June, 2013. Broader Measures of Success: Measuring what matters in education. <http://www.peopleforeducation.ca/measuring-what-matters/wp-content/uploads/2013/10/P4E-MWM-full-report-2013.pdf> (Accessed 5 June 2015.)
- Redecker, C. 2013. The Use of ICT for the Assessment of Key Competences. A Joint Research Centre Science and Policy Report, European Commission, Luxembourg <http://ftp.jrc.es/EURdoc/JRC76971.pdf> (Accessed 4 June 2015.)
- Rogers, W. T. 2014. Improving the utility of large-scale assessments in Canada. *Canadian Journal of Education / Revue canadienne de l'éducation* 37:3.
- Stecher, B. and Chun, T. 2001. School and Classroom Practices During Two Years of Education Reform in Washington State. CSE Technical Report 550, National Center for Research of Evaluation, Standards and Student Testing. <https://www.cse.ucla.edu/products/reports/TR550.pdf> (Accessed 5 June 2015.)
- Sternberg, R. 2003. *Wisdom, Intelligence, and Creativity Synthesized*. Cambridge University Press.
- UNESCO. 2015a. Education 2030: Towards Inclusive and Equitable Quality Education and Lifelong Learning for All. <https://en.unesco.org/world-education-forum-2015/incheon-declaration>.
- _____. 2015b. *EFA Global Monitoring Report: Education for All 2000-2015: Achievements and Challenges*. Paris, UNESCO. <http://unesdoc.unesco.org/images/0023/002322/232205e.pdf> (Accessed 25 June 2015.)