

2014

Proficiency-Based Grading: Can We Practice What They Preach?

Julie Shalhope Kalnin
University of Portland, kalnin@up.edu

Follow this and additional works at: http://pilotscholars.up.edu/edu_facpubs



Part of the [Education Commons](#)

Citation: Pilot Scholars Version (Modified MLA Style)

Shalhope Kalnin, Julie, "Proficiency-Based Grading: Can We Practice What They Preach?" (2014). *Education Faculty Publications and Presentations*. 35.

http://pilotscholars.up.edu/edu_facpubs/35

This Journal Article is brought to you for free and open access by the School of Education at Pilot Scholars. It has been accepted for inclusion in Education Faculty Publications and Presentations by an authorized administrator of Pilot Scholars. For more information, please contact library@up.edu.

Proficiency-Based Grading: Can We Practice What They Preach?

Julie Shalhope Kalnin, Ph.D.
University of Portland

Abstract

The practice of assessing and grading students on their ability to demonstrate proficiency related to a standard has grown significantly in K–12 settings over the past decade. This article invites teacher educators to consider how to respond to this trend by examining an approach for preparing teacher candidates to participate effectively in proficiency-based systems. The process of integrating a proficiency-based grading strand into one undergraduate assessment course for elementary teacher candidates is described to illustrate how an approach advocated in the K–12 environment may be adapted to a higher education setting. Reflections on both the challenges and the benefits of the adaptation suggest that approaching course design from a proficiency-based perspective, while perhaps most valuable in a course that builds candidates' assessment literacy, can also enhance teacher educators' efforts to design purposeful course experiences in other arenas.

***Keywords:* assessment, preservice teachers, proficiency-based grading, teacher preparation**

Kalnin

The phrase “teacher preparation” expresses a central relationship between education programs and the profession. In order to prepare teachers well, programs must be responsive to the contexts that their candidates can anticipate entering. Yet, shifting mandates and emerging practices that could prove either fads or substantive reforms complicate the responsive stance. For teacher educators, teacher preparation is more than job training; it is the education of professionals who will not only work within, but also will have the knowledge, skill, and disposition to influence, the educational system. Tensions can arise as individual teacher educators and programs wrestle with questions of how to be both responsive to, and stewards of, the profession we serve.

Assessment is perhaps the most prominent area in which such tensions have surfaced since the legislation of No Child Left Behind (NCLB) in 2001. Within the NCLB assessment context, teacher educators have thoughtfully explored ways to support their candidates in learning about (Reeder & Utley, 2008), adjusting to (Bates & Burbank, 2008), or complying with NCLB while resisting its testing-focused student characterizations and curriculum constraints (Samuel & Suh, 2012; Taylor, 2010; Wepner, 2006; Whitenack & Swanson, 2013).

Educators beyond the academy have also developed responses to counter the fragmenting effects of NCLB. One response that is increasingly being adopted in K–12 settings is proficiency- or standards-based grading. Essentially, proficiency-based grading emphasizes a more holistic focus on the standards behind standardized testing. The approach bases grades not on factors such as attendance or behavior, but on a student’s ability to demonstrate understanding and skill in relation to a standard.

In a 2013 interview, Robert Marzano concluded, “Standards-based grading is beginning to grow exponentially” (Koumpilova, 2013). A brief internet search can confirm this statement. A search for proficiency- or standards- based grading will yield results from districts from Florida to Colorado, North Carolina to Washington, and run a solid swath through the Midwest and down the Eastern seaboard. During the preparation of this article, 55 post-secondary

Proficiency-Based Grading

institutions in New England formally signed the Collegiate Endorsement of Proficiency-Based Education and Graduation. This straightforward statement supports “proficiency-based approaches to instruction, assessment, and reporting” and signers pledge to accept a wide range of transcripts to ensure that students with proficiency-based records will be in no way disadvantaged (New England Secondary School Consortium, 2014). Evidence that proficiency-based grading has a positive impact for student learning is emerging (Pekel, 2013; Welsh, D’Agostino, & Kaniskan, 2013; Winters & Cowen, 2012). As proficiency-based grading’s acceptance grows, teacher educators need to consider whether—and how—to respond to this assessment movement.

Re-examining Proficiency-Based Grading

Although proficiency-based grading is highly visible today, the precepts are not new. In his 1998 book, *Educative Assessment*, Grant Wiggins advocated assessing and reporting student performance using proficiency-based grading. The practice has roots in mastery learning (e.g. Bloom, 1968; Block & Burns, 1974) and clearly shares theoretical and practical perspectives with outcome-based education (e.g. Spady, 1994). Resnick’s (1999, 2005) discussions of effort-based learning and the need to more clearly define standards are also close associates. Those shared theoretical underpinnings are most clearly defined in the Principles of Learning advocated by the Institute for Learning (IFL), which include a focus on “organizing for effort,” “clear expectations,” and “fair and credible evaluations” (IFL, 2014).

Certainly, teacher educators are no strangers to the concept of assessing using proficiency scales. With the rapid deployment of EdTPA by 34 states and 522 teacher preparation programs across the nation (AACTE, 2014), we can expect to become only more skilled in evaluating our candidates from this perspective. In my experience, however, proficiency assessment is a tool often restricted to clinical placements. Preparing our students for EdTPA may change this restriction, as experiences in coursework are likely to be more intentionally aligned with the performance assessment.

Kalnin

Even without the influence of EdTPA, though, evidence about grade inflation in schools of education (Koedel, 2011; Nikolakakos, Reeves, & Shuch, 2012) should spur teacher educators to reexamine our grading practices.

In this article, I will describe the process I went through to develop an approach for integrating a proficiency-based assessment strand in a semester-long undergraduate course for elementary teacher candidates. I describe an experience that provided me with a compelling rationale for change, outline how I adapted the K–12 proficiency grading model to a higher education setting, and reflect on the challenges and benefits of this approach.

Learning from Experience

I first was introduced to the current iteration of proficiency-based grading in 2007 when I was teaching in a unique university-district partnership between Austin Public Schools and the University of Minnesota. The Hormel Foundation had generously funded a masters' cohort designed to engage the district's teachers in improving literacy, math and science instruction. Intensive summer sessions focused on disciplinary content; year-long courses emphasized implementation of that content through providing curriculum design activities, strengthening collaborative practices such as peer coaching, and infusing data analysis into instructional decision-making.

The first course in the program was a semester-long introduction to teacher leadership and collaboration; the goal was to begin de-privatizing teaching practice (Louis, Marks, & Kruse, 1996) through meaningful professional conversations (Garmston & Wellman, 1998). One assignment was for the teachers to form book study groups. In addition to the books I had offered as options, I encouraged teachers to bring in books they wanted to read with colleagues. One group, Lynn Hemann, Eric Harder and Curtis Bartlett decided to read O'Connor's (2007) *A Repair Kit for Grading, 15 Fixes for Broken Grades*. Their choice stemmed initially from frustration. How could students' unrelenting negotiations about grades be redirected? Through their discussion of O'Connor's text,

Proficiency-Based Grading

though, the group began to consider how to change grading practices for a more fundamental reason—supporting student learning. O'Connor's text stresses that teachers have a responsibility to ensure that grades are accurate reflections of what students have learned, not rewards for attendance or cooperative behavior. When the standards for grades are meaningful, transparent, and fair, grades can support, rather than retrospectively label, the learning process (O'Connor, 2007). O'Connor's ideas provided an impetus for change.

In the next implementation-focused course of their masters' program, Bartlett and Harder, who both taught eighth-grade pre-algebra, opted to prepare for proficiency-based grading by developing a curriculum map that aligned their course materials with state standards; in determining unit objectives the teachers referred particularly to state testing specifications. Then, with district assistance, the two carried out data analysis that further supported their rationale for incorporating O'Connor's recommendations. In comparing their students' semester grades with their scores on state math assessments, it was clear that grades were not correlated with scores on the state test ($r^2=.194$) (Pekel, 2013). The two were particularly surprised to see that a small group of students who had received "A's" in the course had failed to achieve the level of "proficient" on the state test. Bartlett and Harder expressed concern that they had been "lying" to those students and their parents (Bartlett, Harder, & Berglund, 2009). How could a student receive an "A" in a class oriented to state standards and yet not be able to respond with 60% accuracy on the state assessment? After discussion with their principal, Katie Berglund, the two piloted proficiency-based assessment in eighth grade pre-algebra. Their approach adhered to O'Connor's principles. Grades were based only on unit-test scores. While homework was not counted toward the grades, Harder and Bartlett did require that students complete all homework before retaking a test. The two made themselves available to students before and after school and at lunch to provide supplementary instructional support.

The following summer, Harder, Bartlett and Berglund repeated

Kalnin

the correlational analysis and saw two important changes. First, the grades were more strongly correlated with the state test results ($r^2=.42$) (Pekel, 2013) and second, no students who received an “A” in the course had failed to demonstrate proficiency. Using this pilot as a catalyst, Berglund, Harder, and Bartlett launched a school-wide shift toward proficiency-based grading. Throughout, Berglund skillfully supported Harder and Bartlett’s teacher leadership as she cultivated district support, strengthened faculty acceptance, and educated parents about the initiative (see Pekel, 2013 for a full description).

Bringing Proficiency-Based Grading to a Teacher Preparation Course

These educators’ experimentation and documentation had offered such a compelling case for incorporating proficiency-based grading in schools that I saw it as my responsibility to introduce future teachers to this practice. Now at the University of Portland, I decided to apply what I had learned to a new assignment: an assessment course for undergraduate elementary teacher candidates.

When I began planning for the course, I came to see that the resources available for teacher educators were not well integrated. If the popular textbooks on assessment that I reviewed mentioned proficiency-based grading at all, they offered little guidance for how to implement such an approach. Books and articles on grading practices, on the other hand, dealt minimally with assessment issues. If my students were to gain insight into the process, I decided I would have to develop an experiential approach that linked assessment and grading. I couldn’t just give witness to the potential power of the approach; I needed to learn how to practice proficiency-based grading myself.

Defining Proficiency

Whether the practice is termed “standards-based” or “proficiency-based” grading, the principle is the same. A student is judged on her ability to demonstrate, *to a specified level of performance*, understanding or skill as articulated in a standard (or

Proficiency-Based Grading

portion of a standard). The first problem, of course, is what that specified level of performance should be. What will count as proficient?

I used two techniques to define proficiency in the course. In higher education, the process of determining course content is more reliant on the instructor's personal judgment than might be true in PK–12 settings, but that judgment is guided in most colleges by accreditation standards, course objectives, and unit-level frameworks (i.e. conceptual framework). In addition to the course description, I turned to InTASC Standard 6: “The teacher understands and uses multiple methods of assessment to engage learners in their own growth, to monitor learner progress, and to guide the teacher's and learner's decision making.” (CCSSO, 2011). Based on these guidelines and drawing on my selected textbook, Popham's (2011a) *Classroom Assessment: What Classroom Teachers Need to Know*, I established two main areas of focus: general assessment literacy and types of assessments. In a related article, Popham (2011b) defined assessment literacy usefully. “Assessment literacy consists of an individual's understandings of the fundamental assessment concepts and procedures deemed likely to influence educational decisions” (p. 267). I developed an initial list of concepts that I felt met this definition.

The second element that I used to define student proficiency was the concept of cognitive complexity. Designers of large-scale assessments operationalize cognitive complexity variously as “demands on thinking,” “question-level demand” or “depth of knowledge” (Schneider, Huff, Egan, Gaines & Ferrara, 2013). Similarly, from a classroom perspective, defining a performance standard requires that an instructor articulate the level of thinking required by a task. Classroom teachers often turn to Bloom's 1968 taxonomy of learning objectives as a familiar framework for cognitive complexity, and, as discussed above, Bloom's taxonomy is well aligned with the theoretical underpinnings of proficiency-based grading. I chose to use an updated version of the taxonomy (Anderson & Krathwohl, 2001) that provided clear guidelines for distinguishing among factual, conceptual, procedural, and

Kalnin

metacognitive knowledge (Krathwohl, 2002).

Using the concept of cognitive complexity allowed me to differentiate between those concepts that I wanted students to know, those I wanted them to apply, and those I wanted them to be able to act on creatively themselves. For instance, it was important that students could define the concept of content validity, but given the definition of assessment literacy I was aiming for, I also wanted to ensure that students could apply this concept to a realistic scenario they might encounter as professionals. On the other hand, for concepts like “norm”- or “criterion”-referenced tests I determined proficiency sufficient at the factual level (see Appendix A: Column 2). Another teacher educator might make different decisions about what aspects of assessment literacy should be developed to a certain level of cognitive complexity, but the process of choosing, and explicitly articulating my choices, was a prerequisite for implementing a proficiency-based approach in the course.

At the same time, defining the levels of cognitive complexity that I was hoping students to achieve showed that I could not successfully apply a proficiency-based approach to every assignment in the class. Given the goal that students create a wide variety of assessment types (formative assessments, multiple choice, performance assessments, and a portfolio), and the constraint of a fifteen-week semester, I decided to use traditional, rubric-based evaluation and grading for the complex assessments students would be asked to create. I applied the proficiency-based approach in only one strand of the course—the general assessment literacy dimension. This strand incorporated key concepts related to assessment design, purposes of assessment types, and interpretation of assessment information (Appendix A: Sample Student Proficiency Record, Columns 2, 4, & 5). The “proficiency” portion of the students’ grade contributed 25% to the total semester grade. Exceeding proficiency on all items resulted in full-credit; achieving proficiency on all items resulted in 85% credit. Not meeting on one item reduced credit to 80%; not meeting on up to three items reduced the credit to 70%; not meeting on four or more items resulted in a failing grade (60%).

Implementation

Proficiency-based models emphasize the importance of summative tests, but also of allowing for retesting, or eliminating low scores with cumulative assessments. To demonstrate this approach for students, I designated three “check-up” assessments during the course with an optional supplement to the final exam for students who had not yet fully demonstrated proficiency. Each assessment “check-up” focused primarily on the information that had been addressed during that portion of the course. Factual or conceptual level items asked students to identify or define concepts. Application tasks engaged students in realistic problem-solving scenarios such as choosing between different types of assessments, placing students in instructional groups based on assessment results, or interpreting actual score reports (Appendix B: Sample Check-up).

After each “check-up” students received an individual print-out showing whether they had exceeded proficiency (E), attained proficiency (P) or were not yet proficient (NYP) on tested concepts (Appendix A: Column 7). I created the individual proficiency records by first entering the students’ scores in Excel. I entered the data in rows by student name, with item scores in the adjoining columns. This allowed me to interpret how well the class as a whole had responded to a certain item, an analysis I also modeled for the class. I used the spreadsheet to set up a mail merge document in Microsoft Word that generated the individual reports for each student, giving the concepts, item scores, and proficiency related to a concept; proficiency status was reported only for the highest level of cognitive complexity required for that concept.

With this information in hand, if I saw that a large group of students had not demonstrated proficiency on an item, I would establish differentiated groups during class time to offer instructional activities designed to clarify misconceptions. If smaller clusters or individual students were not yet proficient, I set up study times outside of class for a concept, or invited students to meet with me independently.

After the first “check-up,” I demonstrated proficiency-based

Kalnin

grading by individualizing students' future assessments. A student who had demonstrated proficiency on all of the concepts would have only new concepts assessed on the next "check-up." A student who was not yet proficient on an item would be retested on that concept. My own assessment knowledge was stretched as I struggled to write new—and yet equivalent—items at the indicated level of cognitive complexity to re-assess concepts. Again, I employed Excel and the mail merge feature in Microsoft Word to create individualized assessments. Through the proficiency reports, communication about expectations and achievement was clear—not only in terms of a grade, but in terms of progress toward knowledge that students saw as important to them in their future professional practice.

Student results on personalized assessments demonstrated proficiency-grading directed effort toward learning. Students soon observed that not studying for an assessment was likely to mean they could not demonstrate proficiency. The consequence? On the next assessment, in addition to new content, concepts that hadn't been mastered would be reassessed. I remember one student looking down at her status report, noting the concepts she would see again on the next check-up and wryly commenting, "I guess you really want us to learn this stuff."

In two years, all but one of the students in the course did achieve proficiency on all of the concepts by the end of the semester. Every individual in both sections retested in at least one area at least once. Fewer than five in each section retested on multiple concepts multiple times. One student retested all three times on multiple concepts. The teacher candidates experienced the frustrations one of their own students might feel—of having to face a challenging concept yet *again* on the next "check up," but they also experienced the motivating lure of "not yet." Students saw the value in having another chance to show they could master concepts that had initially confused them. At the end of the course, one student wrote, "When I saw the list of assessment concepts at the beginning of the course, I thought there was no way I could ever learn all of that. But after each check-up, I saw my progress and it gave

Proficiency-Based Grading

me confidence. I did learn all of the concepts. I'm proud of myself' (Anonymous comment, Course Evaluation). Based on the results from two sections, incorporating proficiency-based assessment effectively supported my students' assessment literacy.

Implications

While implementing proficiency-based grading was challenging, I found the approach was powerful at illustrating key assessment concepts such as proficiency, content validity, reliability, and the relationship between assessment, curriculum, and instruction. Using my own assessments as illustrations (carefully designed, but imperfect, as all teacher-made assessments are) was invaluable for class discussion. The explicitness with which I was required to define proficiency for my students clearly illustrated to them how thoroughly individual professional judgment or the shared judgments of a professional community are integrated into assessment design and interpretation. By making my own decisions fully transparent to students in the form of their proficiency status reports, I could invite discussion about my decisions. Had I emphasized the concepts and skills that students were finding foundational in their practice? Had I expected enough? Or too much? With our shared experience at the center, the concept of assessment validity—content validity, construct validity, and consequential validity (Popham, 2011a)—became real to students.

Using proficiency-based assessment to teach proficiency-based assessment also allowed me to engage students in analyzing every step of the design process and to examine the logistics of scoring, of recording proficiencies, and of tailoring assessments to individual's proficiency levels. Using this illustration, we were able to brainstorm about how to address these logistical issues in an elementary classroom.

Finally, students saw firsthand that the assessment process was a guide for differentiated instruction as they participated in varied group assignments during class time or attended study groups and individual sessions offered outside of class. We could talk frankly and specifically about the time required to use this approach, but

we could balance that concern with recognition of the impact of the investment. Each of us experienced and interpreted proficiency-based grading from multiple perspectives: learner and teacher. Our collective inquiry brought the dynamic tension between theory and practice—a foundational aspect of teacher education in the liberal arts context—to life.

Conclusions

Implementing a proficiency-based system, even in one aspect of a single course, has taught me even more about the need to continually “plan backwards” (Wiggins & McTighe, 2006). Of course, I had believed that was what I had been doing all along. Pushing myself to specify goals in terms of their priority and to clarify the level of cognitive demand I expected moved me further down a path I thought I already knew well.

Perhaps the most important result of implementing this approach was a good solid serving of humility. As so many of my respected and hard-working colleagues in teacher preparation strive to do, I have always sought to practice what I preach. But honestly, implementing this discrete attempt at proficiency-based grading illuminated unanticipated complexities. Even though I hope that I convincingly advocated for proficiency-based grading through the model I offered, I couldn’t fully show my students how demanding, time-consuming, or controversial implementing proficiency-based grading in their own classrooms or schools might be. I had undertaken a limited innovation in one course. Implementing this work on a yearlong basis, with other competing demands, questioning parents, and supportive or concerned colleagues, would be a wholly different matter.

To answer the question posed in the title of this paper, then, when teacher educators attempt to “practice what they preach” we may want to notice what we cannot achieve as much as what we can. As my experience demonstrates, our teacher candidates can benefit in multiple ways when teacher educators learn from colleagues in PK–12 settings and responsively incorporate their skillful practice in our programs. The first benefit can be immediate, as

Proficiency-Based Grading

our students derive some of the same positive learning effects that PK–12 educators documented in their classrooms and schools. My students’ assessment literacy was deepened because of the proficiency-based approach I had learned from Bartlett and Harder’s compelling example. The second benefit, though, cannot fully be realized until our candidates become teachers in their own right and are further mentored by our PK–12 partners. Preparation is truly an education when candidates are able to carry what they’ve learned into a professional community where they can apply, reevaluate, and extend initial understandings

References

- American Association of Colleges of Teacher Education [AACTE] (2014). EdTPA participation map. Retrieved from <http://edtpa.aacte.org/state-policy>
- Bartlett, C., Harder, E., & Berglund, K. (2009). Implementing school level change from the ground-up: Assessing for proficiency at Austin Middle School. Presentation at University of Minnesota Instructional Leadership Course: July, 2009.
- Bates, A. J., & Burbank, M. D. (2008). Effective student teacher supervision in the era of “No Child Left Behind.” *Professional Educator*, 32(2), 1-11.
- Block, J. H., & Burns, R. B. (1974). Mastery learning. *Review of Research in Education*, 4, 3-49.
- Bloom, B. (1968). Learning for mastery. *Evaluation Comment*, 1(2).
- Council of Chief State School Officers [CCSSO](2011). The InTASC Model Core Teaching Standards at a Glance. Retrieved from <http://www.ccsso.org/Documents/2011/InTASC%202011%20Standards%20At%20A%20Glance.pdf>
- Garmston, R., & Wellman, B. (1998). Teacher talk that makes a difference. *Educational Leadership*, 55(7), 30-34.
- Institute for Learning [IFL] (2014). Principles of learning. Retrieved from http://ifl.lrdc.pitt.edu/index.php/resources/principles_of_learning

Kalnin

- Koedel, C. (2011). Grading standards in education departments at universities. *Education Policy Analysis Archives*, 19 (23), 1-23.
- Koumpilova, M. (2013). Minnesota schools give standards-based grading system a closer look. http://www.twincities.com/ci_22896205/minnesota-schools-give-standards-based-grading-system-closer
- Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into Practice*, 41(4), 212-264.
- Louis, K. S., Marks, H. M., & Kruse, S. (1996). Teacher's professional community in restructuring schools. *American Educational Research Journal*, 33(4), 767-798.
- New England Secondary School Consortium (2014). Collegiate endorsement of proficiency-based education and graduation. Retrieved from: <http://newenglandssc.org/resources/endorsement>
- Nikolakakos, E., Reeves, J. L., & Shuch, S. (2012). An examination of the causes of grade inflation in a teacher education program and implications for practice. *College and University*, 87(3), 2-13.
- O'Connor, K. (2007). *A repair kit for grading: Fifteen fixes for broken grades*. Portland, OR: Educational Testing Service.
- Pekel, K. (2013). *A tale of two grades: An evaluation of grading for learning, a middle school grading reform* (Doctoral dissertation). Retrieved from Proquest UMI Dissertation Publishing (UMI No. 3605363).
- Popham, W. J. (2011a). *Classroom assessment: What teachers need to know* (6th ed.). Upper Saddle River, NJ: Pearson.
- Popham, W. J. (2011b). Assessment literacy overlooked: A teacher educator's confession. *Teacher Educator*, 46(4), 265-273.
- Reeder, S., & Utley, J. (2008). Elementary teacher candidates' understanding of the No Child Left Behind legislation. *Teacher Education and Practice*, 21(1), 89-102.
- Resnick, L. B. (1999, June 16). Making America smarter. *Education Week*, pp. 38-40.
- Resnick, L. B., & Zurawsky, C. (2005). Getting back on course: Standards-based reform and accountability. *American Educator*, 29(1), 8-46.

Proficiency-Based Grading

- Samuel, F. A., & Suh, B. (2012). Teacher candidates reconcile “the child and the curriculum” with “No Child Left Behind.” *Educational Forum*, 76(3), 372-382.
- Schneider, M. C., Huff, K. L., Egan, K. L., Gaines, M. L., & Ferrara, S. (2013). Relationships among item cognitive complexity, contextual demands, and item difficulty: Implications for achievement-level descriptors. *Educational Assessment*, 18(2), 99-121.
- Spady, W. G. (1994). *Outcome-based education: Critical issues and answers*. Arlington, VA: American Association of School Administrators.
- Taylor, R. W. (2010). The role of teacher education programs in creating culturally competent teachers: A moral imperative for ensuring the academic success of diverse student populations. *Multicultural Education*, 17(3), 24-28.
- The taxonomy table. (2005). In D. Anderson & D. Krathwohl (Eds.), *Taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Upper Saddle River, NJ: Pearson.
- Welsh, M. E., D'Agostino, J. V., & Kaniskan, B. (2013). Grading as a reform effort: Do standards-based grades converge with test scores? *Educational Measurement: Issues and Practice*, 32(2), 26-36.
- Wepner, S. B. (2006). Testing gone amok: Leave no teacher candidate behind. *Teacher Education Quarterly*, 33(1), 135-149.
- Whitenack, D. A., & Swanson, P. E. (2013). The transformative potential of boundary spanners: A narrative inquiry into pre-service teacher education and professional development in an NCLB-impacted context. *Education Policy Analysis Archives*, 21(57).
- Wiggins, G., & McTighe, J. (2006). *Understanding by design* (Expanded 2nd ed.). Upper Saddle River, NJ: Pearson.
- Winters, M. A., & Cowen, J. M. (2012). Grading New York: Accountability and student proficiency in America's largest school district. *Educational Evaluation and Policy Analysis*, 34(3), 313-327.

Appendix A: Sample Student Proficiency Report

1	2	3	4	5	6	7
Item #	Concept	Cognitive complexity of task	General assessment literacy	Types of assessment	Item score	Proficiency status
4a	Cognitive complexity	Knowledge	Assessment design and use		2/2	
4b	Cognitive complexity	Application	Assessment design and use	Pre-assessment	2/5	NYP
9	Cognitive validity	Knowledge	Assessment design and use	Summative assessment	2/3	
10a	Cognitive validity	Application	Assessment design and use	Summative assessment	2/3	
10b	Cognitive validity	Application	Assessment design and use	Summative assessment	2/3	P
6	Construct validity	Application	Assessment design and use	Performance assessment	3/5	P
7	Predictive validity	Application	Assessment design and use	Performance assessment	1/5	NYP
1c	Stability reliability	Knowledge	Assessment design and use	Standardized test	2/3	P
3	Validity	Application	Assessment design and use	Pre-assessment	5/5	E
8a	Content standards	Knowledge	Curriculum relation to assess		0/2	NYP
8b	Performance standards	Knowledge	Curriculum relation to assess		0/2	NYP
1a	Percentile rank	Knowledge	Score interpretation	Standardized test	1/3	NYP
2a	Criterion referenced	Application	Score interpretation	Standardized test	2/2	
2c	Criterion referenced	Application	Score interpretation	Standardized test	2/2	E
1b	Grade equivalent	Knowledge	Score interpretation	Standardized test	3/3	E
2b	Norm-referenced	Application	Score interpretation	Standardized test	2/5	NYP
5	Standard error	Application	Score interpretation	Standardized test	5/5	E

Proficiency-Based Grading

Appendix B: Sample Check-Up

Check-Up Number One: Your Name: _____

1. A student's mother asks, "What does it mean when it says that Yvonne's percentile rank is 60?" The most accurate answer would be-
 - a. Yvonne is smarter than 60% of the other students in the class.
 - b. Yvonne's score indicates that she correctly identified 60% of the items.
 - c. Yvonne has mastered 60% of the criteria that are being assessed.
 - d. Yvonne's score is higher than 60% of scores of other test-takers.
 - e. Both b and c.
2. You are working with colleagues to evaluate whether a specific test would be an effective pre-assessment of student's mathematical knowledge and skill. One of your colleagues asks, "Do we have any information about the cognitive complexity of the items?"
 - a. What does cognitive complexity mean in this context?
 - b. Why might knowing about the cognitive complexity of test items be particularly important for a pre-assessment?
3. Carlo's parents are very concerned about his scores on a recent standardized test. They want to know whether he can be retested because they really believe he will get a better score. When you confer with your school counselor, she tells you that this test is very reliable (.7 stability reliability). Explain in terms a parent could understand whether this information indicates Carlo should or should not be retested.
4. A team of teachers are meeting to group all of the students in the grade level using individual test reports like the one below. These groupings will be used to create classes for differentiated reading instruction. Before the group starts to analyze the data, the team lead reminds the group that the standard error for this test was given as +/- 10. One teacher looks at Vijay's score (figure supplied) and says that given the standard error for the test, he should be put in the "meets" group rather than the "exceeds." How would you respond, and why?
5. Briefly define the difference between a "norm-referenced" and a "criterion-referenced" interpretation of test results.
6. On the report below (figure supplied) identify which of the numbered items is a criterion-referenced interpretation of the test results, and which is a norm-referenced interpretation. For each choice, explain your reason.
 - a. Item C1 is a criterion-referenced norm-referenced interpretation because : (Circle one)
 - b. Item C2 criterion-referenced norm-referenced interpretation because (Circle one)
 - c. Item C3 criterion-referenced norm-referenced interpretation because
7. Name two criteria that help to define the concept of "formative assessment."

Kalnin

8. Evaluate the following scenario. Identify whether the scenario “qualifies” as a good example of a formative assessment. Explain your choice.

Randy’s first grade class has just begun a science unit on “sink and float.” After reading some fiction and non-fiction texts about these concepts, the class is engaged in its first “hands-on” activity. Students are working in pairs and creating boats from a wide-variety of materials. Before they can “launch” their craft in the inflatable wading pool Randy has filled at the front of the room, they must come to him or the paraprofessional to answer a series of questions such as “Do you think this will float?” “What makes you think so?” Randy and the para record the information for each one of the groups. The students then must try it out and return to report whether their prediction was correct and why. After school, Randy sits down to look at the student’s responses and plan for tomorrow. He divides the group into the “sinkers” and the “floaters.” The “sinkers” receive the same task with the same materials; the floaters are given different materials that make creating a “floating” craft considerably harder.

This example QUALIFIES or DOES NOT QUALIFY (circle one) as an example of formative assessment because . . .

9. If you said that Randy’s scenario is NOT an example of a formative assessment, what would need to change so that it qualified? If you said it WAS an example of a formative assessment, what would need to be added to make it even stronger OR meet further criteria for formative assessment (e.g. Popham’s levels)?

Julie Kalnin is an assistant professor at the University of Portland, where she teaches courses in curriculum and assessment as well as teacher leadership and school change. Her research interests focus on teacher professional development across the career span, with particular attention to school-university partnerships.