

The impact of an alternative professional development model on teacher practices in formative assessment and student learning

Jack Robinson^a, Steve Myran^{a*}, Richard Strauss^b and William Reed^c

^a*Department of Educational Foundations and Leadership, Old Dominion University, Norfolk, VA, USA;* ^b*Saint Patrick Catholic School, Norfolk, VA, USA;* ^c*Norfolk Public Schools, Norfolk, VA, USA*

(Received 2 May 2011; final version received 4 February 2013)

Among contemporary means of enhancing student learning, formative assessment is perhaps one of the most important and effective. While formative assessment ideas and practices have been shown to have a proven record enhancing student learning, these practices are slow to be fully integrated into teachers' day-to-day classroom practices. This study describes a collaborative effort among university faculty and public school partners to train teachers in the skills and practices of formative assessment. Regarding teachers' involvement in the formative assessment professional development, findings highlight that 1) teachers' participation in the professional development efforts did strengthen their understanding of both general knowledge of formative assessment and the use of formative assessment practices, 2) teachers' plans to use the strategies in the future were related to their understanding of these strategies, and 3) in-depth and comprehensive understanding of formative assessment practices were critical to concrete applications of such practices in their classrooms. With respect to the impact on student learning, an overall effect size of .41 was found for teachers who utilized a formative assessment strategy compared with district averages for similar learning objectives when the practices were not used.

Keywords: formative assessment; professional development; school–university partnerships

This paper describes a professional development collaboration between university professors and public school personnel on developing the skills and knowledge of formative assessment, or what some have referred to as assessment for learning (AFL). This partnership involved two professors, two central office testing and accountability administrators, and a group of subject area specialists, principals, and teachers. Our approach to this collaboration evolved over two years. In the first year we met with teachers and central office staff at one school during teacher grade-level and planning meetings in order to help teachers refine their basic understanding of formative assessment theory and practices and to spark interests in evolving the effort beyond this initial pilot group. These early meetings, while not as productive as we had hoped, proved invaluable in terms of better understanding the climate of the school and the needs of the teachers. While these educators were being asked to increasingly utilize student performance data in formative ways, it became apparent

*Corresponding author. Email: smyran@odu.edu

that teachers lacked adequate backgrounds and understandings of the fundamental principles of formative assessment to fully capitalize on the uses of these practices.

These early observations and feedback from the public school personnel emphasized the importance of the following:

- (1) offering training that fit into the perceived needs of the participating schools;
- (2) recognizing/respecting teachers' background knowledge;
- (3) avoiding being perceived as an add-on to teachers' already heavy workload;
- (4) focusing on strategies teachers could use in their daily work with students;
- (5) providing opportunities for collaborative dialog among teachers;
- (6) providing a means of examining changes in teachers' actual use of formative assessment strategies;
- (7) providing a timely means of examining the impact of these changes in their teaching practices on student learning.

We also came to understand that in order to develop buy-in and be perceived as authentic by public school personnel, we needed to be able to develop a means of collecting reasonably valid data that could be used to evaluate the impact of the effort and demonstrate a means of meeting accountability demands to a broader audience. Furthermore, to receive continued district-level support, meet demands for accountability, and motivate teachers' continued use of AFL practices, there is a need to examine how changes in practice impact student learning.

These observations led to the idea of offering a year-long course on formative assessment, embedded in the context of teachers' daily work, which would utilize student performance data to monitor changes in teachers' practice on student learning. The course was collaboratively designed, monitored, and modified by university and public school staff as it was taught. As we worked informally with teachers and staff the first year and as we taught the course, our understanding evolved leading to changes in how we worked with teachers.

For this study we didn't attempt to address the multiple aspects of effective professional development, but rather to incorporate as many of those elements as is reasonably possible given available resources into the year-long professional development course. We were interested in helping educators develop stronger backgrounds and understandings of formative assessment skills and practices that actually led to changes in teacher practices and how such a year-long course could be altered as an embedded professional development experience. Additionally, we were interested in knowing more about how such changes occur and how educators translate theory into practice and document how participation may have had an impact on teachers' understanding and use of formative assessment practices. Finally, we were interested in how data could be gathered to provide timely information that teachers could actually use formatively while also providing data on the impact of changes in teacher practices on student learning.

We maintain, as have others (Thompson and Wiliam 2007), that professional development efforts need to be situated in the work of real teachers in real schools. For this study the questions we addressed emerged from trying to link theory with practice rather than as a preplanned experimental intervention. Just as Black and Wiliam (1998) noted that ecological validity of studies on the use of formative assessment practices is important in understanding the applicability of the findings to real classrooms, it would also seem important that the approaches to professional

development also need to be embedded in teachers' daily work, rather than predetermined or imposed on teachers. As they note, the work of professional developers needs to be within the context of the teachers' real-world settings, and in order for research to do more than simply 'interrogate existing practices' it has to be ecologically linked (Black and Wiliam 1998). Given our interest in helping educators develop stronger backgrounds and understandings, how they translate theory into practice and how data could be gathered to provide timely information on the impact of changes in teacher practices on student learning, the importance of ecological validity is evident.

Furthermore, as the literature highlights (Erickson 2007), adopting formative assessment practices involves individual and cultural shift in terms of how educators and students think about the nature of knowledge and knowledge acquisition (ontological and epistemological). An external orientation is unlikely to be readily adopted by teachers whose beliefs systems and orientations have developed over time. These largely unexamined belief systems serve as what Erickson (2007) described as a threat to formative assessments' 'formativity' (189). With the understanding of the importance of working within the various ecologies teachers work in, this is a study of an intervention that evolved over time, that emerged from our work with teachers in their actual professional settings; first in identifying their understanding of formative assessment practices and their associated needs, then in terms of how to work with teachers to actually help them transition from understanding to use in practice, and then in terms of evolving research and data collection designs which could be situated in teachers' work.

The potential impact of formative assessment on student learning

Despite an ever-growing research base and numerous articles advocating the use of formative assessment (Andrade and Cizek 2010; Black and Wiliam 1998; Reeves 2007) or AFL practices in teachers' classrooms, actual adoption and use of such strategies lag considerably behind what is being promoted nationally and internationally. Formative assessment ideas and practices are based on multiple theories of learning incorporating behavioral, cognitive, and constructivist approaches, especially the effective use of feedback (Hattie 2007). Motivational theories on self-efficacy (Bandura 1982), self-regulated learning (Zimmerman 1990), and learning orientation (Stipek 2002) are also woven into the rationale used to support these practices. A meta-analysis of over 250 comparative studies examining the impact of formative assessment practices on student learning by Black and Wiliam (1998) showed an average effect size ranging from .4 to .7 in well-controlled studies. That article and later work have been highly influential in articulating key characteristics of formative assessment. Hattie's (2007) meta-analysis of over 500 meta-analytic studies of 180,000 individual studies examining multiple factors influencing student achievement shows an average effect size of .79 for the effective use of feedback. Of the more than 100 factors examined, the effective use of feedback, which he equates with formative assessment, was in the highest 5 to 10 effect sizes reported. The studies reviewed in both the Black and Wiliam (1998) and Hattie (2007) meta-analyses were across all grade levels from K-12 through university-level work, involved a broad range of subject matter areas, and were conducted in a number of different countries worldwide.

Despite all of this evidence and continuing research further supporting the effectiveness of formative assessment practices, these ideas are slow to be fully integrated into teachers' day-to-day classroom practices (Schneider and Randel 2010). In many respects this should come as no big surprise as the kind of shifts in teacher skills and knowledge as well as the organizational infrastructures needed to support these changes represent several substantive changes in educators' routines. Several reasons for this lack of progress might be offered. Teachers may lack basic assessment skills, have only a partial or incoherent understanding of why or how to use assessment formatively, lack the time and support needed to integrate these practices into their teaching over a sustained period of time, or have too little opportunity for active and collaborative learning (Schneider and Randel 2010). Prior research on teachers' response to reform suggests teachers are more likely to gravitate toward approaches that are congruent with prior practices and to focus on surface manifestations rather than deeper pedagogical principles (Coburn 2004; Spillane 2000). The what, how, and why needed to move this agenda ahead involves knowledge of the content of AFL, an approach to professional development that supports teachers' adoption and use of such ideas, and teachers' making sense of how to enact AFL ideas in their classrooms (Thompson and Wiliam 2007). Without a deeper understanding of formative assessment teachers may fail to recognize that using assessment formatively represents a major change in the teacher's role in student learning and a fundamental reorientation of the teacher–student learning relationship on the part of both teachers and students. The opportunity for extensive professional discourse infused with new ideas, the need to personally reflect on and assimilate what is learned into one's thinking, augmented by the kind of professional development (PD) that is embedded in practice and the day-to-day work of teachers, linked to actual changes in student learning, over time, are key to accomplishing such a reorientation and teacher use of formative assessment practices (Black et al. 2003; Thompson and Wiliam 2007).

Characteristics of effective professional teacher learning

An ever-growing body of research on promising practices for effective PD has emerged over the past 30 years with a considerable impact on the field (Gersten et al. 2010) and a growing consensus of what characterizes effective professional learning (Wei et al. 2009). Elements of this emerging model cover a wide range of diverse topics from district- and school-level leadership and support structures; to enhanced teacher curriculum, content knowledge, and pedagogical content knowledge; to teachers' knowledge and use of research-supported best practices; to collaborative work between teachers and staff; and especially teachers' frequent use of assessment data to adopt and adjust instruction. Across these topics the importance of a sustained, collaborative, work-embedded context for professional learning situated within the needs of the school is stressed especially over external events such as one-day workshops or formally presented interventions (Garet et al. 2001; Gersten et al. 2010; Huberman and Miles 1984). Professional development is most effective when teachers engage actively in instructional inquiry in the context of collaborative professional communities, focused on instructional improvement and student achievement (Wei et al. 2009). However, that success is still tempered by teachers being able to make connections between assessment theory and research and how to enact those ideas in concrete and specific ways in their classrooms. There is a continuing need to explore how various approaches to developing this

understanding facilitate teachers' making sense of AFL practices and how their evolving understanding impacts their decisions to adopt and continue using such practices in practical ways in their classrooms.

Ongoing and available relationships between universities and local school divisions coupled with the requirement that teachers accrue continuing education credits offer one such alternative. The Consortium for Policy Research in Education (CPRE) (CPRE 1996; Parise and Spillane 2010) for example has pointed out that in order to meet the growing challenges of developing stronger curricula and transforming professional development, there is a need for substantial and long-term resource commitments as well as working beyond the usual channels and broadening the roles that stakeholders might play. Developing stronger collaborative working relationships between teachers and administrators, among teachers themselves as well as with university faculty and other educators, is a critical element of re-culturing schools to become places that stimulate and support teachers to make important changes themselves.

Accountability and its impact on teacher professional development

Getting K–12 public schools in the United States to commit to or even allow teachers to participate in the kind of sustained professional development recommended above is exacerbated by heightened accountability demands, the pressures of high-stakes testing, and No Child Left Behind. But this is not a unique condition in the United States as countries all over the world implement systematic accountability systems (Anderson 2005). Heightened attention to evaluative uses of assessment contributes to the tension between summative and formative uses of assessment, stifling the latter (Black and Wiliam 1998). Teachers' adoption and students' responses to formative assessment ideas and practices are influenced by their prior assessment experiences, epistemological beliefs about the nature of learning, perspectives on motivation and cognitive learning, and the role played by effort and ability. The influence of these factors as well as others on teachers' understanding and use of assessment formatively has been studied in multiple international contexts such as England, France, Canada, Greece, Belgium, New Zealand, Australia, Scotland, Japan, Portugal, and Norway (Black and Wiliam 1998) to name a few. Teacher development and use of assessment formatively is based on fundamental cognitive and motivational processes that are applicable across cultures.

These demands limit the opportunity for public school districts and universities to continue to work together unless that work can be traced to measured changes in student performance, especially on high-stakes tests. School districts simply are not inclined to let universities experiment with new ideas or approaches to learning or work with teachers in professional development efforts that cannot be shown to raise students' scores on measures of student learning. In addition, there is a growing press by states and the federal government for universities and particularly colleges of education to demonstrate that they can have a positive impact on K–12 education, especially teachers' use of effective practices that enhance student learning. This situation also compounds the problem of providing teachers with readily available information on student learning they can make sense of that at the same time would provide accountability evidence on enhanced student learning to a broader audience beyond the individual teacher's classroom.

While evolving models of professional development are sensible and compelling, there is still mixed or very little empirical evidence on the critical role these promising practices play in enhancing teacher learning of effective instructional strategies (Borko 2004; Garet et al. 2001) and more importantly student learning (Wayne et al. 2008; Yoon et al. 2007). Rarely do studies attempt to link PD to student outcomes, and the few that have attempted to do so have often yielded disappointing results. For example, Garet et al. (2008) found that large-scale PD that included many of these recommended practices did lead to significant increases in teachers' knowledge and observed teaching practices but failed to enhance student reading achievement significantly. Others have noted that seldom have recommended PD practices been widely field-tested and evaluated using rigorous research techniques (Desimone et al. 2002; Guskey 2000), arguing for the latter. In the Yoon et al. (2007) report conducted for the US Department of Education Institute of Education Sciences cited above, only nine studies out of 1300 they reviewed both meet the criteria of using rigorous research designs and at the same time are able to demonstrate a positive effect on measures of student achievement.

However, if the PD offered needs to be embedded in teachers' daily work it also would seem advisable to gather evidence that teachers can make sense of within the context of their daily work. Rigorous comparative research designs may fail to provide timely information, be cumbersome or expensive to administer, or be outside of most teachers' understanding and thus not provide them with the insight they need to make connections between changes in their practices and the impact these changes are having on student learning. A timely approach to gathering data using available resources is needed that makes sense to teachers and that at the same time can provide reasonably valid information on student learning. Such an approach embodies the basic assumptions of formative assessment. In other words, the design needs to fit into the intervention while adhering to as many canons of rigorous research design as are reasonably possible (Thompson and William 2007). As Little (1993) pointed out, many school improvement efforts fail to embody the basic assumptions of the improvement effort, and as such don't accurately model the skills, practices, and habits of mind needed to use the new skills effectively.

The approach to PD taken in this study evolved from working with teachers and school district assessment staff over a two-year period of time. The first-year university staff members met with teachers and central office staff at one school during teacher grade-level and planning meetings to generate interest and to develop a basic understanding of formative assessment theory and practices. These early meetings were essentially unproductive. It became readily apparent that teachers' lack of understanding and the lack of an intensive focus limited how much progress could be made. This led to the idea of offering a year-long course on formative assessment to develop a more in-depth understanding of the theory behind AFL practices and at the same time focus on teachers' actual use of various AFL strategies in their classroom instruction. In the context of this study that meant the course should try to fit into the perceived needs of the participating schools or the schools' improvement plans, avoid being perceived as an add-on to teachers' workload, consider teachers' needs and background knowledge in the area of formative assessment, be offered over an extended period of time with a focus on strategies teachers could use in their daily work with students, stress collaborative dialog and work between teachers, provide a means of examining changes in teachers' actual use of formative assessment strategies, and provide teachers with a timely means of examining the impact

of changes in their teaching practices on student learning. To meet accountability demands there was also the need to collect reasonably valid data that could be used to evaluate the impact of the course for a broader audience.

Data and methods

Working informally with teachers and more formally through an advisory committee composed of university-level staff and with district- and school-level administrators and staff over a two-year period of time provided information on how university courses which typically focus on content coverage needed to be altered to be more effective as a professional development experience for practicing teachers. Twenty-one school district staff participated in the course and in reviewing and teaching the year-long course. In order to address the second and third questions teachers were required to attend course presentations and to actually try out and document the use of various formative assessment practices in addition to responding to self-report surveys on use and intended use of the practices in the future. An embedded means of collecting and analyzing comparative data on student learning utilizing district benchmark quarterly tests provided information for the third question. The design and adaptations to the course were explored qualitatively through observations and discussions with teachers and school staff. Changes in teachers' perceptions and use of strategies on the self-report survey were analyzed quantitatively. Documented use of various strategies and the impact these changes had on student learning were examined quantitatively by comparing student performances on district quarterly assessments when teachers used or did not use an AFL strategy on similar learning targets.

Initial analysis

The first year of working informally with teachers and staff at one of the participating schools prior to actually offering the PD course was instrumental in helping us to recognize teacher needs and in developing a means of assessing the impact of teachers' changes in practice and its impact on student learning. Beginning in the 2005–06 school year, two faculty members from a Mid-Atlantic metropolitan university worked with third- through fifth-grade teachers during professional development sessions, grade-level meetings, and data team planning meetings to facilitate teachers' use of these practices. These grades were chosen by the school principal as they are the earliest grades for which teachers are held accountable for students' performance on the state's high-stakes tests. Initially the faculty members briefly presented ideas at school-level meetings and participated in data team meetings, offering suggestions on how teachers could use formative assessment practices to address problems and issues raised as teachers reviewed data on benchmark quarterly assessments.

The results of these initial forays into working with teachers with the intent of embedding professional development into their ongoing work responsibilities were discouraging. The action plans teachers developed from these data meetings did not contain vocabulary, let alone strategies, associated with formative assessment, and teachers' discussions were devoid of any underlying conceptual understanding of why formative assessment might have merit. It was clear that a more in-depth and structured approach was needed in order to build teachers' basic assessment skills

and their conceptual understanding of formative assessment and how such ideas could be used in the classroom. Working with district- and school-level staff led to recognition of the need for such a course, initial design of the course, and identification of topics to be covered in the course. Having a need to link changes in teacher behaviors and to help teachers make connections between changes in their teaching practices in formative assessment with changes in student learning presented another challenge.

During this time period an opportunity arose to collaborate on two teacher action research studies. As teachers sought to try out some of the new formative assessment practices and sought to examine the effects of such practices on student learning, the need to collect and examine some form of comparative data arose. In order to meet this need, the idea arose of having teachers compare how their students were doing on quarterly benchmark tests on specific learning targets when they used a particular formative assessment practice(s) and how their students did on parallel learning targets when they were not using the practice. In order to control for other sources of potential invalidity, performance in each area would be compared with district averages on those same learning targets. Teachers for example could have picked more or less difficult parallel learning targets on which they did or did not use a strategy, and comparing the difference in performance with district averages would control for such differential selection. How the same students performed on two similar learning targets would provide a within-subject comparison and using district data would provide a between-subjects or normative means of comparing the relative performance of students. The formative assessment practices used by the two teachers involved having students ‘unwrap’ the learning target, students working with the teacher to generate criteria or ‘rubrics’ students would use to critique their work, and students’ use of the rubric to monitor and assess the quality of their work. District tests use a multiple-choice format and are based on the State of Virginia Standards of Learning (SOL) and use a format typical of the high-stakes tests used in most states. The tests focus on recall of declarative knowledge, reasoning and problem-solving skills, and the use and interpretation of information or data presented in the form of pictures, charts, graphs, reading passages, and samples of writing. Teams of teachers and central office curriculum staff develop these tests to be aligned with the SOL and the format used on that test. The tests are then reviewed and revised by central office assessment staff who are responsible for generating the tests, and scoring and reporting on student performance. Table 1 presents information on the students’ performance on the district quarterly test with parallel learning targets when teachers did or did not use the above strategies.

Table 1. Student performance on quarterly assessment: percent of students mastering SOL ($N = 22$).

Source	Non-targeted SOL Va. Sci. 5.4	Targeted SOL Va. Sci. 5.5
District-level performance	55%	56%
Teacher A: inclusion class performance (special ed. students)	44%	80%
Teacher B: gifted/regular class performance	36%	65%
	69%	83%

In the teacher action research studies conducted, Teacher A, whose class was composed of 9 special needs students and 11 regular education students, had an overall increase of 35 percentage points in students' mastery of the SOL being assessed (students scored 11% below the district average on the non-targeted SOL and 24% above the district average on the targeted SOL) and a 28 percentage point difference in mastery for the special education students in her inclusion class. Teacher B, who taught a class of 22 gifted and regular education students, had a 13 percentage point difference in mastery (students scored 14% above the district average on the non-targeted SOL and 27% above the district average on the targeted SOL). Also of note is the overall percentage of students mastering the targeted SOL, where 80% of the inclusion class/regular education class and 83% of the gifted/regular education class reached a mastery level on the SOL, suggesting the potential for formative assessment practices to help close the gap between lower and higher performing students.

These results parallel those frequently found in the formative assessment literature (Black and Wiliam 1998) that shows these practices have a differential effect on students who typically do not perform as well. The design provided teachers with a way of examining the differential impact of the new practice by comparing how the same student performs under two instructional approaches while at the same time comparing those differences in learning with a normative standard of district-level performance. Thus teacher and student differences are held constant and their relative performance can be interpreted as meeting or differing from district standards. Teachers could also explore the differential impact the practice had on students with different entry-level performances. Surprising as it might seem, we have not seen any other examples of evaluation designs that utilize well-established and reasonably rigorous research designs that could be embedded in teachers' ongoing action research studies and that at the same time make sense to teachers as well as meet accountability demands. The result then from our initial work with teachers informally was to identify the need to work with teachers more intensely over an extended period of time to develop a more in-depth understanding of assessment for learning ideas and practices and the development of a reasonably valid action research design that could be used to link changes in practice with changes in student learning.

Developing and teaching the year-long professional development graduate course

As an outgrowth of these experiences, the first year of a professional development course was jointly planned, monitored, and presented the following school year. What made this course unusual was that it was designed specifically as a professional development course that would be tied to and embedded in the daily work of teachers and it would purposely involve teachers in linking changes in their teaching practices to changes in student learning. The initial plan was to focus the content of the course on knowledge and skills typically taught in a classroom assessment course while addressing the needs we had identified while working informally with teachers, those identified in each school's 'School Accountability Plan', and the needs identified from students' performances on the school district's quarterly benchmark tests.

University faculty and district staff would then provide ongoing support for teachers as they planned to integrate new assessment practices into various upcoming units of instruction following each of 15 three-hour sessions of the course offered over an eight-month period of time from October to late May. Teachers would then use these units of instruction in the next one to three weeks in their classrooms, document which practices were used on which SOL and for how long, identify similar non-targeted SOL for comparison, informally reflect on the impact of these units, and share these reflections with their colleagues at grade-level meetings or during course meetings. Each class session involved lectures on key aspects of formative assessment, structured group discussions, group work sessions on integrating formative assessment strategies into their future lesson plans, whole and small group discussions on previous uses of strategies, ‘fish bowl’ demonstrations of various strategies, gallery walks reviewing class work, and guest lecturers from division and university personnel. As data from the district’s quarterly assessments became available, teachers were required to link the formative assessment practices they had utilized with student performance data. Reflecting on and trying out the new ideas and practices presented were the only requirements in the course and there were no course exams. Course and district staff met eight times as an advisory committee and the university staff and district assessment staff met over 80 times with teachers at grade-level team meetings or for special presentations between course sessions over an eight-month period.

The course covered major aspects of assessment and especially formative assessment identified by the Assessment Training Institute to guide our work with teachers (Stiggins et al. 2006). To enhance teachers’ understanding, develop their basic assessment skills, and provide practical strategies for teachers to use related to the above, the course specifically focused on:

- (1) a balanced approach to formative and summative assessment;
- (2) use of formative assessment ideas/strategies;
- (3) use of assessment to motivate;
- (4) use of effective feedback strategies;
- (5) unwrapping learning targets with students;
- (6) unwrapping reasoning-level learning targets with students;
- (7) choosing appropriate means of assessment;
- (8) effectively leading reasoning-level discussions;
- (9) use of goal setting with students;
- (10) use of data for instruction;
- (11) use of data to meet individual student needs;
- (12) use of assessment for students’ self-monitoring;
- (13) appropriately developing selection-type items;
- (14) use of peer questioning/discourse;
- (15) conducting effective discussions;
- (16) use of exemplars and rubrics;
- (17) student participation in developing rubrics.

Participants for the course were recruited from four elementary schools with the support of their respective principals. In total, 46 participants were recruited, 21 were district-level subject area specialists or school administrators, 23 were classroom teachers, and two held joint positions. Twenty-five teachers participated in the

course from four different schools, 10 and 9 teachers from two schools including the one we worked with the first year, and five teachers and one teacher from the other two schools. The schools were inner city schools with 62 to 95% of the students on free and reduced lunches and 75 to 99% from a minority background. Teachers taught in grades 1 through 5 and varied in terms of basic assessment skills, past experience with formative assessment practices, and the degree to which such practices were a priority of professional development in that school. Given these considerations, teachers varied widely in terms of their readiness for developing, using, and reflecting on formative assessment practices. Furthermore, two major areas of concern presented themselves early in the course. While drop and run professional development, whether it be offered by university staff or school district staff, has been heavily criticized in the literature and we sought to avoid such an approach, we still under-estimated the need for teachers to have the time and repeated opportunities to understand and learn to use the new practices being advocated.

As the course progressed it became apparent that following a topical outline typical of university-level courses on assessment, even if it involved considerable discussion and stressed hands-on practice, would not work. Teachers needed more time to try out, reflect on, and revise their respective practices and sought to revisit the same topic as many as five or six times. Crafting ideas and strategies into practice would take time and several iterations. Given the differential needs of each teacher, our need to cover certain essential assessment practices, and varying levels of principal support, it became apparent that it would be impractical to link the use of these practices to critical areas of need as they arose in each school. Hence teachers' past experiences and the context in which they worked influenced their development and use of formative assessment practices.

Data collection

In addition to the information gathered informally as we worked with teachers, three other sources of information were collected for use in data analysis. One source of information was teachers' documented use of strategies on a specifically developed form. A second was teachers' responses to a retrospective self-report course survey given at the end of the course. Twenty-two of the twenty-five participating teachers responded to the survey. A third was district performance data which teachers incorporated into their documentation of use of strategies form and that were also summated across the studies teachers provided. District performance data or performance on specific SOL were only collected and analyzed for those SOL teachers had identified as a focus for using particular assessment strategies.

Teacher-documented use of strategies and student performance data

A requirement and major focus of the course was on teachers' use of the formative assessment strategies taught in the course. To document that use and prompt teacher reflection, teachers were required to submit a form documenting what strategies they used on which SOL, identify a parallel SOL they did not use the strategy on, note how long they taught the SOL and used the strategy, reflect on students' reaction to use of the strategy and any challenges presented in using the strategy, and as data from district tests became available enter these on the form. The primary means of assessing the impact of the professional development course on student learning was

the percentage of students mastering the SOL learning targets as assessed by the district’s quarterly benchmark tests. Each participant was expected to utilize various assessment strategies taught in the professional development course in their classrooms between six and eight time over the course of the year and collect data on each of these attempts. In this way there could be between 300 to 400 sub-studies or what might be considered teacher action research studies. Teachers varied considerably in terms of how much they tried out different practices and wrote these up formally with some teachers only submitting one or two write-ups while a few teachers submitted as many as six or seven. Other difficulties with matching the learning targets with district data eventually led to 46 comparable studies instead of the anticipated 300 or more studies.

End-of-course retrospective survey

In order to examine teachers’ knowledge and use of formative assessment practices, we developed a retrospective survey to examine these questions (see Goedhart and Hoogstraten 1992; Lamb and Tschillard 2005) employing a 5-point numerical rating format. Items were anchored by descriptive statements: a rating of 1 indicated little or no use or understanding; a rating of 5 indicating extensive use or understanding. A sample item is provided below to help clarify the format of the survey items used for each of the areas focused on in the course and the comparative information it provides that is the focus of the data analysis later in this paper. Items asked about either an idea or about the use of a strategy and asked about their understanding, as well as their previous, current, and planned use of the idea or strategy. Other items on the survey also asked about collaboration, the perceived impact on student motivation and learning, and environmental factors that may have enhanced or hindered participation and utilization of the course (see Figure 1).

The reason for using a retrospective survey rather than a pre–post test measure warrants some discussion. It is often the case that researchers assume that participants’ perspectives are invariant across pre- and post-measures; however, because formative assessment is a topic that most educators have limited backgrounds in, a pre-measure of their understanding and use of these strategies risks a misrepresentation of their knowledge. For example, Black and Wiliam (1998) as well as Hattie (2007) noted that studies have shown that when teachers use feedback, nearly 40% of the time it has a negative impact on student learning. When teachers are asked if they provide effective feedback they are likely to respond in the affirmative only to revise that response later after they gain further knowledge, causing them to reassess their earlier responses. Thus participants may over-estimate their knowledge and skills prior to training (Bray, Maxwell, and Howard 1984). In this way, the conventional pre–post test may over- or under-estimate knowledge gains by creating a

A. I do not understand this idea/strategy	1	2	3	4	5	I understand and am well prepared to use this idea/strategy
B. Previously, I did not use this idea/strategy						Previously, I used this idea/strategy extensively
C. Currently, I do not use this idea/strategy						Currently, I use this idea/strategy extensively
D. I do not find this idea/strategy useful and do not plan on using it						I find this idea/strategy very useful and plan to continue using it

Figure 1. Item 5. ‘Unwrapping’ learning targets in ‘student friendly’ language.

response shift bias (Howard and Dailey 1979; Howard et al. 1979; Pohl 1982; Rockwell and Kohn 1989; Rohs and Langone 1997; Sprangers and Hoogstraten 1988). We sought to address this problem by asking teachers to respond comparatively within the survey itself after they had had training and developed a better understanding of what was being asked. Because respondents are rating their experiences at the same time from the same perspective, they should be less likely to have a response-shift bias (Rohs 1999). If respondents are likely to change their perspective or interpretation of the meaning of an item, then the retrospective survey may produce more valid results (Pratt, McGuigan, and Katzev 2000) than the more traditionally accepted practice of using pre–post measures of change.

Results

Impact on teachers' understanding and practices: end-of-course retrospective survey

The retrospective self-report survey given at the end of the course and completed by 22 of 25 participating classroom teachers was used to measure, among other constructs, the teachers' previous and current use of the 17 formative assessment practices, as well as their understanding of and intention to use those strategies. The data were analyzed for mean differences across each of the 17 ideas and practices surveyed and a summated total score was obtained for use, plan to use, and understanding by summing each teacher's response across the 17 items. In addition to these mean comparisons, Pearson correlations were computed between these summated total scores on use, plans to use the ideas/practice in the future, and teachers' understanding of strategies with other selected variables measured on the self-report survey. Cronbach's alpha for the individual scales ranged from .71 to .80 with the most falling near .80. Cronbach's alpha reliabilities for the total summated scales were .894 for current use of AFL strategies, .864 for understanding, and .892 for planning to use the strategies in the future. Also of interest were four items on the survey that seem to capture teachers' perception of student learning which focused on: student participation, student ownership, increased student motivation, and student orientation to learning. The correlations between teachers' ratings on these four items ranged from the mid-70s to the low 80s.

Across all 17 areas of formative assessment focused on in the course, teachers indicated an increase in their understanding or use of the strategy, with one exception all of these differences were significant on a paired *t*-test and produced effect sizes that ranged from .33 to 2.03 (see Table 2 below; negative *t*-values indicate an increase in use or understanding). Twelve of these comparisons were statistically significant at the .001 level, three at the .01 level, and one at the .05 level. One item, use of peer questioning and discourse, approached significance: $t(2.08)$, $n = 22$, $p = .05$. Of particular interest here are those items that teachers reported the largest changes between previous and current use; these include the use of assessment for student self-monitoring, the use of formative assessment ideas/strategies, and unwrapping learning objectives with students. Because the use of formative assessment practices represents a major change or reorientation to the teacher–student learning relationship, the large effect size on these particular ideas/strategies suggests movement in that direction where teachers interactively involved students more. The lowest mean differences in use of data, peer questioning and discourse, and development and use of rubrics likely represent small shifts in these areas as teachers found

Table 2. Changes in teachers' previous and current use of formative assessment ideas/strategies ($N = 22$).

	Mean difference	T	Sig. (2-tail)	Effect size
Balanced approach to formative and summative assessment	1.18	-4.40	.000	1.28
Use of formative assessment ideas/strategies	1.32	-7.37	.000	1.76
Use of assessment to motivate	1.09	-3.72	.001	1.00
Using effective feedback strategies	1.09	-4.62	.000	1.16
Unwrapping learning objectives with students	1.77	-6.55	.000	1.59
Unwrapping reasoning-level learning targets with students	1.14	-4.15	.000	0.95
Choosing appropriate means of assessment	1.18	-4.86	.000	1.14
Effectively leading reasoning-level discussions	1.09	-5.90	.000	1.20
Use of goal setting with students	1.09	-5.27	.000	0.92
Use of data for instruction	.64	-3.31	.003	0.64
Use of data to meet individual student needs	.50	-2.32	.031	0.64
Use of assessment for students' self-monitoring	1.68	-7.58	.000	2.03
Appropriately developed selection-type items	1.09	-4.16	.000	0.99
Use of peer questioning/discourse	.31	-2.08	.050	0.33
Conducting effective discussions	.55	-3.46	.002	0.60
Use of exemplars and rubrics	.64	-2.98	.007	0.64
Student participation in developing rubrics	1.05	-3.80	.001	1.06

these practices difficult to implement (the items had the lowest ratings in current use) or in the case of use of data the district had stressed the instructional use of data for some time (this area had the highest initial mean score on previous use).

Relationships between understanding, use, and selected variables

Table 3 presents correlations between several variables of interest, especially the summated total scores on teachers' current use of AFL strategies summed across the items, teachers' ratings of their understanding summed across items, and teachers' intention or plan to use the practices in the future summed across items. All three of these summated scores were highly correlated with each other, teachers' current use of strategies correlated with their understanding ($r = .822$) and with their plan to use the strategies in the future ($r = .847$). Teachers' plan to use the strategies in the future correlated ($r = .884$) with their ratings of understanding. These high and statistically significant correlations ($p < .01$) clearly indicate that teachers' understanding was related to their use and planned use of the strategies in the future. Of note here is that these are correlations, hence not only were high ratings in each related to high ratings in the other areas, but the obverse would also be true: teachers who responded with a low rating in one area also responded with low ratings in the other areas. Thus a lack of understanding was related to low current and future use of the ideas and practices. The high degree of correlation also indicates considerable variability in teachers' understanding and use of the practices, suggesting the course was not uniformly effective.

Current use of strategies was also related to perceptions of students' ownership of their learning ($r = .434$) and a change in students' orientation to learning ($r = .506$). Teachers' understanding was related to increased student participation

Table 3. Pearson correlation matrix ($N = 22$).

	Total Current Use of AFL Strategies	Total Understanding	Total Plan to Use AFL Strategies	Students Participate More As a Result of Using AFL Practices	Students Have Greater Ownership of Their Learning as a Result	Students Are More Motivated As a Result of AFL	Students' Orientation to Learning Has Changed as a Result	Years Teaching
Total Current Use of AFL Strategies	1	.822**	.847**	.403	.434*	.320	.506*	.320
Total Understanding		1	.884**	.443*	.327	.353	.403	.256
Total Plan to Use AFL Strategies			1	.476*	.437*	.408	.448*	.373
Students Participate More As a Result of Using AFL Practices	.403	.443*	.476*	1	.742**	.891**	.795**	.353
Students Have Greater Ownership of Their Learning as a Result		.327	.437*	.742**	1	.773**	.816**	.110
Students Are More Motivated As a Result of AFL	.320	.353	.408	.891**	.773**	1	.764**	.356
Students' Orientation to Learning Has Changed as a Result	.506*	.403	.448*	.795**	.816**	.764**	1	.096
Years Teaching	.320	.256	.373	.353	.110	.356	.096	1

**Correlation is significant at the 0.01 level (2-tailed).
*Correlation is significant at the 0.05 level (2-tailed).

($r = .443$) and teachers' plan to use the strategies in the future was related to increased student participation ($r = .476$), ownership ($r = .437$), and orientation to learning ($r = .448$). All of these correlations were statistically significant at the .05 level. Given the high degree of relationship between these items as reflective of teachers' perceptions of factors related to student learning, these correlations taken together suggest teachers clearly thought their understanding and use of AFL had a positive impact on student learning.

Table 4 presents correlations between prescribed and flexible uses of formative assessment strategies and teacher current use, plan to use, and understanding of these strategies. The survey item that asked if teacher felt the need to apply formative assessment practice in a 'prescribed manner' was unrelated ($r = .10$ to $.20$) to any current use, planned use, or teachers' understanding. Contrary to what might be expected using the AFL practices in a prescribed manner was not related to the items asking if teachers used the AFL ideas and strategies flexibly as a framework. On the other hand, using formative assessment ideas/practices flexibly as a framework was significantly related to current ($r = .699$) and planned use ($r = .806$) and teachers' understanding ($r = .648$). The latter finding by contrast shows the impact of the course on engaging teachers in formative assessment practices as a framework for practice rather than as a 'bag' of teaching activities. The results clearly indicate a significant increase in teachers' use of the strategies exceeding what normally might be expected. Furthermore, the relationship between current, intended use, and understanding highlights and reinforces our observations and that of other researchers in this area (Black et al. 2003; Thompson and Wiliam 2007) that teachers' understanding is a key factor in their adoption of new practices.

Also of interest was the finding that a teacher's years of experience was unrelated to any of the variables analyzed, indicating that experience was not likely a factor in teachers' adoption or understanding of formative assessment practices. However, the negative and statistically significant negative correlations between teachers' ratings of the impact of other professional development activities on current ($-.472$) and planned use ($-.426$) and their understanding ($-.475$) do suggest other demands on their time limited how much they learned or could participate in using the AFL practices. This is hardly a surprise as time and competing demands are frequently noted as limiting factors in teacher participation in PD activities (Wei et al. 2009).

Impact on student performance

In analyzing student performance data we wanted to assess whether the changes in teachers' understanding and current and intended use resulted in changes in student performance data. Assumptions regarding teacher awareness of project goals resulted in some teachers selecting standards for treatment or comparison that were not tested in a given quarter, or in which comparative standards were considered dissimilar. As a result, the filtering process identified only 58 studies where teachers had used a particular AFL strategy and 46 similar learning targets where teachers had not used an AFL strategy. Descriptive statistics (Table 5) highlight differences between SOL objectives on which teachers used formative assessment strategies and corresponding division averages on these learning targets. On those objectives where teachers utilized formative assessment strategies, students tested 73% of items correct while the division average was 66%, a 7.18% difference (see Table 5). On objectives the teachers had identified as a parallel learning target on which they did not use an

Table 4. Pearson correlation matrix ($N = 22$).

	Total Current Use of AFL Strategies	Total Under- standing	Total Plan to Use AFL Strategies	Other Professional Development Limited My Ability to Use	Felt Need to Enact Practices in a Pre- scribed Way	Have Used AFL Ideas/ Strategies Flexibly as a Framework
Total Current Use of AFL Strategies	1	.822**	.847**	-.472*	.273	.699**
Total Understanding	.822**	1	.884**	-.475*	.219	.648**
Total Plan to Use AFL Strategies	.847**	.884**	1	-.426*	.147	.806**
Other Professional Development Limited	-.472*	-.475*	-.426*	1	.086	-.348
My Ability to Use Practices in a Prescribed Way	.273	.219	.147	.086	1	.157
Have Used AFL Ideas/ Strategies Flexibly as a Framework	.699**	.648**	.806**	-.348	.157	1

**Correlation is significant at the 0.01 level (2-tailed).

*Correlation is significant at the 0.05 level (2-tailed).

Table 5. Percentage of students' mastery for teacher and division.

	<i>N</i>	Mean	SD
Teacher targeted SOL objective	58	73.312	16.4840
District targeted SOL objective	58	66.128	10.5462
Teacher non-targeted SOL objective	46	71.00	14.821
District non-targeted SOL objective	46	68.37	8.161

AFL strategy, the teachers averaged 71% correct compared with a district average of 68% correct, a difference of 2.63%.

To assess these observed differences between the performance on targeted SOL objectives and division-level averages on matched items, a pared samples *t*-test was run (Table 6) resulting in 46 directly comparable items (12 targeted SOL were duplicative or could not be matched). In each case the difference scores between the teacher's average and the district's average were used for targeted and non-targeted SOL learning targets.

There was a significant difference, $t = 3.63$, $DF = 45$, $p = .001$, in the mean difference scores for teacher-targeted SOL objectives ($M = 7.45$, $SD = 12.64$), and division-level responses to the same SOL objectives compared with a mean difference of 2.63, $SD = 10.76$ on learning targets on which the teachers had not used an AFL strategy. In order to quantify the degree of difference for these mean values, an effect size was calculated for teachers who utilized a formative assessment strategy compared with the district averages, resulting in a small to moderate effect (Cohen's $d = .41$).

Limitations

Because there were a limited number of action research studies submitted by teachers in general and considerable variability in the number of studies submitted by each teacher, we could not link teachers' self-reported use of strategies on the survey in any direct or meaningful way with changes in student learning. While the correlational analysis does indicate teachers made a connection between enhanced student learning and their use and understanding of AFL practices, the data are correlational and based on teacher perceptions. Had we anticipated the problem of having teachers uniformly conduct more studies, we could have provided more scaffolded experiences in the course to foster teachers' responses to this course requirement. Furthermore, the small and non-representative number of action research studies submitted and analyzed limits the generalizability of results from the study, suggesting more about its potential rather than actual impact.

Variations between teachers' responses to the survey resulting in relatively high-scale reliabilities and correlations between variables suggest that the PD was not uniformly effective. From informal discussions with teachers we knew that their school principals were not always supportive of their participation and use of the

Table 6. Paired samples statistics: mean difference between teacher and district-level performance ($n = 46$).

	Mean difference	SD
Targeted SOL objective	7.4522	12.64195
Non-targeted SOL objective	2.6283	10.75984

strategies and that often teachers felt pressures from competing demands on their time. While we sought to address this by linking the PD to school-level improvement plans and continued dialog with school principals, this was not as feasible or as productive as we had hoped. We also found that while we stressed teachers working as learning teams in the course this did not always carry over to their daily work at their schools, again with considerable variability. Much work is still needed on structuring such courses to fit with and be supported by school administrators, fostering teachers working as learning teams in the course as well as at their schools, and trying out new practices and collecting and analyzing data. However, as a model for integrating new knowledge, using this learning to change the way teachers work with students, and linking these changes to changes in student learning, we believe the model still has merit and data to support its use.

Conclusions and implications

Retrospective survey data highlighted that teachers' participation in the professional development course did strengthen their understanding of both general knowledge of formative assessment and the use of formative assessment practices. Teachers' plans to use the strategies in the future were related to their understanding of these strategies, supporting the contention that teachers need an in-depth and comprehensive understanding of formative assessment practices in order to make specific and concrete applications of such practices in their classrooms (Wei et al. 2009). Teachers' ratings on the degree to which they 'used formative assessment ideas/practices flexibly as a framework' were significantly related to use and understanding, with correlations ranging from .699 to .806. These findings show the impact of engaging teachers in formative assessment practices as a framework for practice rather than as a prescribed set of practices or 'bag' of teaching activities. University-level courses that attempt to blend theory and practice into the daily work of teachers are well suited as an alternative means of developing such a framework.

However, such adaptations are challenging. Teachers' needs are not always easily identified, may not be congruent with school improvement plans, may be inconsistently supported by school staff, or be in competition with other school professional development projects. University-level courses need to be developed and monitored collaboratively, cognizant of teachers' iterative needs for multiple opportunities to craft ideas into practice. Inquiry approaches involving data collection, interpretation, and use need to be continually supported, meaningfully linked to teachers' practices and integrated with district-level data collection and reporting.

With respect to the impact on student learning, an overall moderate effect size of .41 was found for teachers when they utilized a formative assessment strategy compared with when the same teacher did not use an AFL strategy on similar learning targets. These data suggest a positive impact of formative assessment training on student learning. However, at this point in time the limited number of teacher action research studies that can be linked to self-reported changes in teacher assessment practices and to measured student learning are meager and indirect. Variation in teacher participation and documentation of the impact of changes in practice linked to changes in student learning limit the degree to which the results can be generalized; however, these data do suggest that the approach has merit. It is worthy of being strengthened and further developed into revised programs likely to garner more consistent teacher participation and collection of usable data that can more fully support

and detail outcomes from changes in practice. As school divisions have increasingly come under a host of accountability pressures, partnerships with universities have become contingent on being able to link efforts to measured changes in student performance, especially on high-stakes tests. We think the approach used in teacher participation and collection of data helps to address these needs. Our approach breathes assessment and pedagogical life into the often dry and technical approach to data-based decision making, expands teachers' understanding of how to use assessment formatively to promote student learning, and puts data which are linked to changes in instructional practices in the hands of teachers, allowing them to reflect on and refine these practices.

There is a continuing need to understand the gap between theory and practice and factors that hinder participation, especially school-level participation. The correlations indicating a negative impact from the press of other professional development activities with teachers' use and understanding suggest less than an ideal situation when there are many competing demands on teachers' time. Fostering greater opportunities for educators to have sustained dialog across classrooms and grade levels about new instructional and assessment practices and to see the impact of such practices on student learning should broaden the impact of such reforms. Focusing on a more articulated and structured learning team format (Stiggins et al. 2006; Thomson and Wiliam 2007) embedded in the course and linked to school-wide staff development as well as data on student learning would likely lead to even more effective classroom and school-level changes.

Because formative assessment ideas and practices are slow to be fully integrated into teachers' day-to-day classroom practices, an iterative training process that is linked to student performance data provides a sound model for helping teachers make the major changes in their roles and support a fundamental reorientation to the teacher-student learning relationship on the part of both informed teachers and students. As has been pointed out (Thomson and Wiliam 2007), this reorientation cannot be accomplished through 'canned' professional development. Instead, training models that take a comprehensive, adaptive, integrated, developmental approach, providing opportunities for extensive professional discourse, and longer term support for embedding practices in the day-to-day work of teachers, are needed.

Notes on contributors

Jack Robinson, PhD, is an associate professor at Old Dominion University where he specializes in formative assessment and student learning and motivation. He has worked extensively in public school collaborations training teachers and school leaders on the implementation and use of formative assessment strategies.

Steve Myran, PhD, is an associate professor at Old Dominion University and specializes in instructional leadership, data-based decision-making, formative assessment and school-university partnerships.

Richard Strauss, PhD, is the retired senior coordinator of Assessment Programs in Norfolk City Public Schools in Norfolk Virginia.

William Reed, PhD, is an assessment coordinator for Norfolk City Public Schools. He received The American Association of Colleges for Teacher Education (AACTE) award for outstanding dissertation in 2009 for his study on "Teachers' Education Factors and Their Influence on Formative Assessment Processes".

References

- Anderson, J. A. 2005. *Accountability in Education*. Educational Policy Series No.1. International Academy of Education. Paris: International Institute for Educational Planning.
- Andrade, H., and G. J. Cizek, eds. 2010. *Handbook of Formative Assessment*. New York: Routledge.
- Bandura, A. 1982. "Self-Efficacy Mechanism in Human Agency." *American Psychologist* 7 (2): 122–147.
- Black, P., C. Harrison, C. Lee, B. Marshall, and D. Wiliam. 2003. "The Nature of Value of Formative Assessment for Learning." *Improving Cchools* 6: 7–22.
- Black, P., and D. Wiliam. 1998. "Assessment and Classroom Learning." *Assessment in Education: Principles, Policy and Practice* 5 (1): 7–73.
- Black, P., and D. Wiliam. 2004. "The Formative Purpose: Assessment Must First Promote Learning." *Yearbook of the National Society for the Study of Education* 103 (2): 20–50.
- Borko, H. 2004. "Professional Development and Teacher Learning: Mapping the Terrain." *Educational Researcher* 33 (8): 3–15.
- Bray, J. H., S. E. Maxwell, and G. S. Howard. 1984. "Methods of Analysis with Response-shift Bias." *Educational and Psychological Measurement* 44: 781–804.
- Coburn, C. E. 2004. "Beyond Decoupling: Rethinking the Relationship Between the Institutional Environment and the Classroom." *Sociology of Education* 77 (3): 211–244.
- Consortium for Policy Research in Education. 1996. *Policies and Programs for Professional Development of Teachers: Profiles of the States*. Philadelphia: Author.
- Desimone, L., A. Porter, M. Garet, K. Yoon, and B. Birman. 2002. "Effects of Professional Development on Teachers' Instruction: Results From a Three-Year Longitudinal Study." *Education Evaluation and Policy Analysis* 24 (2): 81–112.
- Erickson, F. 2007. "Chapter 8 Some Thoughts on 'Proximal' Formative Assessment of Student Learning." *Yearbook of the National Society for the Study of Education* 106 (1): 186–216.
- Garet, M., A. Porter, L. Desimone, B. Birman, and K. Yoon. 2001. "What Makes Professional Development Effective? Analysis of a National Sample of Teachers." *American Educational Research Journal* 38: 915–945.
- Garet, M. S., S. Cronen, M. Eaton, A. Kurki, M. Ludwig, W. Jones, K. Uekawa, et al. 2008. *The Impact of Two Professional Development Interventions on Early Reading Instruction and Achievement*. NCEE 2008-4030. National Center for Education Evaluation and Regional Assistance.
- Gersten, R., J. Dimino, M. Jayanthi, J. S. Kim, and L. Santoro. 2010. "Teacher Study Group: The Impact of Reading Instruction and Student Outcomes in First Grade Classrooms." *American Educational Research Journal* 47: 694–739.
- Goedhart, H., and J. Hoogstraten. 1992. "The Retrospective Pretest and the Role of Pretest Information in Evaluative Studies." *Psychological Reports* 70 (3): 699–704.
- Guskey, T. G. 2000. *Evaluating Professional Development*. Thousand Oaks, CA: Corwin Press.
- Hattie, J. 2007. "Developing Potentials for Learning: Evidence, Assessment, and Progress." In *EARLI Biennial Conference*, Budapest, Hungary. Accessed March 11. <http://www.education.auckland.ac.nz/uoa/education/staff/j.hattie/presentations.cfm>
- Howard, G. S., and P. R. Dailey. 1979. "Response-Shift Bias: A Source of Contamination of Self-Report Measures." *Journal of Applied Psychology* 64: 144–150.
- Howard, G. S., K. M. Ralph, N. A. Gulanick, S. E. Maxwell, D. W. Nance, and S. K. Gerber. 1979. "Internal Invalidity in Pretest-Posttest Self-Report Evaluations and a Reevaluation of Retrospective Pretests." *Applied Psychological Measurement* 3: 1–23.
- Huberman, A. M., and M. B. Miles. 1984. *Innovation Up Close: How School Improvement Works*. New York: Plenum.
- Lamb, T. A., and R. Tschillard. 2005. "Evaluating Learning in Professional Development Workshops: Using the Retrospective Pretest." *Journal of Research in Professional Learning* 1: 1–9.
- Little, J. W. 1993. "Teachers' Professional Development in a Climate of Educational Reform." *Educational Evaluation and Policy Analysis* 15 (2): 129–151.
- Pohl, N. F. 1982. "Using Retrospective Pre-Ratings to Counteract Response-Shift Confounding." *Journal of Experimental Education* 50: 211–214.

- Parise, L. M., and J. P. Spillane. 2010. "Teacher Learning and Instructional Change: How Formal and on-the-Job Learning Opportunities Predict Changes in Elementary School Teachers' Instructional Practice." *Elementary School Journal* 110 (3): 323–346.
- Pratt, C. C., W. M. McGuigan, and A. R. Katzev. 2000. "Measuring Program Outcomes: Using Retrospective Pretest Methodology." *American Journal of Evaluation* 21: 341–349.
- Reeves, D. B. 2007. *Ahead of the Curve: The Power of Assessment to Transform Teaching and Learning*. Bloomington, IN: Solution Tree.
- Rockwell, S. K., and H. Kohn. 1989. "Post-Then-Pre Evaluation." *Journal of Extension* 27 (2). Accessed March 11. <http://www.joe.org/joe/1989summer/a5.html>
- Rohs, F. 1999. "Response Shift Bias: A Problem in Evaluating Leadership Development with Self-report Pretest-post-test Measures." *Journal of Agricultural Education* 40: 28–37.
- Rohs, F. R., and C. A. Langone. 1997. "Increased Accuracy in Measuring Leadership Impacts." *Journal of Leadership Studies* 4 (1): 150–159.
- Schneider, M. C., and B. Randel. 2010. "Research on Characteristics of Effective Professional Development Programs for Enhancing Educators' Skills in Formative Assessment." In *Handbook of Formative Assessment*, edited by H. L. Andrade and G. J. Cizek, 251–276. New York: Routledge.
- Spillane, J. P. 2000. "Constructing Ambitious Pedagogy in the Fifth Grade: The Mathematics and Literacy Divide." *The Elementary School Journal* 100 (4): 307–330.
- Sprangers, M., and J. Hoogstraten. 1988. "Response-Style Effects, Response-Shift Bias, and a Bogus Pipeline: A Replication." *Psychological Reports* 62: 11–16.
- Stiggins, R., J. Arter, J. Chappuis, and S. Chappuis. 2006. *Classroom Assessment for Student Learning: Doing it Right – Using it Well*. Upper Saddle River, NJ: Pearson Education, Inc. through Educational Testing Services.
- Stipek, D. 2002. *Motivation to Learn: Integrating Theory and Practice*. 4th ed. New York: Allyn & Bacon.
- Thompson, M., and D. Wiliam. 2007. "Tight but Loose: A Conceptual Framework for Scaling Up School Reforms." Paper presented at the Annual Meetings of the American Educational Research Association and the National Council on Measurement in Education, Chicago, April 9–13.
- Wayne, A. J., K. S. Yoon, P. Zhu, S. Cronen, and M. S. Garet. 2008. "Experimenting with Teacher Professional Development: Motives and Methods." *Educational Researcher* 37 (8): 469–479.
- Wei, R. C., L. Darling-Hammond, A. Andree, N. Richardson, and S. Orphans. 2009. *Professional Learning in the Learning Profession: A Status Report on Teacher Development in the United States and Abroad*. Dallas: National Staff Development Council.
- Yoon, K. S., T. Duncan, S. W.-Y. Lee, B. Scarloss, and K. Shapley. 2007. *Reviewing the Evidence on How Teacher Professional Development Affects Student Achievement*. Washington, DC: US Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.
- Zimmerman, B. J. 1990. "Self-Regulated Learning and Academic Achievement: An Overview." *Educational Psychologist* 25 (1): 3–17.

Copyright of Teacher Development is the property of Routledge and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.