Quality Assurance in Education

Bringing formative classroom assessment to schools and making it count

Edmund W. Gordon Michael V. McGill Deanna Iceman Sands Kelley M. Kalinich James W. Pellegrino Madhabi Chatterji

Bringing formative classroom assessment to schools and making it count

Edmund W. Gordon, Michael V. McGill, Deanna Iceman Sands, Kelley M. Kalinich, James W. Pellegrino and Madhabi Chatterji
(Author affiliations can be found at the end of the article)

Abstract

Purpose – The purpose of this article is to present alternative views on the theory and practice of *formative assessment (FA)*, or assessment to support teaching and learning in classrooms, with the purpose of highlighting its value in education and informing discussions on educational assessment policy.

Methodology/approach – The method used is a "moderated policy discussion". The six invited commentaries on the theme represent perspectives of leading scholars and measurement experts juxtaposed against voices of prominent school district leaders from two education systems in the USA. The discussion is moderated with introductory and concluding remarks from the guest editor and is excerpted from a recent blog published by *Education Week*. References and author biographies are presented at the end of the article.

Findings – While current assessment policies in the USA push for greater accountability in schools by increasing large scale testing of students, the authors underscore the importance of FA integrated with classroom teaching and learning. They define what formative classroom assessment means in theory and in practice, consider barriers to more widespread use of FA practices and address what educational policy makers could do to facilitate a FA "work culture" in schools.

Originality/value – The commentators, representing scholar and practitioner perspectives, examine the problem in a multi-faceted manner and offer research-based, practical and policy solutions to the observed issues in FA. Dialogue among stakeholders, as presented here, is a key first step in enacting assessment reforms in the directions discussed.

Keywords Formative assessment, Classroom assessment, Diagnostic assessment, Educational accountability, Student testing

Paper type Technical paper

Introduction

Madhabi Chatterji, Guest-Editor, Teachers College, Columbia University[2]

Formative assessment (FA), or more generally assessment conducted to support teaching and learning in classrooms and schools, is a notion that is hardly new to education. At different times in the past, we have heard the familiar calls for more FA or for teaching practices associated with FA, such as, giving student-focused feedback based on assessment results (Hattie and Timperley, 2007; Nitko, 1989; Popham, 2008; Shepard, 2000). The conversation was renewed recently and is continuing unabated today (see The Gordon

Quality Assurance in Education
Vol. 22 No. 4, 2014
pp. 339-352
Emerald Group Publishing Limited
0968-4883
DOI 10.1108/QAE-07-2014-0034

Commission, 2013). Despite the calls, however, practices in schools around the world and at *all* levels of education tend to lag behind the rhetoric (see Black *et al.*, 2003). Why?

One reason for this in the USA could be the greater push for more large-scale testing of primary and secondary school children under federally mandated accountability policies (see the No Child Left Behind Act of 2001 at: www2.ed.gov/nclb/landing.jhtml and the more recent, Common Core State Standards reforms at: www.corestandards.org). Other, less obvious, factors could also be in the way.

In unison, the authors of this discussion call for more formative classroom assessment. They help clarify what formative classroom assessment means in theory and in practice, point out why educators, policy makers and the public should value it more and give it a firm place in schools, identify barriers to more widespread formative practices in schools and most importantly, speak to changing assessment policies to facilitate a greater FA culture in schools.

This third moderated discussion, *Bringing Formative Classroom Assessment to Schools and Making it Count*, again features the thinking of leading scholars in education and educational measurement (Edmund W. Gordon, Deanna Iceman Sands and James W. Pellegrino, in order of presentation), counter-balanced against prominent leaders of two US education systems (Michael McGill and Kelley Kalinich, in presentation order). The discussion is excerpted from a recent blog published in *Education Week*, co-facilitated by James Harvey of the National Superintendents Roundtable and myself (see http://blogs.edweek.org/edweek/assessing_the_assessments). In conclusion, I provide my own perspective on FA based on some recent research and summarize authors' recommendations.

Assessment *for* learning: measurement science can do more for education!

Edmund W. Gordon, *The Gordon Commission on the Future of Assessment in Education*[1]

The Gordon Commission on the Future of Assessment in Education was created in January 2011 with a mission to study the best of what we do in assessment, anticipate where education will be and what education will need of assessment by the mid-twenty-first century and to make recommendations concerning how assessment in education can better meet those needs. The commission consists of 30 scholars and thought leaders who elected to privilege as a main purpose of assessment in education to be: *to inform and improve teaching and learning* (see www.gordoncommission.org).

In the recent past, some of us in the Gordon Commission observed with skeptical satisfaction the coalescing of conservative and liberal political forces around the provision in the No Child Left Behind (NCLB, 2002) education legislation requiring more standardized testing in our schools (NCLB, 2002). I was especially attracted to the NCLB notion that all students should be tested annually and that the data from these tests disaggregated to reveal and address disparities in academic achievement between children from different social divisions. I welcomed this change in the interest of social justice.

What I paid insufficient attention to was the possibility that while the liberal forces may have been interested in the reform of education to address achievement gaps, conservative forces were more interested in using the same data for an increased

emphasis on accountability to demonstrate the limitations of public schooling. The main interest of conservatives appeared to be to privatize public education, which has been historically supported by taxpayer funds in the USA, through potential corporate takeovers of schools showing lower test scores.

For many, the documentation of the failures of the public school using "objective" standardized test results have provided the scientific evidence particular political groups needed to reduce support for public education. If there is no basis in reality for my speculation, how very strange it is that the misuse of standardized test results in such a punitive approach to educational accountability (see Chatterji, 2013a, 2013b) should be so aggressively pursued as national education policy! Do wise policy makers not see the downside of this strategy? Could this policy be wrong? Is this national embrace of educational testing to measure, reward and penalize schools and public educators missing a genuine opportunity to use one of the strongest of the behavioral sciences, the measurement sciences in a more powerful way?

We seem more interested in using the measurement sciences for the *assessment OF education*, when we should be using it also to *assess FOR education*. The cutting edges of today's science increasingly demonstrate that assessment, measurement and appraisal can be used to inform and improve the processes and the outcomes of education. How much more enlightened would be *ASSESSMENT IN THE SERVICE OF EDUCATION* as an education policy of a democratic nation committed to developing intellective competence in its people? I contend that we in the measurement sciences may be much too passive in our use of available knowledge and tools in our profession to serve education well. In public education contexts in the USA today, we seem to use test data mostly to select, place and hold educational entities accountable.

Modern measurement science is capable of more! We have emerging models that can be used to measure the status of developed ability in support of accountability *and* to enable teachers and learners assume responsibility for the development of students' intellective capacity. For example, we have a long tradition of formative approaches to assessment, highlighted in the reports of the Gordon Commission (www.gordoncommission.org/index.html). Assessment is best structured as a coordinated system focused on the collection of evidence […] that can be used to inform and improve the processes and outcomes of teaching and learning (The Gordon Commission, 2013). In my view, the Commission's focus on *processes* as well as *products* of learning offers a promising difference from NCLB-driven testing.

Calfee and colleagues (in press) and Haertel (2013), writing independently in their contributions to the commission, present a more detailed picture of the possibilities of FA to foster learning. In Haertel's (2013) vision of the future of schooling, classroom assessment will be truly integrated with instruction based on student pursuits that are educationally useful and intrinsically meaningful in the classroom context. Assessment inferences will be informed by observations of the processes as well as the products of student learning activities. Records of students' actions will be captured and analyzed to support high-level inferences about their reasoning and expertise.

A similar view is also given by McManus in a Council of Chief State School Officers' report (CCSSO, 2008), who states: "Formative assessment is a process used by teachers and students during instruction that provides feedback to adjust ongoing teaching and learning to improve student achievement of intended outcomes" (CCSSO, 2008, p. 5). Other characteristics that deserve attention are:

- FA is a process rather than a particular kind of assessment […]. There is no such thing as a "formative test".
- The FA process involves both teachers and students […], both of whom must be actively involved in the process of improving learning.
- *Learning progressions* or maps that depict progressively challenging learning goals, provide teachers with the big picture of what students need to learn, along with sufficient detail for planning instruction to meet short-term goals.
- Teachers must provide students with the criteria by which learning will be assessed […] using readily understood language and realistic examples of what meets and does not meet the criteria.

The above considerations (McManus, 2006; CCSSO, 2008) significantly advance and reshape previous conceptions of FA in schools. There are other newer assessment approaches that could also be used to support learning in the classroom. Among these are the qualitative analysis of teaching and learning transactions; adaptive testing; portfolio development and analysis and assessment as instruction (for more details, see our reports at: www.gordoncommission.org/index.html).

Clearly, measurement science is capable of serving teaching and learning better than we have in the past. We have watched as the nation has veered toward upholding less than productive purposes of assessment in education. The privileging of a punishment and reward approach to accountability testing in schools is proving to be disastrous as national education policy (see www.GordonCommission.org). We can do better!

K-12 assessments are more art than science

Michael V. McGill, *Bank Street College of Education, previously Scarsdale School District, NY*[1]

Gordon describes the "liberal" and "conservative" forces that have shaped federal and/or state education policies (Gordon, this issue; The Gordon Commission, 2013). It may be useful to clarify the fact that these forces do not align neatly with the outlook of either major political party in the USA. For generations, most people saw education as an interest that transcended partisanship; in recent years, that vision has become a dark caricature of itself. The misguided high-stakes testing movement would not have its current reach were it not for the support of a broad political coalition of Republicans and Democrats, a consensus that reaches back to the Clinton administration in the 1990s.

Left-wing collusion in enforcing a high-stakes testing policy of shaming and blaming teachers and schools has been at least as responsible for the damage done to schools as the Right-wing interest in privatizing education to cut costs. This unlikely alliance reflects largely the influence of a powerful business community that continues to perpetuate two highly questionable premises: that education's main purpose is to serve the nation's economic machinery and that business methods are the schools' salvation. Fueled by self-interest (some of the most ardent proponents of "corporate-style education" stand to profit mightily from it) and supported by a largely uncritical media, the strategy grinds implacably onward (Ravitch, 2013).

Policy makers do not see alternatives for many reasons. The dominant reform narrative is simple and easy to understand; alternative perspectives are complex, nuanced and difficult to grasp. For example, "Hold teachers accountable for test results"

is a seductive sound bite. It is a lot harder to explain why meaningful education gains depend on far more complex changes in the quality of the teacher applicant pool, teacher education, professional development and school culture (see also, Ravitch, 2011).

There is a general lack of understanding of the issues and mistrust at every level: federal to state, state to local districts and schools and local to teacher levels. Research has little impact on policy because there is little two-way communication or collaboration among the different parts and levels of a fragmented "system" of governments, policy makers, researchers, politicians and practitioners (Ravitch, 2011, 2013).

Further, public funding for enlightened school reform has diminished at every level, and private money has filled the resulting vacuum. Much of it comes from foundations that support some corporate agenda. Lack of funding and staff reductions have weakened professional organizations and other groups that have traditionally influenced education policy. Businesses and foundations have significant resources and have now grown even more powerful (Ravitch, 2011, 2013).

Also, as Gordon (this issue) aptly notes, those with a responsibility to speak out have been largely silent. He specifically mentions those involved with the "measurement sciences". In a better world, school superintendents, principals and teachers also would have risen up. We have been reluctant to take a stand, however. Fear of criticism, fear of authority, fear for our reputations or for our employment have all played a part. University professors outside measurement, by and large, have been quiet as well. They did not have a dog in this fight, it seemed; now the same ineluctable specter of metrics and "analytics" that has haunted elementary and secondary education over these last years is dogging them.

The past does not have to be prologue. Public backing for current practices is remarkably shallow. Supporting high standards is not the same as supporting today's government policies. For example, recent public poll respondents think that test results should be used to help children learn and not for high-stakes decisions. Still, those who promote the current accountability policies are politically strong (see www.ecs.org/clearinghouse/16/21/1621.htm).

We need *less*, not more, mandated testing, and we should be extremely cautious about externally imposed, embedded assessments that create their own curriculums de facto. In the process of trying to lift the average quality of teaching and learning by standardizing both, we run the serious risk of crushing the individuality, the initiative and the pursuit of student and teacher interests that make education a memorable and valuable process for students.

The challenge of building a program of assessment *for* teaching and learning is to blend accountability to external standards with local curriculum and measures that can reflect emerging events and knowledge taught as well as student and teacher interests. Practitioners must be able to integrate assessment into their instruction seamlessly or with minimal intrusion. The particulars of this work are often situational, so I will not address them in detail here.

I will mention two practical problems. First, it takes time to develop and implement curriculum and assessment. And principals and teachers do not have a lot of free time today, especially, given the extensive claims of mandated, summative testing. Second, surprisingly few practitioners have significant experience in creating or using high-quality, embedded FAs. To realize the promise of "assessment for learning",

therefore, we need to invest significantly in professional development and give practitioners time to create these tools and use them effectively. If the Federal and State Governments in the USA want to make a meaningful contribution that is where they could deploy their resources.

In sum, good assessment is valuable, and it should be a more integral part of what teachers do each day. However, as all assessments are imperfect, we would do well to recall that the "science" of measurement adds value to education only when we recognize its limitations and inform it with sound human judgment. Likewise, we would do well to remember that we are engaged in a team enterprise where success depends on collaboration.

Expanding what FA means

Deanna Iceman Sands, *University of Seattle*[1]

In previous contributions to this Special Issue, the concept of FA was raised on several occasions (see Kingston, this issue; Gordon, this issue). Kingston (this issue), for example, explicitly described the concept of FA as "tests" or products and assessment systems. Other authors suggested further that teachers need support and professional development to design assessments that are connected directly to their curriculum as opposed to content typically captured in district, proprietary benchmark, state or national, standardized assessments (see McGill, this issue).

Emerging new definitions of FA move the concept along. One well-known definition was proposed by the Council of Chief State School Officers (CCSSO, 2008). The Council defined FA as "a process used by teachers and students during instruction that provides feedback to adjust ongoing teaching and learning to improve students' achievement of intended instructional outcomes" (CCSSO, 2008, p. 3). Notably, this definition describes FA as a *process* rather than an assessment instrument, tool or product.

A FA process framework. My colleagues and I have developed a framework to explicate the process of FA through the Laboratory of Educational Assessment, Research and InnovatioN. LEARN (see www.ucdenver.edu/academics/colleges/SchoolOfEducation/CentersPartnerships/LEARN/Pages/LEARN-Home.aspx). The project is based at the University of Colorado, Denver, and involves researchers and doctoral students from several institutions, such as Seattle University, the University of Colorado Boulder, and the University of Washington. The FA framework rests on an extensive base of theoretical, conceptual and research literature. The framework makes operational a concept of FA that involves "four assessment activities - clarifying learning expectations (goals), collecting information, interpreting information, and acting on/using information collected to move students closer to the learning goals".

All *four activities* are necessary for an occurrence of FA to transpire. Further, our structure considers FA as an ongoing, continuous process that occurs along a continuum of formal and informal practices. It involves planned and more spontaneous actions, one or more students and occurs in a social context that involves norms, routines, tools and roles/responsibilities for the teacher and students.

This FA framework and accompanying assessment activities build on socio-cultural and cognitive psychology theories of learning. It is situated in the daily life of classrooms and allows teachers to seamlessly engage in assessment connected to their planned and implemented curriculum and instruction. The framework and assessment

activities provide a venue for enacting evidence-based teacher practices, such as explicating clear learning goals and targets, questioning, observing, providing feedback, modeling and planning instruction based on individual student needs, develop and use instructional tools (e.g. progress monitoring sheets, advanced organizers) to scaffold and support student learning and establish norms and routines to create a culture of advancing assessment itself.

FA conceptualized in this way promotes self-directed learning for students, as they engage in goal setting, self-assessment, self-monitoring and self-regulation of their learning strategies. This framework encourages building classroom contexts for collaboration between teachers and students (and among students) by applying practices, such as peer-assessment, peer-mentoring and coaching. Through our continued work, LEARN colleagues hope to use this framework to provide protocols that can be used to identify when, and the quality by which, FA occurs in classrooms.

Classrooms are dynamic systems in which teacher and student decision making vary by content area as well as by day-to-day tasks and activities. Teachers must be equipped to understand how to set clear learning targets and outcomes and collect, interpret and act on information, in concert with their students, to plan, facilitate and support learning tied to standards and curriculum goals and daily objectives. FA, conceived broadly, yet framed explicitly, provides a structure for a multitude of strategies (including the use of data from "tests") to inform and advance student learning, critical thinking and achievement. As important, FA provides teachers with a process and set of practices that more decisively and completely situates them to carry out the critical role they play in their students' intellectual, communicative, socio-emotional and overall well-being.

FA in the real world

Kelley M. Kalinich, *Kenilworth School District, Illinois*[1]

In many ways, the world we all live in has become one big FA. Consider a purchase of an item via the Internet and you are asked to complete a consumer survey. When you return to the Web site, a list of "potential items" for you to purchase appears, based on your prior purchases. This same dynamic is in play in the classroom when teachers seek meaningful information about their students' learning to inform differentiated paths of instruction. As Deanna Sands and others observe, FA is a *process* (see Gordon; Sands, this issue). This multistep process to gather information about student growth, progress and engagement is critical to the art and science of teaching and learning.

In my 33 years as an educator, I identify the topic of "assessment" as one of the major areas where I have observed teacher growth as well as overall growth in the education field. Early in my career, assessment included standardized tests, end-of-unit tests and teacher observations. Goals for students were based on these three assessments, which were more summative in design. Typically, very little was done with the information when a unit of instruction was completed. On multiple occasions, an innovative application of "mastery learning" (see Guskey, 2010) attempted to use FA, but fell short in regards to higher levels of learning on Bloom's Taxonomy, such as, analysis, synthesis and evaluation.

That is not what classrooms look like today. FA is professed to be the foundation for high-quality teaching. It should be ongoing and fluid, building from data set to data set

to guide a teacher's instruction for groups of students as well as individuals. In my observations, many classrooms are showing evidence of these practices.

An example of classroom practices. I will offer a glimpse into what this looks like in practice from a third-grade classroom I recently observed in my district. The teacher started a reading lesson by letting students know that they would be detectives gathering evidence to identify the more "needy" character, Charlotte or Wilbur, in the book *Charlotte's Web*. This seemed simple enough, but how the teacher managed her FA of the students' grasp of the concept, as well as ownership for their reading, was quite demanding.

Students gathered their evidence and wrote it down on their iPads through a shared Google document. They discussed their ideas and debated their positions, while also posing clarifying questions to one another. The teacher made note of the key points of conversations she observed and recorded these in a Venn diagram. When it came time for students to articulate their conclusions in the large group, the teacher surveyed them using a show of thumbs held close to their bodies, to let the students respond in a more private manner. She recorded their responses, and then led them through deeper questioning and application of Charlotte's and Wilbur's characteristics to a real-life situation they had read about earlier in the week. The lesson ended with students performing a "quick write" in their journals about their concluding opinion and evidence to support the opinion. When the students left the classroom for lunch, they had to present their journals and answer the question, "What evidence was the most important to support your conclusion?"

With this 45-minute lesson, the teacher had a wealth of information from which to understand her students' learning and to develop and design the next reading lesson. This is the dynamic type of classroom that Sands (this issue) advocates, and it is the classroom that all of our students deserve. Marzano (2010) notes that the more accurate information teachers have about a student's progress or the performance of the total class, the better the teachers' judgments will be regarding meaningful instruction. Assessment should be ongoing and fluid, but most importantly, it should connect to the work our teachers do in the classroom every day to result in quality learning for our students.

Learning from reform mistakes of the past

James W. Pellegrino, *University of Illinois, Chicago*[1]

Those who ignore mistakes of past education reforms are destined to repeat them. We can equally well anticipate that those who ignore the present implementation challenges of the Common Core and Race to the Top (RTTP) assessment programs in the USA (see Common Core at www.corestandards.org; see RTTP at www2.ed.gov/programs/racetothetop/index.html) are destined to repeat these mistakes when it comes to implementing the assessment of the Next Generation Science Standards (NGSS; see www.nextgenscience.org). One lesson of the current experience is that it has been too fast, especially in a climate of high-stakes testing and accountability. Such are the arguments and evidence emerging from various quarters, including New York State.

Let us be clear: the problem is not that new and better common standards are not sorely needed; nor is the problem that with new and more demanding standards, the assessments must change accordingly, regardless of their proposed uses. The problem

is that large-scale change cannot come all at once, and it cannot be implemented solely from the top down. Educators need time and resources to make changes in their practice, and assessment developers need time to properly design and validate the assessment materials and tools needed by teachers and others to evaluate student learning effectively and fairly.

Need to change how we implement reforms. Previous contributions, such as Gordon's (this issue), Sands's (this issue) and McGill's (this issue), have pointed out that assessment should be designed to support teaching and learning and not undermine it. This is not to say that what the two assessment reform consortia have been developing for these past three years and field-tested this spring will not be of high quality relative to the standards they purport to assess. The jury is still out on that. We will have to wait and see what the four-year RTTP assessment program has wrought and how it will impact students' learning and teachers' lives.

The point is that we started in the wrong place. Our policy mandates at the federal level dictate that we build large-scale assessments for purposes of accountability and then ask school systems to make the transition while struggling under the weight of mandated testing, using assessments not yet aligned to the new Common Core curriculum standards.

We have seen the results of this approach and heard arguments against it from multiple states from New York to California. Basically, California chose to throw down the gauntlet by choosing to move forward and not backward. And many states and public school educators, including associations representing teachers and administrators, are now calling for delayed implementation of high stakes decisions based on student performance on the new assessments (Klein, 2014; Williams, 2013). Note, they are not necessarily rejecting the Common Core curricular reforms nor the need for high-quality assessments that can inform and support the system as it tries to make progress toward college and career readiness.

But what if things were different? What if instead of leaping to develop large-scale tests, as the Partnership for Assessment of Readiness for College and Careers and the Smarter Balanced Assessment Consortium have been funded to do, our government had taken the $350 million and used it in a different way? What if they had invested that money in developing and validating assessment tools and resources to help teachers in the classroom focus on the student performances and forms of deeper learning that are at the heart of the Common Core standards?

NGSS and assessments. Well, we might have a chance to answer that "what if" question in the case of science education. In spring 2013, "Achieve" issued the Next Generation Science Standards (NGSS; see www.nextgenscience.org) grounded in work done by the National Research Council (NRC) to develop a K-12 Science Education Framework. And in December 2013, the NRC issued a report focused on Developing Assessments for the NGSS (Pellegrino *et al.*, 2014). The NRC report argued that the competence called for substantially changes the game for science teaching and learning. Assessment poses a major design challenge, given that the standards focus on students' capacities to reason with and apply core disciplinary ideas in various areas of science.

While the design challenge is great, as it is for the Math and English Language Arts standards, the NRC Report does not argue that we should abandon developing assessments that monitor how well the system is doing in educating our youth. What it does argue for is a balanced system that includes three components:

(1) classroom assessments to support teachers and students;
(2) monitoring assessments for use at the state policy level; and
(3) indicators of opportunity to learn.

Of equal significance is the argument that the assessment system should be built from the bottom up or inside out – starting with the classroom level and working toward the monitoring level – just the reverse of what we have done with Common Core English Language Arts and Math assessments. By focusing on the classroom level first, we would better understand what the NGGS entail in terms of student knowledge, understanding and performance as well as the forms of evidence that are needed to support claims that students are developing the intended competencies. This can support development of the full range of assessments needed to support various classroom teaching and learning purposes. Not least among these are diagnostic assessments supportive of teacher and student enactment of FA practices.

Have we learned enough from our experience with Common Core and RTTP reforms so that the teaching, learning and assessment of the new science standards can profit from hindsight and move forward with a bit of foresight? Many of us certainly hope so.

Conclusion: balancing formative classroom assessment with assessment for accountability

Madhabi Chatterji, *Teachers College, Columbia University*[1], [2]

I belong among the growing number of scholars and practitioners in education calling for more widespread use of formative classroom assessment in America's schools (see Gordon; Kalinich, McGill; Pellegrino; Sands; and others, this issue). To connect research to practice and policy in this area, however, we need far more work and consensus building. "Formativity" in classroom assessment still means different things to different people, as some of the commentaries attest.

A diagnostic FA model. My kind of FA, appropriately conducted, involves tasks and exercises that are *diagnostic* in purpose and helpful in shedding light on specific, fine-grained learning needs of students as they grapple with particular domains. To close detected gaps effectively, diagnostic assessment tasks must be coupled with necessary follow-up *processes* that engage teachers and students together.

For example, take long division (LD) in math, a focus of our recent research in selected elementary schools in New York, supported by the National Science Foundation (Chatterji *et al.*, 2009; Chatterji, 2012). We confirmed that LD concepts pose significant challenges to many young learners, and some of these cognitive blocks eventually lead to persistent learning gaps at higher levels when students need facility with decimals or fractions. We identified several specific types of learning needs:

- students fail to grasp what a given LD problem is asking of them;
- they cannot begin the LD computational process;
- they cannot recall or apply the multiplication required;
- they misplace partial answers in the wrong place value column during the LD process; or
- they fail to see mathematical equivalents between answers expressed as fractions or decimals.

A big issue was in applying *place value* concepts during the computational process or in applying LD to solve real-life problems. Effective FA tasks reveal mental blocks like these that precipitate student errors, and they do so in ways that are *instructionally useful to teachers* while teaching and learning is still going on. This property is essential for the diagnostic and mediation processes that must follow.

With respect to *processes*, what kinds of activities would teachers who successfully close learning gaps undertake? Research suggests that teachers should be able to:

- embed diagnostic assessment tasks into their daily instructional repertoires;
- interpret student errors meaningfully with respect to the domains of expected learning outcomes;
- provide timely feedback to students to help them overcome specific learning barriers;
- mediate or re-teach concepts/skills in which learning needs are observed; and
- provide follow-up practice exercises to help students consolidate new learning (Chatterji, 2012; see also Sands, this issue).

But teachers cannot do it alone. Students' own awareness of what they are trying to learn, self-reflection on achieving learning goals and capacity for regulation of their own learning are key factors that influence student development and expertise building in a domain (Chatterji, 2012). My kind of FA, then, would be defined by both the *assessment products* (i.e. suitably designed and scorable diagnostic assessment items, performance tasks, quizzes or more formal assessment systems incorporating these) as well as contingent *assessment processes* (practices that teachers would typically facilitate, but that would involve students directly). We are talking about classroom assessment practices that are:

- proximal (*instead of* distal or removed from ongoing teaching–learning environments);
- diagnostic (yielding fine-grained profiles of learner strengths and weaknesses or errors in defined domains, *instead of* only summary scores on student progress);
- positive and supportive (whereby mistakes are viewed as opportunities to learn and improve, *rather than* as causes for stigma and fodder for punishing learners, teachers or schools);
- cyclic and ongoing (*instead of* simply culminating exercises, conducted when a lesson, unit, school term ends); and
- co-owned by teachers and learners and supported by schools and external agencies (*instead of* assessment programs that outside agents, schools or policymakers adopt and ask teachers to implement).

We coined the label "Proximal Assessment for Learner Diagnosis" (PALD) for a proposed model of diagnostic classroom assessment based on these five criteria. PALD draws on the talent development theory in medicine, sports and music and the latest knowledge bases from the pedagogical, cognitive and assessment sciences (Chatterji, 2012; see also Pellegrino *et al.*, 2001; Pellegrino *et al.*, 2014). PALD is based on the principle of proficiency-based and developmental learning and founded on the belief that everyone can succeed in a domain. When well-conducted, this form of assessment carries the potential for closing learning gaps

during the teaching learning process. Therefore, we believe, it can eventually lead to more equitable distributions of achievement in valued domains for *all* learners (Chatterji *et al.*, 2009; Chatterji, 2012).

We encounter evidence of FA being practiced creatively and innovatively in pockets (see Kalinich, this issue). But, like several commentators here, I do not believe teachers, schools, students or parents in American schools have enough education today to be discerning about sound and unsound formative classroom assessment. So, a first commitment of policy makers has to be toward allocation of *resources*, *training* and *supports* for all relevant stakeholders in this area, particularly teachers and school leaders (see Gordon; McGill; Pellegrino; and Sands, this issue).

Summarizing recommendations. It would be a mistake if today's anti-testing rhetoric in American public education pitted assessment for accountability against formative classroom assessment. We need *both* kinds of assessment in our schools. Each serves a different but complementary function in public education. Doubtless, each needs further development and improvement.

Authors have offered helpful recommendations for moving the field forward. In summary, they are:

- a shift in assessment priorities toward policies that promote *assessment for learning*, instead of undermining it (see Gordon; McGill; Pellegrino, this issue);
- a greater *commitment from the measurement profession* toward assessment for learning (Gordon, this issue);
- more *teamwork and collaboration* among policy makers/political leaders, measurement community, education scholars and school-based practitioners to bring about FA changes in schools (McGill, this issue);
- allocation of *resources* and *professional development* of educators and other stakeholders involved (Chatterji, Kalinich, McGill, Sands and others);
- utilization of frameworks and models that highlight *FA processes* with a recognition that FA involves *both* processes and products (Chatterji, Kalinich, Sands, this issue); and
- movement toward three-pronged assessment reforms implemented from the classroom upward and "inside out", with emphases placed on *classroom assessment*, *education system monitoring* and *opportunities to learn* for students (Pellegrino, this issue).

A search for the right balance, where well-done formative classroom assessment is coupled with better systems of school evaluation and accountability must therefore continue. New models of school evaluation and quality assurance should not conflict with or erase a genuine FA culture in schools. Rather, they should embrace it as a prerequisite condition for schools to be able to offer learning opportunities more equitably for all students to succeed.

Notes

1. Authors listed in presentation order. Biographical information follows at the end of the article.
2. Moderator. Biographical information follows at the end of the article.

References

Black, P., Harrison, C., Lee, C., Marshall, B. and Wiliam, D. (2003), *Assessment for Learning: Putting it into Practice*, Open University Press, Buckingham.

Calfee, R., Wilson, K., Flannery, B. and Kapinus, B. (in press). "Formative assessment for the common core literacy standards", *The Gordon Commission*, New York, USA.

Chatterji, M. (2012), "Development and validation of indicators of teacher proficiency in diagnostic classroom assessment", *The International Journal of Educational and Psychological Assessment*, Vol. 9 No. 2, pp. 4-25.

Chatterji, M. (Ed). (2013a), *Validity and Test Use: An International Dialogue on Educational Assessment, Accountability, and Equity*, Emerald Group Publishing, Bingley.

Chatterji, M. (Ed). (2013b), "When education measures go public: stakeholder perspectives on how and why validity breaks down", *Teachers College Record*, Vol. 115, No. 9.

Chatterji, M., Koh, N., Choi, L. and Iyengar, R. (2009), "Closing learning gaps proximally with teacher-mediated diagnostic classroom assessment", *Research in the Schools*, Vol. 16 No. 2, pp. 59-75.

Council of Chief State School Officers (2008), "Attributes of effective formative assessment: a work product coordinated by Sarah McManus, NC Department of Public Instruction for the CCSSO Formative Assessment for Students and Teachers (FAST) SCASS", CCSSO, Washington, DC.

Guskey, T.R. (2010), "Lessons of mastery learning", *Educational Leadership*, Vol. 68 No. 2, pp. 52-57.

Haertel, E. (2013), "A vision for the future", Report presented at a Symposium conducted by the Gordon Commission on the Future of Assessment in Education.

Hattie, J. and Timperley, H. (2007), "The power of feedback", *Review of Educational Research*, Vol. 77 No. 1, pp. 81-112.

Klein, A. (2014), "NEA criticizes "botched" common core implementation", *Education Week*, available at: http://blogs.edweek.org/edweek/campaign-k-12/2014/02/nea_criticizes_botched_common_.html. (accessed 19 February 2014)

McManus, S. (2006), "Attributes of effective formative assessment", The Council of Chief State School Officers, available at: www.ncpublicschools.org/docs/accountability/educators/fastattributes04081.pdf

Marzano, R.J. (2010), *Formative Assessment and Standards-Based Grading*, Solution Tree, Bloomington.

Nitko, A.J. (1989), "Designing tests that are integrated with instruction", in Linn, R.L. (Ed.), *Educational Measurement*, 3rd ed., American Council on Education and MacMillan Publishing, New York, NY, pp. 447-474.

No Child Left Behind (NCLB) Act of 2001 (2002), Pub. L. No. 107-110, § 115, Stat. 1425.

Pellegrino, J.W., Chudowsky, N. and Glaser, R. (Eds) (2001), *Knowing What Students Know: The Science and Design of Educational Assessment*, The National Academies Press, Washington, DC.

Pellegrino, J.W., Wilson, M.R., Koening, J.A. and Beatty, A.S. (Eds) (2014), *Developing Assessments for the Next Generation Science Standards*, The National Academies Press, Washington, DC.

Popham, W.J. (2008), *Transformative Assessment*, Association for Supervision and Curriculum Development, Alexandria, VA.

Ravitch, D. (2011), *The Death and Life of the Great American School System: How Testing and Choice Are Undermining Education*, Random House Publishers, New York, NY.

Ravitch, D. (2013), *Reign of Error: The Hoax of Privatization and the Danger to America's Public Schools*, Random House Publishers, New York, NY.

Shepard, L.A. (2000), "The role of assessment in a learning culture", *Educational Researcher*, Vol. 29 No. 7, pp. 4-14.

The Gordon Commission (2013), "Assessment for learning: a public policy statement", available at: www.gordoncommission.org/rsc/pdfs/gordon_commission_public_policy_report.pdf

Williams, C.S. (2013), April 8), "Allow more time for implementation of common core state standards", *Learn First Alliance*, available at: www.learningfirst.org/allow-more-time-implementation-common-core-state-standards

Author Affiliations

Edmund W. Gordon, Teachers College, Columbia University, New York, NY, USA and Yale University, New Haven, CT, USA

Michael V. McGill, Bank Street College of Education, New York, NY, USA

Deanna Iceman Sands, Seattle University, Seattle, WA, USA

Kelley M. Kalinich, Kenilworth School District, Kenilworth, IL, USA

James W. Pellegrino, University of Illinois at Chicago, Chicago, IL, USA

Madhabi Chatterji, Teachers College, Columbia University, New York, NY, USA

About the authors

Edmund W. Gordon is the John M. Musser Professor of Psychology, Emeritus at Yale University, Richard March Hoe Professor of Psychology and Education, Emeritus at Teachers College, Columbia University and Director Emeritus of the Institute for Urban and Minority Education at Teachers College, Columbia University.

Michael V. McGill just retired from the post of Superintendent of Schools at the Scarsdale School District, New York, and is now the Director of Programs in District Leadership and Reform at Bank Street College of Education. He was named the New York State Superintendent of the Year in 2007.

Deanna Iceman Sands is the Dean of the College of Education at Seattle University. Her more recent research interests are in formative classroom assessment.

Kelley M. Kalinich has served as Superintendent of Schools in the Chicago area for the past 15 years, following roles as Principal, Assistant Principal and Special Education Teacher/Facilitator. She holds a doctorate from Peabody College of Education at Vanderbilt University.

James W. Pellegrino is Co-director of the Learning Sciences Research Institute, Liberal Arts and Sciences Distinguished Professor and Distinguished Professor of Education at the University of Illinois at Chicago.

Madhabi Chatterji is Associate Professor of Measurement, Evaluation and Education and Founding Director of the Assessment and Evaluation Research Initiative at Teachers College, Columbia University (AERI@TC). Madhabi Chatterji is the corresponding author and can be contacted at: mb1434@tc.columbia.edu