REVOLUTIONIZE ASSESSMENT

Empower Students, **Inspire** Learning

RICK STIGGINS

Foreword by Michael Fullan

© Hawker Brownlow Education

Contents

Foreword v
Michael Fullan

Acknowledgments vii

About the Author ix

1. **A Hidden Crisis in American Education** 1
2. **To Begin With, We Must Understand This Realm of Educational Assessment** 11
3. **Understanding Our Assessment Trap** 19
4. **These Are Troubled Times in the Realm of Educational Assessment** 29
5. **A New School Mission Demands a New Assessment Vision** 39
6. **Assessment for Truly Effective Schools Requires Local School District Leadership** 57
7. **A Concluding Message to the Measurement Community** 81

References 91

Index 93

© Hawker Brownlow Education

© 2014 Hawker Brownlow Education • 9781760012960 • CO2960

Foreword

Rick Stiggins has spent his professional life working with students, teachers, school leaders—and assessments—and this book recounts the evolution of his thinking about how best to use assessment to improve education. Stiggins understands the uses and misuses of assessments, but he sees there has been much more of the latter—especially with the increasing use of high-stakes annual standardized tests. He explains the problems and their solutions in such a comprehensive and compelling way, we may finally get some action on the phenomenon of assessment.

Drawing on detailed research evidence and his own close work with practitioners and researchers, Stiggins leaves no stone unturned. He shows how and why assessment, which started off as a great potential source of improvement for what ails schools, has become the single biggest albatross in the reform package. The first four chapters not only document the problem of current assessment, but also show how and why we have become mired in a vicious circle of wasted effort and resources in the billions of dollars. No one can read these chapters without increasing dismay.

Chapter 4 brings this bad pre-history into focus, by addressing head on the issue of why annual testing has not improved schools. The culprit is the naïve assumption of policy makers about the role of testing combined with a massive lack of assessment expertise at all levels. The bottom line is that there is almost no focus in the system on examining the impact of testing on day-to-day teaching or on learning and student success. Having revealed the "assessment emperor" as having no clothes, Stiggins spends the remaining three chapters mapping out a fundamentally different role for assessment.

© Hawker Brownlow Education

We are then treated to a wonderfully clear and inspiring chapter on a new assessment vision for schools, including a powerful, insightful section on "the emotional dynamics of being evaluated." With this vision as a foundation Stiggins turns to strategy, starting with the roles of school and district leaders. He formulates three "local priorities"—very clear, specific, and doable.

In the final chapter Stiggins offers specific suggestions about how to implement useful assessments that contribute to student learning, this time speaking specifically to the measurement community and testing industry. *Revolutionize Assessment* is both a deeply personal and professional book, and it weaves these two dimensions together beautifully. The message is clear, it is essential, and it is both a cry of frustration and a solution delivered on a platter. It is time to re-position assessment. Rick Stiggins has given us the ideas and tools to do so—a great legacy.

Michael Fullan
Professor Emeritus
OISE/University of Toronto

© Hawker Brownlow Education

4

These Are Troubled Times in the Realm of Educational Assessment

Failure is simply the opportunity to begin again, this time more intelligently.

Henry Ford

Until the middle of the last century, society had been satisfied evaluating its schools based on the quality of the instruction provided. The implicit assumption was that if good teaching was provided, the achievement outcomes would take care of themselves. However, that confidence was shaken in the 1950s and 1960s, by social upheaval at home and the space race with the Soviet Union, beginning with the USSR's launch of the first satellite, Sputnik. We began to ask serious questions about the quality of our schools, and this spawned a desire to evaluate them based on actual achievement results. Those in positions of civic leadership came to believe we could base that evaluation on evidence derived from annual standardized tests and, in doing so, we could

© Hawker Brownlow Education

improve the quality of our schools. What follows is a brief history of the implications of that naïve belief.

My purpose in reviewing this history is not to challenge annual accountability testing or advocate for its removal. This testing strategy, in fact, has had little impact on instruction for several reasons. We must understand that its lack of influence does not arise from the inappropriateness of the tests, but rather the insufficiency of the information they provide. We have relied on them too heavily when we should have been supplementing them with other applications of assessment that can provide the additional evidence needed to improve student learning.

There are those who would contend that we never expected our annual standardized tests to do the school improvement job by themselves. But, in fact, this is the only application of assessment in which we have invested. As the development of annual standardized tests outlined below will reveal, the investment has been immense. If we had believed that other applications of assessment could help improve schools—such as day-to-day classroom assessment—would we not have invested as heavily in their quality, development, and use? Clearly, we have not done that. Training programs for teachers and school leaders have remained devoid of any helpful classroom assessment training. We have not even invested in preparing practitioners to use the standardized tests we have banked on so heavily. Neither our evaluations of the quality of teaching and teacher performance nor our evaluations of school leaders have ever included an examination of the quality of their assessment practices. Until very recently, as a society we have invested in only one vision of excellence in assessment in our schools. Let's take a look at that history.

Our story begins with college admissions tests. From the 1950s we have been led to believe they predict academic success and so have made them the gatekeeper for access to our system of higher education. If fact, they provide only a very modest prediction of freshman grade point average and do not predict GPA beyond the first year. High school grades are as predictive as these test scores, yet as a society, we rarely question the lofty status of SAT scores.

In the 1950s and 1960s, in response to the growing demand for more effective schools, local school boards adopted districtwide standardized testing programs to compare the scores of their schools to national norms. The resulting test scores were reported in the local media in the hope that this would promote school

© Hawker Brownlow Education

improvement. It did not matter that local educators had little idea what was being covered on these published tests or how those expectations related to their local curriculum; they had absolutely no training in how to interpret technically complex scores, let alone how to use tests to improve schools.

In recounting this history of the "accountability movement," allow a cash register to ring up in your mind accumulating the long-term costs. Each new level of testing was added on to those that preceded it. Again, my intent is not to use this as an argument to stop the testing. Rather, it is to reveal why these testing practices have done so little to promote instructional improvement.

In the 1960s, we began ranking states based on average college admissions test scores and reporting the results in the news media, assuming that low ranking states would be sufficiently concerned that they would work harder to improve their schools. However, those mandating such rankings had no idea that few educators had any idea what was tested on college admissions tests or, indeed, whether what was tested could be influenced by improved instruction. Indeed, these were referred to as tests of scholastic aptitude and the conventional wisdom was that this student characteristic could not be improved with study. Regardless, the mere fact of the public reporting of those relative test score ranks was considered sufficient to drive school improvement.

In the 1970s, we added state-wide testing programs. We began that decade with very few state-wide testing programs and ended the decade with programs in the majority of states. Surely, those in leadership positions believed, if we compare local districts' performance with each other and report the results to the public, districts will work harder to be sure their schools improve.

In the 1970s and 1980s, we added a national assessment program. Education leaders adopted and began testing a common set of achievement expectations so we could compare states on student performance with the intent of helping low performing states. The ranking of states has remained quite stable.

In the 1990s we became enamored of international testing programs comparing nations, and the United States finished in the middle of the pack. Surely, political and school leaders announced, we can and should do better. We were (and are) embarrassed and that should begin to improve things, they hoped. Again, it did not matter that few at the local level had any idea what was tested or, therefore, knew how to improve instruction in

© Hawker Brownlow Education

© 2014 Hawker Brownlow Education • 9781760012960 • CO2960

the service of greater success. Our place in the rank order of countries remains essentially unchanged.

In the 2000s, political pressures gave rise to our first-time-ever, national, every-pupil testing in the service of "leaving no child behind," as control over the accountability movement and its drive to promote better schools completed its journey from local school boards to state legislatures and on to the federal government.

More recently, we see a movement to attach teachers' and administrators' salaries and employment status to changes revealed in annual standardized test scores. Once again, it appears that those setting such policies are unaware of the fact that most teachers teach at grade levels and in subjects where no standardized test is available and, even when they are, these annual accountability tests:

- have not been designed or validated for this purpose,
- often do not align with an individual teacher's assigned teaching responsibilities,
- are not sensitive enough to detect the impact of an individual teacher on these test scores, and
- yield scores that are influenced by many factors that are beyond the control of teachers.

In short, they cannot serve productively in this capacity.[1]

Once a layer of testing was added, it remained. All the layers remain in place today, with another half a billion dollars currently being spent to create and implement new Common Core Standards in national tests. Seven decades later, the time has arrived to reflect on and talk about the cost of this intervention—in terms of time and money—in relation to its contribution to school improvement. In some instances, students can spend as many as 30 or 40 instructional days or more of a 180-day school year taking tests.

Despite this enormous investment of time and money, scores on our national examinations have changed little, achievement gaps have not narrowed as fast as we would like, our place in international rankings remains fairly static, dropout rates remain unacceptably high, and graduation rates remain unacceptably low. Surely, there are many causes contributing to these outcomes, but as we reflect on causes, it is relevant to ask, *Has annual testing delivered on its promise of improved schools, especially at a level commensurate with its costs?*

When school improvement researchers develop new instructional ideas, they are expected to carry out rigorous scientific

© Hawker Brownlow Education

studies to demonstrate that implementation of their intervention will enhance student learning. They are expected to demonstrate the impact of their teaching idea by showing how much better students perform once the new approach is in place. In school improvement research circles, this has become the coin of the realm for defending new ideas. I can find no place where this often-demanded scientific research has even been proposed, let alone conducted for standardized testing. Think about what this says about our blind faith in the utility of these tests.

Some will contend that these tests are not supposed to be the *cause* of school improvement. Rather, their intended purpose is to measure the *effects* of school improvement interventions. But when we witness the immense costs associated with this practice, I must don my taxpayer hat. We should demand evidence that this has been a sound investment of our educational tax dollars. Without this kind of cost/benefit analysis, how do we judge value in relation to other school improvement alternatives? Rather than using the absence of this research to argue for ending the testing, we should use it to understand the limits of this form of testing and, thus, to keep its expected benefits in perspective. Our assessment resources should be allocated across assessment strategies or applications in proportion to the expected contributions to school improvement.

THE CAUSES OF OUR TROUBLED TIMES

This brief retrospective alludes to some of the reasons for our chronic inability to connect testing to instructional improvement. They arise from the professional environment that has surrounded assessment practices over the decades. Let me spell out those causes explicitly and in some detail. They define both the reasons for our decades of assessment discontent and the remedies needed to create a new and productive vision of excellence in assessment of our schools.

Lack of Assessment Expertise

In recent decades, annual test scores have often been delivered to local teachers and administrators who are unschooled in what is being tested, what the scores mean, how to interpret them, and how to link them to instructional decisions, but it does not stop there.

© Hawker Brownlow Education

© 2014 Hawker Brownlow Education • 9781760012960 • CO2960