



## Redefining “Fair”

**Assessment and Grading  
for the 21st Century**

*Damian Cooper,  
Ken O'Connor,  
with Nanci Wakeman*

*The authors identify four challenges to improving assessment and grading practices for the 21st century. The good news is that none of these challenges are insurmountable as the writers present recommendations for improving learning for all students.*

In the 1980s, as they faced the realities of the information age, educational leaders began to question the traditional sifting and sorting function of schools. Literacy, numeracy, and problem-solving had clearly become the new 3 R's; but societal and economic forces meant that these skills were now essential for every student, not just for those heading to colleges and universities. The “outcome-based education” movement recognized that the focus of schooling needed to shift from teacher “inputs” to student “outputs”. In the ensuing decades, although the terminology changed, the focus on results remained. In 2008, the missions, visions, and values of school districts across Canada and the United States echo a common theme: *student success, success for all, failure is not an option*, and, of course, *No Child Left Behind*. The message to teachers is clear: they must work in ways that opti-

mize learning for ALL students.

But after more than two decades, and millions of dollars spent on developing and communicating new assessment and grading policies, as well as endless conferences and workshops that focus on results and student success, many teachers' beliefs and practices have changed very little. More often than not, we see practices, in middle and high school classrooms especially, that are intended to sift and sort the “A” students from the “D” students, or in Ontario, “My Level 4's from my Level 1's”.

Why has so little changed?

Having worked in schools and school districts around the world for several decades, we can identify four main challenges to change in Canada and the United States. The good news is that none are insurmountable and pockets of success are appearing, suggesting that significant change is possible.

### Challenge 1

As evidenced by their reluctance to abandon traditional grading and reporting formats, state and provincial departments of education have not demonstrated the political will to change grading and reporting systems from a norm-referenced model to a criterion-referenced model.

### Challenge 2

Despite the resources that have gone into supporting teachers, there is little accountability to ensure that “no teacher is left behind” in terms of acquiring and implementing the skills necessary to improve learning for ALL students.

### Challenge 3

Teachers cling to outdated beliefs about motivation, discipline, responsibility, and fairness that lead

to inappropriate assessment and grading practices.

### Challenge 4

The lack of communication between secondary and tertiary education systems supports teachers in holding on to these outdated beliefs.

### Challenge 1 and how to fix it

How do grading and reporting procedures need to change to align with an educational system whose mission is to ensure that all students learn and achieve at a high standard? There are five major changes that must occur:

1. Percentage grades must be replaced by levels of achievement or by letter grades representing levels of achievement.
2. The practice of averaging marks over the course of a term must be replaced by summary grades that reflect more recent achievement.
3. Single subject grades must be replaced by grades for each of the essential learning goals or standards. Ideally, there would be no overall subject grades, but we recognize there will be pressure to retain them in grades 11 and 12.
4. Data about behaviours and attitudes must be gathered and reported separately from achievement data.
5. Procedures for dealing with missed critical assessments must result in the completion of such work, NOT the use of zeroes.

### 1. Replacing Percentage Grades with Levels of Achievement or Letter Grades

Percentage grades are an artifact of norm-referenced grading. They are directly related to the normal curve and the statistical procedures associated with it. Percentage grades are not linked to predetermined curriculum targets. If they were, then a student who received 78% on an essay would be able to ask legitimately, “What is the 22% that I didn't get correct?” No teacher we have ever met would be able to answer this question!

In a criterion-referenced system, the symbols used to identify proficiency must be clearly stated and known to students before they begin work. Simply put, students must know what they have to do to “clear the bar”.

Educational systems that employ percentage grades do not clearly specify either the content to be learned or the performance standards to be met by students. They simply report, for example, “Science – 75%”. Such grades typically represent a hodge-podge of science content knowledge, skills, attitudes, and behaviours. Furthermore, the components of this grade may vary in significant ways from school to school, or even from class to class in the same school. In such systems, a grade average of 85% across all subjects may be quite arbitrarily identified as the “standard” for admission to a post-secondary science program.

Globally, both the International Baccalaureate (IB) and Advanced Placement (AP) programs utilize criterion-referenced performance standards that use achievement levels, not percentage grades, to certify proficiency: IB employs a seven-point scale ranging from “very poor” to “excellent” and AP employs a five-point scale, ranging from “no recommendation” to “extremely well-qualified”.

Clearly, if school systems are to be successful in educating all students to high standards, then percentage grades comprising an unspecified mix of learning variables must be replaced by levels or letter grades that represent clear, appropriate, and consistent measures of achievement.

## **2. Replacing the Mean with More Appropriate Summary Grades**

The practice of averaging a set of

marks gathered over a term is common in norm-referenced systems; however, this practice is inappropriate in a criterion-referenced system. Imagine if, after taking your driver's test for the third time and demonstrating mastery on all of the required skills, the examiner informed you that after averaging in your two earlier attempts, you would NOT be awarded your license!

In criterion-referenced grading systems, the final grade is based on more recent performance. This means using only summative assessment tasks for grading purposes. In a well-designed program, there will be a sufficient number of summative tasks to adequately sample essential learning. For example, each unit may end with a summative task and the final grade will be determined on the basis of the trend in student achievement on this set of tasks.

## **3. Replacing Single Subject Summary Grades with Grades for each Learning Goal**

Single subject summary grades have little value since many students achieve well on some learning goals and less well on others. The valuable information is provided by grades for each essential learning goal or each category of learning goal as is provided in Ontario on the provincial report cards for grades 1 to 8 for English, Mathematics, and Second Language. There is no over-all pass/fail or credit.

## **4. Separating Data about Achievement from Data about Attitudes and Behaviours**

While there is often a strong correlation between attitudes/behaviours and achievement, these are two distinct domains of learning. Achievement data must communicate to students and their parents

what concepts and skills have been learned and how well they have been learned. Such learning is to be found in the lists of subject-specific *learning goals, standards, or curriculum expectations* appearing in provincial and state guidelines. Attitudinal and behavioural data concern the generic habits and skills that are associated with school, and later, workplace success. These include *participation, effort, co-operation, homework completion*, etc. Teachers need to gather data about achievement and about attitudes and behaviours. But the data from each domain must be recorded and reported on separately. Teachers, students, parents, post-secondary institutions and potential employers all need to be able to review both kinds of information in order to make decisions about areas for improvement, promotion to the next grade, entry into college or university, or possible employment.

## **5. Responding to Late and Missed Work**

A criterion-referenced system identifies the critical evidence required to demonstrate that a student has learned the essential concepts and skills for a given subject and grade. This means that *all essential* evidence must be submitted before the teacher can make this determination. Criterion-referenced assessment planning requires that teachers agree upon the essential evidence needed to certify students as competent and knowledgeable in a given subject and then requires that students complete all of these essential tasks.

By contrast, some norm-referenced systems typically set the cut point for a passing grade as low as 50%. This method encourages mediocrity and also sends the message to students and parents that 50% of the required evidence is not important. If we are to change the culture of learning and assessment from one



that encourages mediocrity to one that demands high quality work from all students, then the message has to be communicated to students that all essential work must be completed to a high standard.

So what is an appropriate response when a student does *not* complete an essential piece of work? In the words of Douglas Reeves (2004), “The penalty for not doing the work must be that you have to do the work.”

School policies that permit the use of penalties for late work and zeroes for missing work and other behavioural infractions, such as academic dishonesty, stem from a belief in the power of threat and punishment. But as Guskey and Bailey (2001) point out, “...no studies support the use of low grades or marks as punishments. Instead of prompting greater effort, low grades more often cause students to withdraw from learning.” And while many teachers will argue that the threat of late penalties and zeroes motivates students to complete all work and to do so on time, a quick survey of work completion in any middle or high school will indicate the fallacy of this argument.

So how should teachers determine the report card grade for a student who has not completed one or more of the summative assessment tasks? If the use of zero is not an appropriate response to missed tasks, does this mean that summary grades should be determined by considering only those tasks that have been completed? Absolutely not! If all summative tasks are required to provide evidence of essential learning, then if one or more of these tasks are missing, the report card must indicate “Incomplete”. The message communicated to both student and parent is that the missing evidence of learning must be submitted before a summary grade can be determined.

### Challenge 2 and how to fix it

If teachers are expected to change their assessment and grading practices to improve student learning and improve communication about learning to students and parents, “pressure and support” (Fullan & Hargreaves, 1992) are equally important. High quality training must be available to enable teachers to implement the required changes and time must be made available for this training to occur. Such training must go beyond one-day workshops occurring off-site (Guskey, 2000). It must be followed by coaching and mentoring in individual classrooms, with the mentors coming from the ranks of teachers' peers, not outside experts. Such training must occur over several years – three at a minimum. And finally, teachers must be held accountable by principals for implementing the required changes. There can be no “opting out”.

Current approaches to voluntary professional development see already highly skilled teachers becoming ever more skilled at their craft as they attend a growing array of summer institutes and after-school workshops. The result: an ever-widening gap between highly skilled and less skilled teachers. This situation must change, but it can only change if the training in critical areas of practice such as assessment are made mandatory and are integrated into teachers' yearly routines.

### Challenge 3 and how to fix it

When asked to implement practices such as test and assignment re-writes, many teachers become indignant. Some argue, “That's not fair on the students who got it the first time.” Others complain, “How are we preparing students for the real world if they get to do tests and assignments over again? If you're a pilot landing an Airbus and you

blow the landing, you don't get a second chance. There are dire consequences.” Yet others charge, “If there's no longer any failure, what does that say about our standards? We no longer have any.” And still others argue, “We will no longer be able to motivate students to complete their work if the threats of late penalties and zeroes are removed.” These often heard objections to assessment reform reflect deeply ingrained norm-referenced attitudes about teaching and learning. Let's examine each of the arguments in turn.

Firstly, the “fairness” issue. In a **norm-referenced** system, fairness demands *uniformity*. Since the purpose of the system is to sift and sort students in order to have 'fair' competition and to be able to make valid comparisons, all students have to be dealt with in the same way. In practice, they all need to complete the same assessments under the same conditions.

By contrast, the purpose of a **criterion-referenced** system is to ensure that all students achieve proficiency. And because students are different, fairness requires not uniformity in testing but *equity of opportunity*. This means that the number and design of the assessments for a given subject or course may need to vary to enable all students to achieve success. Such a system still demands that the 'what' and 'how well' of the learning (content and performance standards) do not change; however, the context and the nature of the assessments may differ for some students.

Secondly, what of the charge that *we are not preparing young people for the real world if students are permitted to redo assessment tasks?* The Airbus pilot argument simply doesn't hold up. The pilot DOES get to “redo” the poor landing – not once, not twice, but hundreds of times – in the flight simulator at pilot training

*school*. Note the word, “school”. Our classrooms, including the senior years, represent the simulator, NOT the Airbus. Our assessment and grading practices must *prepare* students for entry into the outside world; they should not *be* the outside world. Besides, most “real-world” situations *do* provide adults with more than one opportunity to demonstrate competency.

Thirdly, the argument that a “*no failure*” policy represents a lowering of standards reflects a serious misunderstanding of the practices associated with criterion-referenced grading. Such a system does NOT mean that sub-standard work will receive a passing grade; it means that sub-standard work must be improved until it meets the standard. Rather than lowering our standards, current assessment reforms, when implemented as intended, lead to *improved* standards of student work.

In a criterion-referenced grading system, mediocrity is not acceptable. Implemented properly, such a system represents a significant raising of the bar. Quality, rather than being an option, is non-negotiable – and what is needed to certify proficiency is accurate evidence of each student's level of achievement, based on clear performance standards. The message to students is “Excellent, proficient, or not there yet – do it again!”

If we succeed in moving beyond these counter arguments, there are three changes that must occur at the classroom level before a criterion-referenced grading system can be implemented effectively:

- a manageable curriculum that focuses on essential learning;
- a deep understanding by teachers of the different purposes for assessment as reflected in detailed assessment plans;

- teacher willingness to differentiate instruction, as well as the skills to differentiate to meet the needs of all students.

### **1. A Curriculum that Focuses on Essential Learning**

The greatest challenge teachers face when they demand that all students demonstrate proficiency on essential learning is TIME. First, let's accept that there will never be sufficient time for teachers to *cover* everything, let alone time for students to *learn* everything. Furthermore, if teachers are going to *differentiate* their instruction to meet the strengths and needs of groups of students, as well as provide re-assessment opportunities for some students, lack of time becomes an even greater challenge.

The solution? Backward design (Wiggins & McTighe, 1998). Teachers, working collaboratively in grade or course teams, must engage in long-range planning to decide what is essential learning. With curriculum documents and classroom resources in front of them, they must reach agreement about the essential concepts and skills that students must acquire and demonstrate by the end of a term, semester or year. By “backward designing” units of study, teams of teachers ask the question, “What concepts and skills are critical for students to acquire and retain long after instruction has ended?” Once this question has been answered, teachers must next ask, “What critical assessment evidence must be gathered to demonstrate that students have learned these concepts and skills?”

### **2. Understanding the Different Purposes for Assessment**

The answer to the assessment

question in the preceding paragraph is the first step in developing an assessment plan for the unit or term. Three types of assessment must appear in any assessment plan:

1. Diagnostic assessment – the various assessments that occur early in a unit to determine students' understanding and skill levels *before* instruction begins.
2. Assessment OF learning – the summative tasks that students must complete to demonstrate that they have learned the essential concepts and skills.
3. Assessment FOR learning – the ongoing or formative assessments that help build students' understanding and help them to refine their skills leading up to the assessments OF learning.

The assessment OF learning tasks must be communicated to students and parents at the beginning of each term so that everyone knows precisely what evidence of learning is required. These tasks must be meaningful, relevant and, where possible, authentic. Report card grades will be determined on the basis of students' achievement on these tasks. The diagnostic assessments and assessments FOR learning are for instructional purposes. Data from these are used throughout the teaching/learning process to adjust instruction in ways that maximize learning for all students in a class.

### **3. The Need to Differentiate Instruction**

Referring to the assessment plan, the diagnostic assessments occurring at the beginning of a unit or term enable the teacher to group students according to their initial levels of skill and understanding. These assessments provide teachers with the information necessary to



differentiate instruction to ensure that each student learns within his or her “zone of proximal development” (Vygotsky, 1978).

In a criterion-referenced classroom, one size of instruction does NOT fit all. While the goal of a specific lesson may be the same for all students, teachers know they have to adapt the conditions of instruction – pacing, complexity of resources, number of practice questions assigned, level of support, etc. – for different groups within the class.

#### **Challenge 4 and how to fix it**

Communication between secondary and post-secondary institutions must improve significantly, to improve alignment of their grading procedures and to ensure that information about graduating students is valid and reliable. Currently, some secondary teachers justify many of their assessment and grading practices in terms of preparing students for the rigors of college and university. In our work, we frequently hear from teachers who agree with all of the approaches suggested in this article, but then do not implement them because “my students won't be prepared for the reality of post-secondary education.”

This argument is problematic for several reasons. Firstly, many post secondary institutions are now beginning to embrace assessment for learning practices, as well as beginning to rate student achievement against known standards. Secondly, post-secondary education is fundamentally different from K to 12 public education. Finally, *preparing* students for post-secondary destinations will not be accomplished through “sink or swim” approaches that set students up for failure rather than for success.

That said, communication between

secondary and tertiary educators needs to be significantly improved. Such communication is necessary to improve the alignment between the grading and reporting procedures used at each level. It is necessary to establish, to the satisfaction of both groups of educators, precisely what kinds of assessment tasks are valid in order to determine students' preparedness for the demands of college and university education. It is necessary to reach agreement on the standards of performance required for each of these assessment tasks, based on the performance levels that will be required for success in college and university. In short, discussions need to focus on replacing a student's overall average as a measure of success with a set of clear *criteria* that will serve to better predict student success in tertiary education.

#### **Conclusion: Changing the Culture of Assessment, Grading and Reporting**

We have identified four challenges that currently stand in the way of improving assessment and grading practices in our schools. We have also suggested ways to remove these challenges in order to improve learning for all students. However, our suggestions demand that educational leaders make a number of difficult decisions which will then require many teachers to make significant changes to their current practice. But as Dylan Wiliam (2007) has recently asked, “The question is whether we have the courage to reach.”

Sections of this article first appeared in the *Educators Resource Guide: Assessment, Evaluation and Reporting of Student Achievement* produced by The Ottawa Carleton District School Board, 2008.

Guskey, T.R. & Bailey, J. (2001). *Developing grading and reporting systems for student learning*. Thousand Oaks, CA: Corwin.

Guskey, T.R. (2000). *Evaluating professional development*. Thousand Oaks, CA: Corwin.

Fullan, M. & Hargreaves, A. (1992). *Teacher development and educational change*. Toronto: Routledge.

Reeves, D. (2004, December). The case against the zero. *Phi Delta Kappan*, 86 (4), 324.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Wiggins, G. & McTighe, J. (1998). *Understanding by design*. Alexandria, VA: ASCD.

Wiliam, D. (2007). Content then process: Teacher learning communities in the service of formative assessment, in D. Reeves, ed., *Ahead of the curve*. Bloomington, IN: Solution Tree.

**Damian Cooper** is an independent education consultant who works across Canada, the United States and overseas, helping schools and school districts improve their instructional and assessment skills. He spent many years as a teacher, working in academic environments, vocational settings and schools serving students with severe learning disabilities. Damian has a wealth of experience at the classroom, district and state levels. His published works include *Getting Assessment Right: Mathematics and Getting Assessment Right: Language*. *Damian's latest resource*, *Talk About Assessment: Strategies and Tools to Improve Learning* is published by Nelson.

**Ken O'Connor** is an educational consultant. Ken's teaching background includes over 30 years of experience from grades 7 to 12 in six schools in Ontario and Victoria, Australia. Formerly Curriculum Coordinator with the Toronto District School Board in Ontario, Canada, his responsibilities included student assessment, evaluation and geography. Ken has authored extensively, including co-authoring the highly successful resource book, *Assess for Success*, published by the Ontario Secondary School Teachers' Federation in 1994. His other publications include *How to Grade for Learning: Linking Grades to Standards (2nd Edition)*, Corwin, 2002 and *A Repair Kit for Grading*, ETS Portland, 2007.

**Nanci Wakeman** is a freelance writer and editor. After a career as an educator and education writer, she attended Sheridan College's Print Journalism program. She has recently been published in *Markham Business Magazine*, *Healthy Living Magazine*, the *Markham Economist* and *Sun and Red Point Media's Medical Concierge Magazine*.