

Progressive Student Assessment Practices A Blue Sky Paper (Excerpt on Student Evaluation and Assessment)

Barbara J. Smith

Great schooling needs to do more than simply list expectations with engaging ways to address them. Teachers can often find exceptional activities that align nicely with state, provincial, or national expectations or standards, but such resources tend to be remiss in offering quality assessment of either the formative or summative kind. Assessment is a key component of an engaging curriculum. There are three key assessment forces in schools that address student learning:

- classroom assessment
- report card grading and reporting practices
- standardized test scores

In different ways, each of these forms of assessment can improve and worth the investment to reform them.

Classroom Assessment

Probably the most critical form of assessment is what happens when the teacher gathers evidence of learning in the classroom. This can happen by examining projects, quizzes, interviews, and observations. Some teachers rely on commercially-designed assessment tools that often accompany textbooks, while others make their own tools for assessing learning at the classroom level. Teachers may assess to determine mastery, rank their students from weakest to strongest, or some may do both. Teachers may use an assessment tool for the whole class at the same time (i.e. a scheduled test) or they may wait for a time when each student is ready to be assessed, that is, ready to demonstrate their mastery. While it may be convenient for students to hand things in on the same day, and be tested on one single test day, the reality is that students are not the same, nor do they learn at the same rate. A mastery approach would mean that the idea of a single test day, or assignment due date for all students, would be rare.

Some of the more urgent issues around assessment link to the discussions of penalties for lateness and the awarding of 'zeros'. Much assessment has been side-tracked by 'late' policies in schools. As Gonzalez (2018) points out, there are students who turn in projects a day late only to receive a half-credit based on the teacher's 'no excuses' late policy. Granted that people are required to meet deadlines in the real world, this should not mean that every subject and every assignment in a school must teach lessons of promptness. Handing in things on time is a life skill, and as such, I would argue, that patterns of timeliness should be deliberately taught and measured within a 'Lifeskills' class. I seriously doubt every Science, Math or English teacher has taught the lesson on being prompt, and if it has not been taught – how is it, that it can be assessed? Whether, or not a grade is reduced, the acceptance of late work can lead to backlogs of grading for the teacher. Many educators feel the

"If we teach today as we taught yesterday, we rob our children of tomorrow." ~ John Dewey

need to impose a penalty for lateness, to teach the life skill, but without the actual teaching of the topic of 'promptness', it is not exactly fair to teach a lesson via a penalized grade. Grades need to reflect learning, late to the table or not. Concerned about this practice and the impact of late penalties on student grades, Gonzalez (2018) noted: "...a student who turns work in late can have very low grades, even if they have mastered the content." She added: "In classes where late work is penalized, a grade is a reflection of the student's time management, or of stress, or perfectionism, or dozens of other possible factors. What it isn't, is a reflection of learning."

And what about giving scores of zero? A student misses an exam in college and appeals of the 'sleeping in' nature tend to be met by shrugged shoulders. Administrators, hiding behind the cloak of fairness, shatter student GPA's at the stroke of a policy – and many students, so well brain-washed by the system, believe they deserve the zero scores. Consequently, to all who view the transcript, it is interpreted that the student does not know as much as another student, when it is more about a college professor teaching a student a life lesson. Bottom line, the zero is a reflection, not of what the student has learned, but rather, a fixed mindset about what matters more than the syllabus itself. As Sackstein (2016) suggests: "Don't give up on the kids. Giving a zero is giving up and almost expecting them to do the same." Unfortunately, few challenge such practices, so they continue.

By examining the actual tools for assessment, it is possible to understand how we can make significant changes that support student learning at the classroom or micro level. Teachers gather data from multiple sources including observation, diagnostic, rubrics, and tests. They need to be cognizant of the volume of tasks that students are asked to do, as educators often have more data than they need, often at the expense of exhaustive student work.

What are the costs of improved assessment practices? Once again it is time; time afforded to modify existing assessment tools and time saved by reducing the volume of excess assessment. A detailed look at the most common assessment practices reveals how fresh approaches can lead to more engaging results.

Rubrics - Frequently, rubrics are cited as best practice because such tools function to outline the assessment process for students in advance of their work. Figure 14 reveals a sample rubric adapted from expectations laid out for the Scholastic Aptitude Test (SAT), a popular university test used by many US college admission teams.

"If we teach today as we taught yesterday, we rob our children of tomorrow."~ John Dewey

Figure 14: Adapted SAT Writing Rubric

1	2	3	4
Inadequate	Partial	Proficient	Advanced
The response has little or no cohesion	The response demonstrates little or no cohesion and limited skill in the use and control of language.	The response includes a central claim or implicit controlling idea.	The response is cohesive and demonstrates a highly effective use and command of language.
The response has an inadequate clear central or controlling idea	The response may lack a clear central claim or controlling idea or may deviate from the claim or idea over the course of the response.	The response has a clear central claim or controlling idea,	The response includes a precise central claim.
The response lacks a recognizable introduction and conclusion.	The response may include an ineffective introduction and/or conclusion	The response includes an effective introduction and conclusion.	The response includes a skillful introduction and conclusion.
The response does not have a discernible progression of ideas.	The response may demonstrate some progression of ideas within paragraphs	The response demonstrates a clear progression of ideas both within paragraphs and throughout the essay.	The response demonstrates a deliberate and highly effective progression of ideas both within paragraphs and throughout the essay.
The response lacks variety in sentence structures; sentence structures may be repetitive.	The response has limited variety in sentence structures; sentence structures may be repetitive.	The response has variety in sentence structures.	The response has a wide variety in sentence structures.
The response demonstrates general and vague word choice; word choice may be poor or inaccurate.	The response demonstrates general or vague word choice; word choice may be repetitive.	The response demonstrates some precise word choice.	The response demonstrates a consistent use of precise word choice.
The response may lack a formal style and objective tone.	The response may deviate noticeably from a formal style and objective tone.	The response maintains a formal style and objective tone.	The response maintains a formal style and objective tone.
The response shows a weak control of the conventions of standard written English and may contain numerous errors that undermine the quality of writing	The response shows a limited control of the conventions of standard written English and contains errors that detract from the quality of writing and may impede understanding.	The response shows a good control of the conventions of standard written English and is free of significant errors that detract from the quality of writing.	The response shows a strong command of the conventions of standard written English and is free or virtually free of errors.

To examine such a rubric with the task of improving it, begs many questions:

- Is it necessary to repeat “The response” thirty-two times?
- What does it mean to “have use and control of language”?
- What is an “implicit controlling idea”?
- What is the difference between a “clear” and “precise” central claim?
- What is the difference between an “effective” and “skillful” introduction/ conclusion?
- What is the difference between a “clear progression of ideas” and a “deliberate and highly effective progression of ideas?”
- What is the difference between a “variety of sentence structures” and a “wide variety in sentence structures?”

"If we teach today as we taught yesterday, we rob our children of tomorrow." ~ John Dewey

- What is the difference between a “lack of formal style and objective tone” and “may deviate noticeably from a formal style and objective tone”?
- Why are multiple expectations mixed together? (i.e. “strong command of conventions of standard written English with, virtually free of errors”)

If rubrics are to be fair and support learning, they must be clear. Think about how you might revise the SAT Writing Rubric so that teachers and students would be able to replicate more consistent assessments of the same writing piece.

The value of rubrics can also be limited when the jargon and often lengthy list of expectations within a single rubric block, make it difficult to use rubrics effectively. It can also be complicated for students and teachers when there are so many quality parameters, such as those used for the SAT writing rubric: ‘inadequate’ (1), ‘partial’ (2), ‘proficient’ (3), and ‘advanced’ (4). How can it be “adequate” to demonstrate ‘partial’ learning?

Gonzalez (2014a) admitted she did not read all the cells on rubrics, “having been on the receiving end of multi-page, multi-cell rubrics stuffed to the gills with 9-point font.” She noted: “I looked at the third and fourth columns, where *expectations met* and *exceeded expectations* were described, and I did everything I could to make my work satisfy those criteria. The other two columns got little more than a glance.” So why do rubrics include descriptions of degrees of less than adequate work?

Often a piece of writing judged by different teachers can elicit different scores. On a typical 4-point scale, regardless of the descriptors, just as many teachers will allocate a ‘2’ score as a ‘3’ score, indicating that the writing is not exceptional, but not weak either. A 4-point scale introduces more levels of subjectivity. Students are less trusting of a tool that has a 50% chance of judging their work as a “2” or “3”. Teachers should also be less trusting of such a tool. According to Ardivino, Hollingsworth & Ybarra (2000), “the percentage of rate disagreement increases as the number of distinctions increases...a lower number of performance levels is more effective” (pp. 25 & 26). A piece of writing, when viewed on a three-point scale, might score a ‘3’ if it stands out as exceptional, a ‘2’ if deemed average/proficient, and a ‘1’ for partially meeting the expectation. Multiple viewers would usually yield more consistent results using such quality parameters. Clarity of expectations means that students understand a task “in terms of volume and quality of work.” Figure 15 is example of a simplified inquiry project rubric.

Figure 15: Inquiry Project Rubric

Self-Score I....	PERSONAL COLLEGE PROJECT Rubric (2 = “mastered”; 1 = “started”; ? = “no evidence”)	Teacher Score You...
	documented with rationale, initial perspectives about best colleges	
	explained reason for narrowing focus on specific programs at college	
	gathered data about programs in 30 colleges or more from web sites and other sources	
	gathered data from a balance of local and out-of-state or country universities	
	completed a mock Common Application and made a list of areas of strengths	
	made list to help improve experience in advance of submitting Common Application	
	interviewed or surveyed at least one student in college	

"If we teach today as we taught yesterday, we rob our children of tomorrow." ~ John Dewey

	analyzed the strengths and weaknesses of existing college ranking systems	
	critiqued website and marketing of researched colleges	
	generated a personal ranking scale with clearly defined categories to support decisions for choosing a college for self or others	
	prepared solid argument for college ranking with data/details to support ranking categories	
	explained shift in understanding of college choices from beginning to end of college inquiry	
	generated suggestions for filling in college applications (including Common Application)	
	shared narrative paragraph about own personal college choices	
	shared limitations of inquiry (what could have been improved?)	
	shared Personal College Project findings with classmates using PowerPoint effectively	
	gathered evidence of classmates learning following presentation	
	worked independently on each part of inquiry project	
	asked for help and/or supported others (during the inquiry)	
	asked questions and took notes and stayed on task during presentations	
	<i>Total</i>	<i>/80</i>

Such a tool was designed to support student mastery. Different than a single point rubric/checklist with a simple pass (or not) distinction, the simplified checklist rubric, as in the case of the Inquiry rubric example (Figure 15), can offer differentiated scores. In this way, students can be rewarded for demonstrating degrees of mastery work. As Copp & Smith (2010) noted the clarity of assessment tools “should serve to motivate students to demonstrate both mastery and creativity” (p. 83).

Educators need to have a clear idea of what constitutes mastery, so assignments, tasks and tests can be designed with a fair balance of expectations. Given that there is a base of knowledge that can support more complex understandings, it makes sense to weave together assessment criteria that blends varied levels of complexities.

The reality is that rubric tables, often filled to the brim with word density, repetition, and subjectivity, line many assessment sections of textbooks, regularly used by teachers in many classes. I could not find, however, research that confirmed, whether (or not), students read rubrics, or furthermore, used rubrics as a tool to improve their work.

A new approach to creating rubrics can provide an efficient space for students to weigh in on their work. The best practice research provides a healthy selection of studies that confirm the value of self-assessment activity: Black & Wiliam, 1998; Orsmond, Merry & Reiling, 2002; Brookhart, Andolina, Zuza & Furman, 2004; Andrade & Valtcheva, 2009; Bryant & Carless, 2010; and Dyer, 2014. It is rare, however, to find combined teacher and student assessment tools. Stiggins (2008) claimed: “Turn learners into assessors...Adults aren’t the only ones who assess students” (p. 4).

By viewing outlier approaches to rubric design, it is possible to see how such tools can evolve and support student learning and thus contribute to school improvement. No question, many templates and evaluation worksheets that exist today are good, but paying serious attention to how we can make them great is how we can sharpen our tools to make them better.

Classroom Tests - While educators need to critique rubrics, and have the capacity to revise them, they also need to know how to examine and create classroom tests. Teachers need to be sure of the intent of a test. Should it be to find out what students know, or what they don’t know? Are ‘applied’

"If we teach today as we taught yesterday, we rob our children of tomorrow."~ John Dewey

questions, really applied, or are they items that have not been taught? And what is a fair way to address the question of re-testing?

Referring to the idea of giving extra chances to retake an exam, Schwartz (2017c) noted that one high school teacher required students “to go over the first test with him and discuss study strategies before they set up a retake date” (pp. 7-8). In a different high school, another teacher prepared a world history test. When he discovered many students did not do well on the World War I section, he asked me if I thought it was okay if he re-tested those students (who did not achieve mastery) on that section. I said “hell, yes!” This teacher understood that while students had scored 80% or more on the other sections of the test, this test, for some students, could provide both formative and summative feedback. The students with weaker understandings of the First World War were not given a free pass to move on to the next area of study. The ‘not yet’ portions of the test (less than 80% score), would be formative until they could reach mastery of all expectations. Unfortunately, it is rare for teachers to hand back tests for re-testing purposes. They often work with formulas that build-in reductions for second or more attempts. Wormeli (2012) shared the story of Martin, who was given a second chance after first receiving an F grade:

After re-learning and a lot of hard work, the second attempt results in an A. We trust the exam to be a highly valid indicator of student proficiency in the subject, and Martin has clearly demonstrated excellent mastery in the subject. When the two grades are averaged, however, the teacher records a C in the grade book—falsely reporting his performance against the standards.

The score reveals, not what the student knows, but some random formula that blends the element of speed into the mix. From a student’s perspective, it can feel pointless to try again, when past mistakes are factored in to the final score. When students take longer to achieve mastery, it can often be challenging for educators to determine a score. However, Wormeli (2012) reminded us that the next generation of students could be liberated from the “oppression of averaging.” He noted: “We’re working with real individuals, not statistics. Our students have deeply felt hopes and worries and wonderfully bright futures. They deserve thoughtful teachers who transcend conventional practices.”

Ideally teachers will design tests before the beginning of a unit, and then revise their tests and quizzes along the way to align with classroom experiences. When students sit down to write a test, there should be no surprises. Teachers should complete the test in advance, to have an idea of the timing, question volume, blending of challenging and more challenging expectations, as well as ‘fair’ checking to determine if what is being tested has been taught. Classroom tests should take time to grade, and ideally, they should include open-ended questions that prompt students to explain their understandings more fully than a standardized multiple-choice type test. Testing for mastery means that students will take the tests when they are ready, which means a teacher may need to have a bank of tests handy. Having multiple versions of a test can support differentiation by permitting students to work at their own pace.

Observation and Day-to-Day Learning Evidence -When teachers design various tasks for students to work on in their classrooms, they can observe how students are setting and meeting learning goals.

"If we teach today as we taught yesterday, we rob our children of tomorrow."~ John Dewey

Ideally, tasks are not developed in pockets of short time spans <less than 60 minutes>. Some extended projects can take several days or weeks to complete. A teacher who conferences with students through the various stages of their work, can document, adjust, and guide learning. Keeping track of what students say in a conferencing scenario, by jotting down notes, or when listening in on group or paired activity, can be powerful sources of data. Given that students learn in different ways, and can demonstrate their understandings in a variety of contexts, we need to pay more attention to teacher observation as a source of evidence of learning. Wall (2011) claimed: "As part of the ongoing recording and monitoring system within early years' settings, the usefulness and power of observation and assessment are sometimes overlooked as they may be deemed time-consuming" (p.86). While observational data is typically part of the primary school teacher's toolkit, it is often omitted by teachers of older students. Classroom observations can be a significant source of data for students of any age, but the gathering and interpreting of such material should be systematic and supportive. Effective observation practices should include a plan to ensure all students have similar opportunities to demonstrate their understandings. Gonzalez (2018) recognized that "grades are inherently imperfect." She added:

To truly assess our students' learning, we need to get to know them, observe them, and study a wide sampling of their work over time. When we reduce all that to a single measurement for the sake of efficiency, we lose that bigger picture (p. 11).

Quality observations should require students to explain their ideas or answers on a quiz. As Lynch (2017b) suggested: "While it may not be as efficient to grade answers that go beyond filling in a bubble, these are the questions students need to answer to apply their knowledge in real-world applications". Seattle Schools effectiveness consultant, Johnnie McKinley (1999), claimed that classroom-based observations are "much more valuable than what can be discerned by a standardized test, which offers only a snapshot." Educators who understand and use a full range of assessment tools can see how their teaching influences learning.

Diagnostic and Transformative Data Collection - According to the OECD, *Quality and Equity* document (2012), the use of diagnostic tools can lead to improved learning. While most educators agree that diagnostic testing before students begin a unit or project is a good thing to do, students rarely take part in such a practice. They may exist at the beginning of units, for instance in Math textbooks, but often such initial benchmark data is not collected. A few teachers build their own diagnostic tools, but such practice is often spotty.

Having students understand the goal of finding out what they do not know can help teachers reduce repetitive teaching and allow them to concentrate on what each student needs to master. A well-designed diagnostic tool can help teachers develop individual programs that support differentiated instruction, teaching that is tailored to each student. Diagnostic testing is rarely shared with students as a goal setting tool and seldom revisited at the end of the unit to discuss students' new understandings. As teachers and students become more adept at action research, I sense they will become more comfortable with using before and after data. In this way, they can celebrate learning as a transformation.

"If we teach today as we taught yesterday, we rob our children of tomorrow." ~ John Dewey

When elementary school boys were asked *'what is inquiry?'*, prior to beginning their third, fourth and fifth grade inquiry projects, the students typically responded with two single words: "research" and "questions." At the end of their project, most responded with extensive descriptions featuring: "seek/ explore, plan, gather, observe, analyze, graph, conclusions, identify bias, recommendations, and ask new questions." The students were expected to transform their original understanding. By comparing their before and after responses, they could also see and articulate how their understanding transformed. Students did learn to be responsible for building on their own knowledge; they were not empty vessels to be given facts from afar. Such assessment also provided some key data for examining if the teaching experience, in between, had an impact on learning.

Reporting Practices and Report Card Grading

There tends to be fewer surprises when parents have access to assessment on a regular basis. Typically, schools will distribute report cards three or four times a year, but the conferencing is often scheduled weeks later, losing the value of immediate response and support for improved immediate action. Often schools limit the time for conferencing to five or ten minutes, which makes the experience somewhat like *going through the 'communication' motions* on a speed date. As Copp and Smith (2010) note: "It is difficult for parents and teachers to have two-way conversations in a five-minute time frame" (p. 91).

Some progressive schools support student-led conferences when report cards are distributed. In this way, the teacher has already conferenced with the student to help them plan their presentation for their parents. In this way students, parents, and teachers can be on the same page, aware of 'next steps' for learning in a timely manner.

The grading on report cards generally refers to the overall synthesis of multiple assessment evidence. Student work over a term or semester tends to be compiled in a series of grades recorded as evaluations that communicate achievement at a given point in time. While experts agree that the timed curriculum is not equitable, few people question the finite or fixed nature of the traditional report card. Students who grasp concepts quickly, tend to be rewarded with higher grades or percentages. There is a general acceptance that students do learn at their own pace, and that teachers need to differentiate their teaching to help all students learn. However, when the report card is not differentiated, and not fluid between grades, students who need more time will probably continue to be assessed with scores populating the bottom curve of the class. According to Adams & Masters (2017): "inequity comes about when teachers ignore "the uniqueness of individuals." They added that it is a mistake "when we infer an individual's learning needs from their age or year level" (p. 6).

At the classroom level, summative evidence of student learning is used to populate the student report card. Traditional reports list subjects with a space for a grade or percentage representing achievement; some reports have room for 'effort' scores, as well as a comment section to further describe the rationale for the grade. Teachers spend considerable time preparing these documents for students and parents. Some on-line systems for tracking grades can make the tabulations less cumbersome, but the writing of commentary tends to take up a significant investment of time. The consolidated report, with one grade per subject, may save room in file drawers, but its effectiveness at moving students to improved learning action has not been, in my

"If we teach today as we taught yesterday, we rob our children of tomorrow."~ John Dewey

experience, fruitful. Common reports that are designed and controlled at the state, provincial, or national level, tend to follow such a format. While it may seem commonsense, that is, that all reports should look the same, there is no comparative research that indicates that such efforts to streamline reports have resulted in any improvements, whatsoever. It seems that moving towards a 'fixed' way of reporting has somewhat consolidated a dated and conventional form of reporting. Wormeli (2012) claimed that: "Average," "above average," and "below average" are norm references, but in today's successful classrooms, we claim to be standards - (outcomes) based." He added: "Thankfully, many schools are moving toward disaggregation in which students receive separate grades for individual standards...No more mind games; students have to learn the material."

Standards-Based Grading (SBG) can provide a progressive alternative to conventional school report cards. According to teacherease.com, "SBG breaks down the subject matter into smaller 'learning targets.' Each target is a teachable concept that students should master by the end of the course." The emphasis on specific skills helps students and parents know exactly what essential skills have been mastered, or have yet to be mastered. Grading for SBG reports is based on a demonstration of mastery. Often the SBG scales are illustrated by a number rating: "1's indicate that students have little understanding of a concept, and consequently cannot demonstrate any mastery...As students learn, they can demonstrate partial mastery, and score 2. Once they meet a target, they score 3" (teacherease.com).

Just as students and teachers embrace change in the quest for school improvement, so too, should we accept the need to change and improve reporting practices. At Accelerated Schools in Denver, students receive a daily report card to ensure a "close and constant screening of the student's progress toward fulfilling identified needs":

Students are graded by letter grades, A through C. No failing grades are given. If a student has not mastered a subject adequately, an "incomplete" is given until the student meets the mastery requirements of the subject. Daily grades are based upon the average of the day's reports (acceleratedschools.org).

Similarly, in several charter and independent elementary and secondary schools, we had the autonomy to develop in-house report cards. Rather than offer the traditional range of options, 'A-B-C-D-F' or the numerical '1-2-3-4' designation, we developed a three-layer approach to assessment. The three layers broke down assessment of learning into novice, apprentice, and expert levels. In a high school setting, the novice level assessment was indicated with an 'I', referring to incomplete. More recently, we have used 'NY', for "not yet", relating to Dweck's (2006) growth mindset work. Students, who received NY, meant they needed more time to master understandings. The goal in all subject areas was to move past the novice stage to one of two levels of mastery:

- A Level = demonstrated exceptional skills and understanding (at least 90%)
- B Level = demonstrated excellent skills and understandings (between 80-89%)

"If we teach today as we taught yesterday, we rob our children of tomorrow."~ John Dewey

All students who did not demonstrate at least an 80% understanding were given additional support. Often students with 'I's' or 'Not Yets' were asked to revise their project work and re-write sections on tests to make sure they had every opportunity to achieve mastery.

Even though there is “no such thing as a standard mind” (Rose & Ogas, 2016, p. 9), assessment can distinguish qualities of achievement. Mastery does not have to mean pass/fail; it can reveal levels of excellence. Anything less than a substantial level of learning requires more work, mimicking real expectations in the workplace. There is talk of a shift in how businesses view college transcripts: “Over the past few years, Google and Deloitte adopted individual-focused methods for evaluating employees” (Rose & Ogas, 2016, p.7). Referring to Molenaar’s ‘jaggedness principle’, they suggested that “all human qualities that educators and employers care about — including intelligence, character, talent, and potential — are multidimensional and cannot be reduced to a single score, type, or rank” (Rose & Ogas, p. 7). They further claimed:

There is no “one-size-fits-all” classroom format...Our standardized system of education reinforces the view that learning is like climbing the rungs of a ladder: Every student must climb the same rungs in the same order...to ascend to the same destination (the diploma). Colleges even penalize students who climb a rung at a slower pace than average...it usually makes no sense to rank students on how quickly or slowly they advance (Rose & Ogas, 2016, p. 8).

At a Science conference dedicated to increasing the status of STEM in Canadian schools, I was surprised by the lack of attention paid to assessment, particularly summative grade scores. There are some STEM teachers, who see a range of scores as compelling; in other words, a course may be considered rigorous when there are significant numbers of students whose grades fall below 80%. Establishing such an evaluation culture within STEM disciplines can act as a serious deterrent to learning and engagement. It is highly likely that students avoid coursework that has a high probability of leading to poor grades. When students are aware that their grades do not reflect mastery, this can also limit their confidence in STEM subjects. Poor grades could probably be the most significant reason why many students opt out of STEM programming at the earliest possible point in their high school career. Simple surveys that give students a chance to say why they chose or did not choose, STEM courses, would certainly support or negate such a claim. What seems common sense is that students stay in school and continue to take classes that they are successful in. My sense is that the reason for a poorly STEM-educated society might have more to do with the lack of encouragement in terms of grades, than simply a lack of hands-on experiences or Science role models. Science in the real world is a profession, where failure can lead to success, yet the assessment in schools seems to be more about weeding out kids who make mistakes, and have difficulty digesting facts quickly.

The cost of changing reporting practices comes down to the will to change the purpose of assessment. Should we assess to rank or gather evidence of mastery? Teachers, students and parents all value assessment, so it matters that we strive to do assessments better. Changing a report card can be costly, but such changes are long overdue. The system of evaluating student understandings must be revisited. Teachers and students need to be responsible for moving C, D, or less scores to a minimum of 80% understanding. Systems that allow the escape routes of lesser scores are fueling the increased

"If we teach today as we taught yesterday, we rob our children of tomorrow."~ John Dewey

perception that certain subjects are only for 'others'. Whether it is the Ontario 1-2-3-4 labels, percentage options, or letter grades, the option to pass with less than a solid base of understanding is indeed a serious barrier to building an educated society.

Standardized Testing

Lynch (2018) questioned school cultures where students work inside individual test packets “filling in bubbles on scan sheets with sharpened pencils.” He added: “the tests are virtually the same boring layout that they were when many of us took our standardized tests as K-12 students.” Students can be exposed to standardized tests designed to gather comparative data at the classroom, school, district, state, national or international levels, but making such testing the central goal of education should not be the emphasis in schools.

The SAT and the ACT (American College Testing) are probably the most familiar standardized tests that are typically used to gain university entrance. Both tests cover a broad range of questions, but one notable difference is that the ACT does not subtract wrong from right answers. While both test results function to help school admission teams send out their offers of acceptance, they are not a comprehensive measure of student achievement. Most college admission departments indicate that standardized test scores are only one metric they consider. Many even suggest that there are “no hard and fast rules” (Yale Admissions) as to what a student needs to score to be considered for entry. The following represents the range of scores that were shared on Yale’s Admission web page:

There are no score cutoffs for standardized tests, and successful candidates present a wide range of test results. During the most recent year, test score ranges (25th to 75th percentiles) for enrolled freshmen were:

- SAT-Verbal: 710-800
- SAT-Math: 710-790
- SAT-Writing: 720-800
- ACT: 32-35

The tracking and publishing of such scores indicated that standardized tests matter. Even though college admission teams claim that they only view the highest scores when student repeat these tests, many ask students to submit all their scores for each attempt. It baffles me how students in their senior high school years can maintain their sanity amidst the tsunami of tests and re-tests that fill their calendars. These tests function as encyclopedic score cards that serve as a screen to differentiate who can squeeze as much information as possible into their brains. Our minds should be much more than storage devices.

The collateral damage of high stakes testing affects students in different ways. The pressure to achieve can be so overwhelming for young people. Kamenetz (2017) was concerned about close to three million stressed-out high school students taking the AP exams. She noted that such courses are perceived as “a mark of the aspirational, a promise of higher standards and, occasionally, a more expensive alternative.” The claim that such coursework offers a higher end education has yet to be

"If we teach today as we taught yesterday, we rob our children of tomorrow."~ John Dewey

proven: "remarkably little independent research has been conducted on the academic benefits of AP" (Warne, 2017). Regardless of undocumented evidence of benefits, there is little doubt the business of testing is a lucrative one. Referring to the recent boom in AP popularity, Kamenetz (2017) added: "if you spend \$93 to take the exam, plus often hundreds of dollars for textbooks and lab fees, they may be exchangeable for college credit... Participation doubled in the last 10 years..."

Standardized testing happens in different kinds of settings. On a global scale, the Program of International Student Assessment (PISA) conducts tests every three years. Upwards of half a million students take this two hour survey that evaluates education systems worldwide. According to the PISA, the U.S. continues to lose ground. In 2012, PISA results ranked the U.S. below Finland, various European nations, Asian countries, Canada and Australia (Lurie, 2013). By 2015, the U.S. ranked 31st among the 35 Organization for Economic Cooperation Development (OECD) industrialized member nations (Barshay, 2016), putting the U.S. in the bottom 11.4% of these nations. According to Lynch (2017): "We know that American students lag behind other developed countries when it comes to math and science achievement." His book highlights additional studies that examine such weak trends in US education:

South Korean and Singapore students consistently outrank U.S. students in basic and advanced Math and Science course achievements	Rich (2012)
Business leader surveys concerned about their employees' lack of basic writing and communication skills	Nolop (2013)
27% of US 12th graders were proficient in writing (based on 2011 National Assessment of Educational Progress)	Leal (2012)

At the national level in the US, the *Partnership for Assessment of Readiness for College and Careers* (PARCC) testing was developed to see how to make improvements in the *Common Core State Standards*. When such initiatives are tied to funding, it can lead schools to worship a 'practice testing' model of curriculum. According to Lynch (2016a), "Teachers find themselves stuck in a no-win situation ... where they must reach certain benchmarks to receive recognition and funding". He added, "they have to forsake learning in wider scopes to make it happen." Given that testing and practice testing have engulfed and re-shaped the curriculum in a way that over-rides greatness, the necessity of such practices must be challenged. As Lynch (2017b) claimed: "For many teachers, the way that they want to teach and the way that they are forced to teach vary greatly, and much of that is due to unreasonable accountability standards that include student performance on standardized tests."

In Alberta, Canada, *Student Learning Assessments* were introduced to replace their *Provincial Achievement Test* in 2015: "The assessments aim to provide students, teachers, and parents with information at the beginning of the school year to be used as a reference to plan and support student learning during the year" (OECD, p. 13). Conducting such tests at the beginning of the school year differs from the conventional state-wide tests that tend to happen at the end of the academic year.. What data can be gleaned from standardized tests, however, should be viewed for what it is, a slice of what learning is going on in classrooms, not the full picture. Nevertheless, the data, if used with integrity, can help determine where to invest resources and supports for learning.

"If we teach today as we taught yesterday, we rob our children of tomorrow."~ John Dewey

Community members and educators run the risk of viewing the results of standardized slices as actual full specimens of learning, when such results represent a small cell of a much larger learning organism. As Lynch (2017c) claimed: “Not all students are natural-born test takers,” so to assume that standardized tests have captured the learning is simply naïve. Schwartz (2014) concluded that new tests “don’t measure effective oral communication, collaboration, learning how to learn or develop academic mindsets.”

The costs of writing, printing/or preparing software, and administrative tests are costly, especially on an annual basis. The investment in such testing practices should be challenged. It can cost hundreds of thousands of dollars for school districts to opt for annual testing. It would be well worth it to consider the viability of testing every three to five years if systems must adhere to such practices. Making it less permanent on an annual basis would also slow down the grind of having practice tests bleed so far into the fabric of teaching. There is little time to learn when students are being tested so much. While the practice testing business that brings in billions in revenue would be cut back, interruption of deep learning would not be so widespread.

An engaging curriculum integrates content, skill, and understanding expectations with assessment. An education system with a rudder firmly focused on high stakes testing, misses the mark on deep learning and preparing young people to be innovators in the future. As Fullan & Langworthy (2014) so elegantly claimed: “whole system change still faces significant barriers in most places. These barriers reside primarily in the student assessment, teacher evaluation and school accountability regimes that currently define success for our education systems.” They further suggested: “Until we find new ways to define and measure success – ways that measure schools’ adoption of new pedagogies and students’ achievement of deep learning outcomes – crucial system factors will stand in opposition to innovation” (p. 3). There are many strong voices who challenge the over-reliance of testing. Lynch (2016b) claimed: “Assessments turn living, breathing students into machines, of sorts, who must be programmed to spit out the right answers at the right time.” Sornson (2017a) pointed out that many systems have failed to improve scores, yet what remains is the push toward “more standardized, superficial, fragmented and meaningless learning” (p. 4). In another article, he noted, “We deliver more content than is reasonable in the time allowed, hoping that by ‘covering’ the standards that might be included on the mandated assessments our students will get better test scores” (2017b, p. 2). In his view:

We have stripped away much of the sense of community, joy, play, and social learning from our schools as we hold onto an archaic educational system that was never designed to help large numbers of students become quality learners for life (p. 2).

How it is that standardized tests are not subject to such accountability for improvement that their results impose on schools, on teachers and on students? Is it not time to turn the tide and look at the standardized testing process as part of a much larger view of what makes a school great? If we are aiming for quality, then we need to ensure that qualitative metrics are understood as far more accurate data than what the current test score ‘slices’ can reveal. Lynch (2016a) adds two questions to the debate: “How can all students be measured with the same yardstick – and how can punishments and rewards be handed out using such a scale?” with a significant conclusion: “standardized

"If we teach today as we taught yesterday, we rob our children of tomorrow."~ John Dewey

assessments are flawed.” He revealed the following issues embedded in the blind acceptance of standardized testing:

- Inadequate sampling of material being tested;
- Indirect, rather than direct, observation of what a student is truly learning;
- Too narrow a scope of knowledge;
- Not enough exceptions made for regional/cultural differences within a state;
- Too many lasting inferences made about the students taking the tests that are based on very little merit;
- Too much emphasis on a punishment mentality, and not enough on what can actually be improved; and,
- No accounting for socioeconomic or disadvantaged barriers that hinder a teacher’s potential.

Relying on a single cell of an organism to shed light on the entirety of school learning, and using that data to define and compare greatness, is frankly sloppy work. In the long run, it can only serve to diminish a society’s intellectual return. An investment of time to garner a true picture of learning needs to be the priority. Recognizing the need for balanced assessment, Stiggins (2008) noted: “We have come to a tipping point in American education when we must change our assessment beliefs” (p. 1). Concerned that decades of obsessive reliance on standardized tests has not been enough, he added, “The time has come to acknowledge the extreme limitations of these tests” (p. 2).

Students and teachers often fall into the trap of assessment, which rarely addresses how students behave like experts; rather, the pencil and paper, or now, on-line tests, tend to be all about how much knowledge can be recounted within a multiple-choice test designed to trick students with two stronger answer choices. In Science courses, for instance, where are the portfolios of lab experience? Where is the evidence of habit mastery? To prepare for Sciences in college, many grade 11 and 12 Science classes try to emulate the college experience – that is, a highly dependent focus on content testing. Many students, who cannot absorb the content, can be frustrated by the trickster testing and drop such courses, deemed as more rigorous, as lower potential Science grades can limit their chances of getting into the university of their choice. Even if they struggle through the dense high school senior Science curricula, they are met with an unreasonable amount of content in their first year.

Why are universities not recognizing that mastery approaches can lead to a much more educated scientific culture? It is simply a mindset that accepting fewer students exemplifies rigor. Just as journal editors can punch out the number of rejections, assuming this means their publications are better, many universities herald their low percentages of acceptances, as some sort of status symbol.

How can leadership at the post-secondary level become more aware and think more deeply about assessment? Even though the target of college acceptance plays a significant role in the way secondary and elementary schools are designed, we can improve schools by integrating what students may need after college, in work and life. How we assess students reflects our teaching values. The capacity to see beyond the sameness of existing assessment tools can contribute to building a culture of improvement. It takes time to consider other assessment possibilities, time to adjust current practices and tools, and time for studying the impact of assessment tools on student learning. Time

"If we teach today as we taught yesterday, we rob our children of tomorrow."~ John Dewey

for such a shift in thinking requires significant professional learning funding, to help educators see and embrace alternative perspectives with regards to assessment practices.

Grappling with Ideas

This chapter on assessment is a lengthy read, but given the depth and breadth of issues surrounding the development of quality assessment tools and the range of ways assessment can be implemented, it is clearly a key part of curriculum, and worthy of significant change. Quality classroom assessment is complex, and therefore takes time to unpack. Begin grappling with some of the following questions:

- How often should we assess?
- Should students be afforded zero scores or be docked for lateness?
- Should we assess student leadership or co-curricular involvement?
- How much should we invest in testing? Can we save on assessment costs?
- Review this chapter and make a list of assessment practices you currently do, and a list of ideas for improving those practices.

References

Accelerated Schools, Denver Retrieved from <http://www.acceleratedschools.org/academic-programs>.

Adams, R. & Masters, G. (2017). The school curriculum: about time. *Teacher Magazine*. June 26, 2017. Retrieved from <https://www.teachermagazine.com.au/columnists/geoff-masters/the-school-curriculum-about-time>.

Andrade, H. & Valtcheva, A. (2009). Promoting learning and achievement through self-assessment. *Theory into Practice*. October 7, 2009, 48:1, 12-19. Retrieved from <https://doi.org/10.1080/00405840802577544/>.

Ardivino, J., Hollingsworth, J. & Ybarra, S. (2000), *Multiple Measures: Accurate ways to assess student achievement*. Thousand Oaks, CA.: Corwin Press.

Barshay, J. (2016). U.S. now ranks near the bottom among 35 industrialized nations in math. *The Hechinger Report*. December 6, 2016. Retrieved from <http://hechingerreport.org/u-s-now-ranks-near-bottom-among-35-industrialized-nations-math/>.

Black, P. & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappa*. October, 1998. Retrieved from <https://www.rdc.udel.edu/wp-content/uploads/2015/04/InsideBlackBox.pdf>.

Brookhart, S.M., Andolina, M., Zuza, M. & Furman, R. (2004). Minute math: An action research study of student self-assessment. *Educational Studies in Mathematics*. September, 2004, Vol. 57, Issue 2 (pp. 213-227).

Bryant, D.A. & Carless, D.R. (2010). Peer assessment in a test-dominated setting: Empowering, boring or facilitating examination preparation. *Educational Research for Policy and Practice*. February 2010, Vol. 9, Issue 1, pp 3-15).

Copp, J. & Smith, B (2010). *Mining for Gems: A Casebook of Exceptional Practices in Teaching and Learning*. St. Catherine's: Canadian Educational Standards Institute.

"If we teach today as we taught yesterday, we rob our children of tomorrow." ~ John Dewey

Common Core State Standards. Retrieved from <http://www.corestandards.org>.

Common Goal Systems. Retrieved from teacherease.com.

Dweck, C. (2006). *Mindset: The new Psychology of Success*. New York: Ballantine Books.

Dyer, K. (2014). Proof that student self-assessment moves learning forward. NWEA. March 20, 2014. Retrieved from <https://www.nwea.org/blog/2014/proof-student-self-assessment-moves-learning-forward/>.

Fullan, M. & Langworthy, M. (2014). *A Rich Seam: How New Pedagogies Find Deep Learning*. Pearson, February 22, 2015. Retrieved from https://issuu.com/splinterdesign/docs/pearson_rich_seam.

Gonzalez, J. (2014a) Your rubric is a hot mess: Here's how to fix it. *Brilliant or Insane: Education on the Edge*. October 9, 2014. Retrieved from <http://www.brilliant-insane.com/2014/10/single-point-rubric.html>.

Gonzalez, J. (2018). "How accurate are your grades?". *Cult of Pedagogy*. April 22, 2018. Retrieved from <https://www.cultofpedagogy.com/accurate-grades/>.

Kamenetz, A. (2017). 4 Things We Don't Know About AP Tests. *nprEd*. May 1, 2017. Retrieved from <https://www.npr.org/sections/ed/2017/05/01/525073237/four-things-we-dont-know-about-ap-tests>.

Leal, F. (2012). Report U.S. students lack writing skills. *The Orange County Register*. September 14, 2012.

Lurie, S. (2013). Why doesn't the constitution guarantee the right to education? *The Atlantic*. October 16, 2013. Retrieved from <https://www.theatlantic.com/education/archive/2013/10/why-doesnt-the-constitution-guarantee-the-right-to-education/280583/>.

Lynch, M. (2016a). Ineffective assessments, part II: Where assessments stand today. *The Edvocate*. December 28, 2016. Retrieved from <http://www.theedadvocate.org/ineffective-assessments-part-ii-assessments-stand-today/>.

Lynch, M. (2016b). Ineffective assessments, part III: Why Common Core fails. *The Edvocate*. December 29, 2016. Retrieved from <https://www.theedadvocate.org/ineffective-assessments-part-iii-common-core-fails/>.

Lynch, M. (2017). *Understanding Key Education Issues: How We Got Here and Where We Go From Here*. Routledge.

Lynch, M. (2017b). Ineffective assessments, part V: More critical thinking needed. *The Edvocate*. January 2, 2017. Retrieved from <http://www.theedadvocate.org/ineffective-assessments-part-v-critical-thinking-needed/>.

Lynch, M. (2017c). Ineffective assessments, Part VII: Better cultural and learning understanding. *The Edvocate*. January 4, 2017. Retrieved from <http://www.theedadvocate.org/ineffective-assessments-part-vii-better-cultural-learning-understanding/>.

Lynch, M. (2018). Why all assessments should be digital. *The Tech Edvocate*. March 12, 2018. Retrieved from <https://www.thetechedvocate.org/why-all-assessments-should-be-digital/>.

Nolop, B. (2013). Our college graduates can't write! *Wall Street Journal*. October 9, 2013.

OECD (2012). "Equity and quality in education: Supporting disadvantaged students and schools. Spotlight Report: Netherlands. OECD Publishing. Retrieved from <http://www.oecd.org/education/school/49603617.pdf>.

"If we teach today as we taught yesterday, we rob our children of tomorrow." ~ John Dewey

- OECD (2015). Education Policy Outlook, Canada. January 2015. Retrieved from <http://www.oecd.org/education/EDUCATION%20POLICY%20OUTLOOK%20CANADA.pdf>.
- Orsmond, P., Merry, S. & Reiling, K. (2002). The use of exemplars and formative feedback when using student derived marking criteria in peer and self-assessment. *Assessment & Evaluation*. Vol. 27, 2002, Issue 4.
- Partnership for Assessment of Readiness for College and Careers (PARCC). Retrieved from <https://parcc-assessment.org>
- Rich, M. (2012). U.S. students still lag globally in math and science tests show. *New York Times*. December 11, 2012.
- Rose, T. & Ogas, O. (2016). The faulty foundation of American colleges. *Medium*. Retrieved from <https://medium.com/@ltoddrose/the-faulty-foundation-of-american-colleges-29312f7e7158>.
- Sackstein, S. (2016). The perils of late work and how to make it count. *Education Week Teacher*. January 19, 2016. Retrieved from http://blogs.edweek.org/teachers/work_in_progress/2016/01/the_perils_of_late_work_and_ho.html.
- Schwartz, K. (2014). More progressive ways to measure deeper levels of learning. *Mind/Shift*. April 9, 2014. Retrieved from <https://www.kqed.org/mindshift/34821/more-progressive-ways-to-measure-deeper-level-of-learning>.
- Schwartz, K. (2017c). How to build self-assessment into jampacked high school classes. *Mind/Shift*. November 7, 2017. Retrieved from <https://www.kqed.org/mindshift/49222/how-to-build-self-assessment-into-jampacked-high-school-classes>.
- Sornson, B. (2017a). Time for disruptive innovation in our schools. *Community Works Institute*. January 19, 2017. Retrieved from <https://medium.com/communityworksjournal/time-for-disruptive-innovation-in-our-schools-a3e5bb170a25>.
- Sornson, B. (2017b). Hasn't Worked, Can't Work, Won't Work. *Community Works Institute*. February 1, 2017. Retrieved from <http://blog.communityworksinstitute.org/2017/02/01/hasnt-worked-cant-work-wont-work/>.
- Stiggins, R. (2008). *Assessment FOR learning, the Achievement Gap, and Truly Effective Schools*. Portland: ETS Assessment Training Institute. September 8, 2008. Retrieved from https://www.ets.org/Media/Conferences_and_Events/pdf/stiggins.pdf.
- Wall, K. (2011). *Special Needs and Early Years: A Practitioner's Guide*. London: Sage. Retrieved from <http://dx.doi.org/10.4135/9781473957862>.
- Warne, R.T. (2017). *Research on the Academic Benefits of the Advanced Placement Program: Taking Stock and Looking Forward*. Sage. Retrieved from <http://journals.sagepub.com/doi/full/10.1177/2158244016682996>.
- Wormeli, R. (2012). It's time to stop averaging grades. *Association for Middle Level Education*. Retrieved from <https://www.amle.org/BrowsebyTopic/WhatsNew/WNDet/TabId/270/ArtMID/888/ArticleID /21/Its-Time-to-Stop-Averaging-Grades.aspx>
- Yale University (<https://admissions.yale.edu/counselors>).

"If we teach today as we taught yesterday, we rob our children of tomorrow." ~ John Dewey