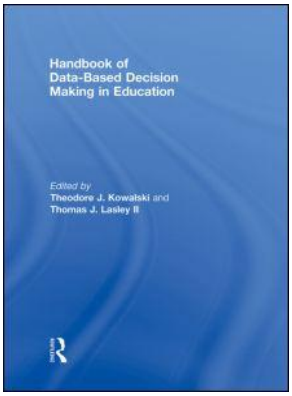

Handbook of Data-Based Decision Making in Education

Theodore J. Kowalski, Thomas J. Lasley

Evidential Reasoning and Decision Support in Assessment of Teacher Practice

Publication details
https://www.routledgehandbooks.com/doi/10.4324/9780203888803.ch22
Arthur M. Recesso, Sally J. Zepeda
Published online on: 13 Oct 2008

Handbook of Data-Based Decision Making in Education

Edited by
Theodore J. Kowalski and
Thomas J. Lasley II

First published 2009
by Routledge
270 Madison Ave, New York, NY 10016

Simultaneously published in the UK
by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

This edition published in the Taylor & Francis e-Library, 2008.

"To purchase your own copy of this or any of Taylor & Francis or Routledge's collection of thousands of eBooks please go to www.eBookstore.tandf.co.uk."

Routledge is an imprint of the Taylor & Francis Group, an informa business

Library of Congress Cataloging-in-Publication Data
Handbook of data-based decision making in education / Theodore J. Kowalski & Thomas J. Lasley II, editors.
p. cm.
Includes bibliographic references and index.
1. School management and organization–Decision making–Handbooks, manuals, etc. I. Kowalski, Theodor 1943– II. Lasley II, Thomas J. 1947–
LB2805 .H2862 2008
371.2 22

ISBN 0-203-88880-4 Master e-book ISBN

ISBN10: 0–415–96503–9 (hbk)
ISBN10: 0–415–96504–7 (pbk)
ISBN10: 0–203–88880–4 (ebk)

ISBN13: 978–0–415–96503–3 (hbk)
ISBN13: 978–0–415–96504–0 (pbk)
ISBN13: 978–0–203–88880–3 (ebk)

22

Evidential Reasoning and Decision Support in Assessment of Teacher Practice

Arthur M. Recesso and Sally J. Zepeda
University of Georgia

Introduction

Applications of evidential reasoning as an approach to assessment of teacher practices have gone unnoticed in education. Other sectors, such as law, use evidential reasoning methods for systematic interpretation of evidence to increase certainty and decrease doubt about events. Teaching is a complex series of events and it is difficult to explain what it means to enact effective teaching or standards-based practices, for example. Novice teachers, those who are preservice or in the induction phases, find it difficult to operationalize such global terms or understand what it means to enact standards-based strategies. Development of pedagogy and pedagogical content knowledge requires substantive feedback from those who assess and guide growth. Evidential reasoning methods and tools bring forward to education an opportunity to overcome this and other challenges. Deconstructing practice into its constitutive elements allows the assessment to be focused and manageable. Thus, an instructional leader is able to discuss collaboratively with a teacher the extent to which a standard of teaching is present or absent.

Our intent here is to present an instantiation of methods and a tool in the context of assessment of teacher practices for growth and support. Through close collaborations with educator preparation and now local schools, we have discovered ways in which principled approaches to systematic use of evidence inform continuous growth and support. We discuss a series of processes and a tool that enhance assessment of teacher practice extensible to all areas of educator preparation, educational assessment, and decision making. From our ongoing research to develop a methodology that tightly couples performance assessment with decision making emerges a principled and systematic approach to using evidence. First, our attention focuses on our purpose to examine teacher assessment and building leader decision making.

National Focus on Teacher Assessment

The No Child Left Behind (NCLB) accountability specifications require continuous monitoring and improvement of student achievement (NCLB, 2001). Much of this accountability falls to individual school principals, who assume many roles and

responsibilities (Leithwood, 1994; Sebring & Bryk, 2000). Paradoxically, many of the provisions of NCLB force leaders to make decisions about the quality of teaching with a limited set, even sole source, of evidence (student achievement data). Leaders are handcuffed by the narrow definition of "evidence-based" adopted from the clinical trial methodologies used in medicine. Clearly, student achievement data are a valuable form of evidence that can explain learner progress toward expected outcomes. However, researchers have found other forms of evidence such as video capture of teachers' practice to have high levels of force in explaining what happened in the classroom leading to learning outcomes (Pea & Hoffert, 2007; Sherin & van Es, 2005; van Es & Sherin, 2002). Furthermore, NCLB requires the leader to train teachers in the uses of evidence-based instructional strategies. This assumes that the leader fully understands the problem based on an analysis of student achievement, and the leader is prepared to make a decision (selecting an appropriate intervention). Other sectors such as law have provided more powerful and useful ways to reach decisions with greater certainty in high stakes situations or making a more accurate diagnosis.

We are suggesting that we have the correct diagnosis—teachers need more support, but the mandated treatment (use of evidence-based interventions) is the very last step in the decision-making process. We propose that school-building leaders can make considerable improvements to the quality of teaching by using evidential reasoning and more appropriate tools to capture evidence to inform decision making. Selecting the most appropriate solution depends on how well informed the decision is. The probability of making the most appropriate decisions increases through the use of logical and systematic inquiry into the relationships between evidence and events and other evidence (Kadane & Schum, 1996; Schum, 1994). The very core of an argument is to increase the certainty of knowing what occurred during the act of teaching. Hence, this certainty allows a better choice of treatment aligned with explanations generated and more purposefully focused on needs. The certainty unfolds more clearly from not only the evidence but also through the methods of collecting and interpreting evidence. The utility of interpretation emerges from the validity of the instruments used to amplify fine grain elements of practice and apply a gradient (Kane, 1992, 2004; Messick, 1987).

Given effective leadership and opportunities to access a wide array of evidence of their influence on student achievement, teachers will improve daily implementation of effective instruction (Stigler, Gonzales, Kawanaka, Knoll, & Serrano, 1999; Teale, Leu, Labbo, & Kinzer, 2002; Zepeda, 2005). Typically, leaders are challenged by local capacity (e.g., funding) to provide support (e.g., instructional supervisors). Often the principal, for example, is overwhelmed with the sheer number of evaluations that must be conducted before the end of the school year. Hence, there are significant hurdles for implementing processes to clarify for the teacher which activities actually influence (or should influence) student learning priorities. Just as beginning teachers assume their approaches to be effective, school leaders assume that their support is helpful, lacking evidence of the impact of their teacher support on student achievement. The cycle becomes self-perpetuating as teachers teach and school leaders evaluate instead of assess—each fulfilling their respective roles without direct evidence of the link between teaching practice being implemented and defined student

needs, and the collective impact of their respective practices on student learning. Leaders can no longer rely solely on informal observation, anecdotal recollection, and unsubstantiated perceptions in their interactions with teachers. We need to break this cycle.

Systemic reform through systematic use of a wide array of evidence can improve leadership quality, teacher quality, and student achievement (Spillane, Halverson, & Diamond, 2001). We propose Evidential Reasoning and Decision Support (ERDS), a systematic approach designed to link leadership, teaching practices, and student achievement, to provide tools through which these links can be made productive in the mutual support provided, and to use evidence of student learning and teaching practices as the cornerstones of planning, implementing, and assessing quality teaching-learning practices (Recesso et al., in press).

Using Evidence for Growth and Support

In this chapter, we focus on the integration of ERDS methods into teacher assessment. Through close collaboration with local schools and preparation programs, we elucidate how evidence-informed assessment practices that systematically identify, collect, and interpret evidence to assist in making decisions. Evidential reasoning methods are combined with instruments to measure progress and to delineate success or deficiency in classroom practices. Feedback is generated from interpretations of multiple pieces of convergent evidence known to embody critical attributes of effective practice presented in a teaching standards framework. Teaching standards that define what a teacher should know and be able to do are widely accessible including the Interstate New Teacher and Support Consortium (INTASC), National Board of Professional Teaching Standards (NBTS), and subject-specific standards such as the National Science Teaching Association (NSTA) and National Council of Teachers of Mathematics (NCTM) Standards. Using ERDS methods and a technology-based tool both teachers and instructional leaders can communicate growth or areas in need of improvement with greater certainty (e.g., Zepeda, 2005, 2006, 2007; Zepeda & Ponticell, 1998).

Foundations of Evidential Reasoning

Evidence is defined as all matter (physical or neurological) that can be detected by some device (organic or mechanical) for the purposes of explaining events linked to a hypothesis (Bentham, as cited in Twining, 1985; Schum, 1994). Something becomes evidence the moment educators passively receive it (someone tells the principal about discipline problems in a classroom) or purposefully seeks it out (classroom observation) to make sense of an event (classroom discipline) and build an argument (discipline rules are not being implemented). Evidence such as instructional practices and learning processes are what are captured through procedures such as classroom observations. Through our work, we have been able to enhance classroom observation through the use of video-capture and more extensive uses of data collection tools (Recesso, in press; Zepeda, 2005, 2007).

Four Stages of Evidence Methodology Related to Assessment

Evidential reasoning, drawing conclusions from multiple sources of evidence, has a complex history. What has become an approach to using evidence in many fields evolved from the early days of philosophers who were contemplating if the Earth was round to the contemporary psychology and Forensic sciences (Twining, 1985). During the last 100 years, health, intelligence, and legal sectors have experienced the greatest influence from the evolution of ideas, methods, and tools grounded in evidential reasoning. Schum (1994) presents one of the most well-articulated frameworks for studying and using evidence across any sector. Recently, educational researchers such as Mislevy, Steinberg, and Almond (2003) and Almond, Steinberg, and Mislevy (2002) have acknowledged the influence of Schum's work in developing innovative methods and technology-based tools for assessing student learning. Evidence Centered Design (ECD) has emerged as a principled way to structure student assessments embedded with tasks to generate evidence of learning (Mislevy, et al., 2003). Furthermore, ECD has been instantiated in Principled Assessment Designs for Inquiry (PADI), a technology-based tool that combines the theoretical foundations of assessment with technology applications. Making the design of quality assessments accessible to researchers and practitioners shows promise for increasing the education sector's capacity to generate evidence known to explain when and to what extent learning occurs.

Making sense of evidence in ways that inform growth and success is the foundation for the ERDS approach. Herein, we present a four-stage methodology for using evidence in systematic ways that inform decisions to improve or replicate effective teaching practices. After introducing the methods, we demonstrate how the stages are integrated into individual teacher performance assessments.

Stages of Evidential Reasoning and Decision Support

Stage 1. Trigger A trigger is a signal or cue that emerges from an event that causes a person to think about support for teacher growth and success.

People do not just make decisions. There is always something, an event or a piece of evidence, which cues us to take notice, think, and then act. Principals do not just assess teachers' practice—they have acquired varying levels of sensitivity based on the source (e.g., report of student achievement data) and situation (e.g., school has not made Academic Yearly Progress) in which triggers are generated. Common sources of triggers associated with teacher assessment are multidimensional and can be categorized as follows.

- Intentional stimuli—these stimuli are emitted purposely and are directed at or sought by the receiver (may be solicited or unsolicited).
  - Directive—mandates, rules, laws, explicit directions, "you must"
  - Suggestive—unsolicited suggestions, "you should"
  - Implied—reports from others with the intention of precipitating decision; scores

Examples—NCLB mandates the use of student achievement test scores.
The School Board passes a policy to adopt new curricula.
Sixth-grade math scores have increased this year.

Unintentional stimuli—these stimuli are emitted without the intention of precipitating a decision on the part of the receiver and are not directed purposely at the receiver (stimuli are unsolicited).

Incidental—observations, reports, experiences, unintentional learning

Natural stimuli—these are naturally occurring, unsolicited, undirected, and unalterable (e.g., demographics, time).

Examples—You observe an innovative way of teaching writing.
Newspaper article claims birth rates are on the rise.
Discipline reports are on the rise from one classroom.

The ability to manage and discriminate between levels of importance of each trigger helps gauge sensitivity to events. Principals must be attentive to the signals (i.e., student discipline problems on the rise) to provide assistance when teachers are in need. In contrast, reacting continuously to many triggers leads to constantly trying to put out fires and getting very little accomplished. The middle ground, finding the most appropriate level of sensitivity in an era of multiple simultaneous reform efforts, enables the principal to focus strategically on a few critical needs. In sum, the triggers linked to teacher assessment are important as they bring attention to what to focus on during a classroom observation. Managing triggers is a learned process and is critical to help one focus on explaining how a specific part of teaching practice is impacting student achievement.

Using a Common Framework to Frame Triggers in Practice

All 50 states have standards for teacher practice and student learning standards (Council of Chief State School Officers, 2003). Furthermore, each of the subject domains (e.g., science) has a standards framework espousing what are valued practices and integral knowledge to enact effective practice. These documents are useful resources for linking triggers to expected practices and generating clear focus—one that is centered on a fine grain attribute, the critical constitutive elements that make up a practice. Assessment of "effective teaching" or even inquiry-based instruction is too global. The concepts are so ambiguous that it is very difficult for teachers (especially novices) to grasp what is meant by effective teaching and, thus, nearly impossible to observe and determine if it occurred. Hence, using a standards document such as those available from national accreditation organizations (NCATE, SACS) and state agencies (Department of Education or Board of Regents) one is able to link a trigger (test scores show low assessment of new learning standards) to the domain-attribute-item (Planning and Instruction: teachers use instructional strategies reflecting a standards-based classroom) and establish the focus of the assessment (observe teachers' enactment of instructional strategies in the classroom). With a clear focus statement, an assessment becomes manageable and purposeful. Once both the principal and the teacher know what will be assessed, they can focus on how to capture evidence of the practice.

A principal might capture a trigger during an informal walkthrough of a classroom (see Figure 22.1). In this case, the principal has documented when the event took place and what was witnessed. Upon returning to the office, the principal was able to align the event to a specific attribute. The process refines what other teachers will be asked to replicate and show about this practice. It becomes clear about what in practice is important. In fact, other teachers can use the same instrument to observe their own or practices of others.

Stage 2. Marshalling Evidence Marshalling evidence includes identifying, collecting, and using a wide array of evidence known to help explain events (Schum, 1994) such as planning (lesson plans), teaching (video of practice), and learning (student work samples or student achievement data). Evidence classification is based on its properties including credibility, relevance, type, and form.

Credibility of evidence rests primarily with the source (Schum, 1994). Self-assessment is often not a factor in recertification decisions because of the high stakes involved. In fact, in law and medicine evidence generated from human memory must be bolstered with other complementary or harmonious evidence to strengthen the direction of an argument. The force of evidence, the extent to which it explains an event with a known level of certainty, is attributed, in part, to knowing the credibility of the source. Video as evidence of an event is often free from bias and thus carries a high level of credibility.

Evidence is relevant when it is known to represent an event (Schum, 1994). A properly positioned video camera (back of the classroom) provides evidence (video of instructional strategies) with high relevance (video from back of the classroom is known to be appropriate for capturing whole class instruction). Student achievement data are evidence of student learning and therefore relevant. The same data are relevant to measuring the effectiveness of instruction but less than other evidence such as video capturing actual enactment of instruction. Hence, to the extent that

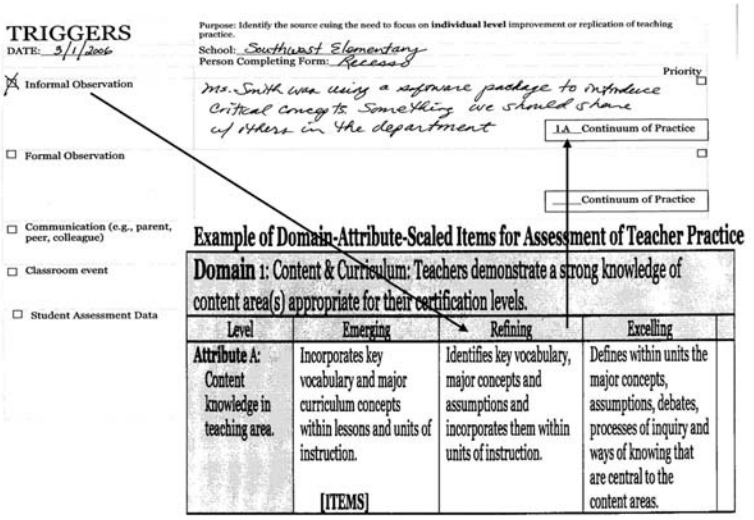

Figure 22.1 Trigger aligned to a common framework.

more relevant evidence is available it would be the first choice in the basis for decision making.

Very rarely does evidence provide absolute certainty (especially a sole source). Variance in levels of credibility of sources and relevance of evidence are important factors leading decision makers to consider how to marshal and combine evidence known to help explain events with a high level of certainty. An understanding of the types and forms of evidence further assists in selecting the best available evidence known to help explain the relationship between leadership decision making, teacher practice, and student performance. Table 22.1 provides a sample of the types and forms of evidence our school and teacher education program collaborators found to influence assessment of teacher practice.

Read Table 22.1 from bottom to top. You will notice a link between the leader's vision setting, teacher's enactment, and learner's demonstration of knowledge. Hence, the evidence is an embodiment of an event (vision setting, enactment, demonstrating learning). This is not to say that all of the evidence must be collected and analyzed. Depending on the question being asked, you may purposely select the best evidence to explain the events. Concentrating on the teacher's practice, for example, the principal may collect a lesson plan, video of enactment, and student work sample as evidence that can easily be shared with other teachers to see the success.

Stage 3. Interpretation: Making Sense of the Evidence Sense making can be enhanced by using instruments that are aligned with a common framework of teaching practices that amplify critical attributes known to influence student learning.

Interpretation extracts information from evidence to generate explanations. The power of an explanation increases through the use of instruments (tests, rubrics, observation forms) known to increase certainty and decrease doubt (Mislevy & Gitomer, 1996; Shafer, 1976; Yang & Singh, 1994). That is, through the use of well-designed valid instruments one can say with confidence that a classroom event (a teacher's use of software for students to construct knowledge about ratios) leads to student mastery knowledge. Using the metaphor of a lens to amplify attributes of practice or suppress the noise of extraneous information, we can operationalize how an instrument helps explain those parts of practice that need improvement (Hill,

Table 22.1 Evidence of teacher and learner performance.

Type	*Form*
Learner performance	Standardized test scores Student work samples Responses to open-ended questions
Teacher	Video enactment of strategies Lesson plan with strategies defined Pre-planning discussion about strategies to use
Leader practice	Observation of teacher's practices Funding professional learning on instructional strategies Vision for teaching and learning

Hannafin, & Recesso, 2008). Principals who assess teacher practice are often observing teachers outside of the subject domains in which they are familiar (Zepeda & Mayers, 2004). Even for principals who do have familiarity with a subject area, inquiry can look very messy where practices are non-traditional. Hence, using a lens created to focus on the attributes of a specific practice (students working in small groups and using simple machines to learn the concepts of mechanical advantage) improves the probability for quality feedback given to the teacher (Zepeda, 2005, 2007; Zepeda & Mayers, 2004). The power of interpretation is further enhanced when an instrument is applied to combinations of evidence, each contributing a given level of certainty, from which one derives an explanation that can be tightly aligned with an intervention (professional learning) or useful resource (appropriate instructional materials).

Stage 4. Course of Action Select and implement intervention or resource appropriate for improving deficiencies and replicating success in teaching practices.

The final stage in the first iteration of the cycle is making the decision. A decision is the act of selecting an intervention and resource from existing knowledge and then implementing it in an effort to support changes in practice or replicate success. Interventions may involve focused professional development and other supports such as peer coaching. The decision is to link the teacher's needs (e.g., content knowledge, pedagogy, pedagogical content knowledge) with a support mechanism for learning and then enactment demonstrating progress toward meeting an expectation of increased knowledge and understanding about practice. A solution may be to adopt a new curriculum known to increase student achievement. It is critical not to operate from a deficit model. There are great successes in classrooms that are often not captured. Clearly, we need to do a better job of establishing processes for capturing and codifying through appropriate lenses what it means to enact standards-based practices in classrooms for others to see and learn. Having such a resource affords all teachers access to concrete deconstructed examples of success to support their learning and to improve teaching in ways that impact student learning and achievement.

Affordances of Video Evidence and Technology-Based Tools

Video is a powerful form of evidence for capturing actual events for live or post-event analysis. The utility of this and other kinds of evidence is realized through the use of Web-based technology to relieve the user's cognitive load when using an intricate methodology. Furthermore, the advent of user-friendly graphic interfaces and embedded guides aids in gaining efficiency. In this section we demonstrate how existing technologies give instructional leaders a tool to systematically link triggers, foci, and video evidence to understand with increased certainty and explain events as they really occurred. After providing an introduction to one type of tool for using ERDS methods and interpreting video we will present a case from our work in the field.

Video Capture Tools

Video capture equipment has been used in education since the 1960s. Video of classroom practices has since been a tool used in teacher education programs as a way of promoting reflection (Fuller & Manning, 1973). For example, student teachers often practice in front of the camera or even capture actual classroom practices to bring back to the university for collaborative reflection with peers and faculty supervisors. More recently, video has become a part of mentoring processes for beginning teachers (Pea & Hoffert, 2007; Sherin & van Es, 2005; van Es & Sherin, 2002). As advances in technology make video annotation easier and more affordable, the pace for developing tools also quickens. A catalyst for this work was the U.S. Education Department's Preparing Tomorrow's Teachers to use Technology (PT3) Program. Recesso and Hannafin (2003), for example, developed a video annotation tool with the intent that it be integrated into preservice preparation programs as an instrument to support assessment and continuous growth.

The Video Analysis Tool (VAT) implements standardized Web-based interface functions to populate a robust database and infinitely expandable storage system (Recesso et al., in press). Teachers have two options for getting their video into VAT. Video cameras connected to a computer network, called Internet Protocol (IP) video cameras, capture live events and immediately transmit video for live or post-event analysis. A great benefit of this technology is in the design of the camera. No tapes are required, there are no batteries to replace, and the teacher does not have to be concerned with the video camera configuration. Once the configuration of the camera is set through the interface in VAT, the video camera starts capturing, the video file is sent to storage, and the video camera turns off as stipulated in the teacher-defined settings.

As a second option, video of teaching and learning may also be uploaded manually into the system for analysis. This provides end users with flexibility to use any video recording equipment available to them. Teachers may choose to use any typical consumer-line video camera that is most readily available, capture the video, and upload it from tape or video camera embedded hard-drive to the VAT system. If the teacher chooses to use the live streaming functionality, the system permits the teacher to allow someone access (e.g., instructional leader) to view and code video of the practices live (as the events actually take place). Video uploaded post-event also requires that the teacher give permission to others (e.g., peer) who may simultaneously view and code the video. If the viewers so choose, they may code and annotate the video, all of which is stored in the database with the video. All of the metadata associated with segmenting video (e.g., coding using a rubric) into manageable parts (e.g., a three-minute video clip as compared to an hour), the video, and descriptions provided by users are stored in a database ready for search, retrieval, and analysis on demand.

Users access a video collection through the Web interface and choose to analyze or view videos. Clips are segmented through the *analyze video* function by depressing start time and end time (see Figure 22.2). Annotation of the clip is added through the standard text entry *comments* window on the right side of the browser. Users who have preselected a *lens* are also able to associate the *clips* (bottom left corner of browser) with attributes of practice and a level of performance (e.g., basic, advanced, proficient).

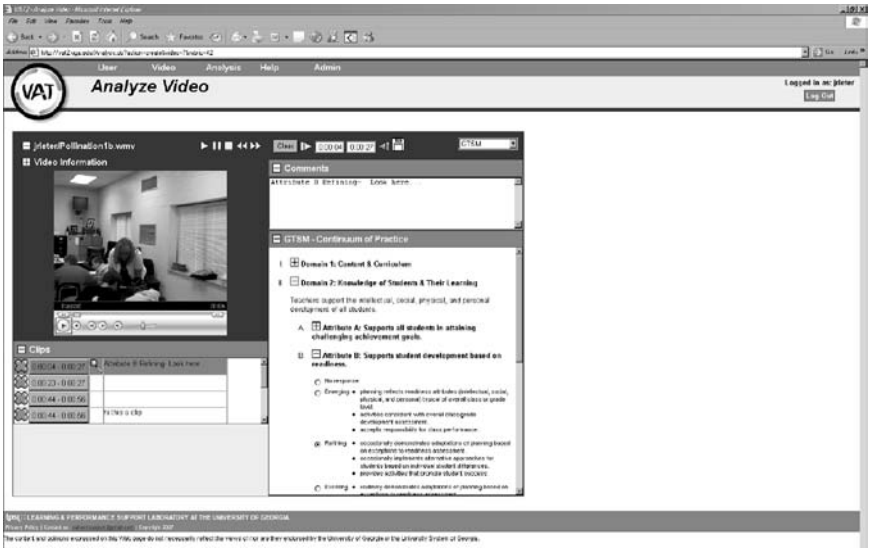

Figure 22.2 User analysis of video to segment clips in VAT.

Video owners may choose to view codified video themselves and/or share clips with others through the *view analysis* function. The principal and teacher, for example, may view the clips together or asynchronously each providing insight into the events as they actually unfolded in the classroom. Mentors may also request the use of the teacher's videos as exemplars (e.g., standards-based practices) to be shared with other preservice teachers. In VAT, teachers are owners of the video they capture, and they have complete control of who views or analyzes their practices. By accessing the *show analysis* function, teachers grant or deny permission for someone else to access the video (see Figure 22.3). With permission, other teachers are able to access the video and associated codes to see what constitutes practice (e.g., mastery of questioning strategies).

As a third option, users may compare and contrast events and/or the coding associated with video clips using the *view analysis* function and selecting two video files (see Figure 22.4). A multitude of uses have emerged for the dual-view function. Using VAT to create exemplars, videos coded by experts in the field, one can test his or her knowledge of practice (see if the same attributes of effective teaching) can be found or even compare his or her own practices. Teachers using VAT have uploaded a video of their classroom practice and multiple videos of students working in small groups. Using the view analysis function they can simultaneously view teaching and students' learning. This has been a powerful view for teachers to compare and contrast practice (e.g., expertly coded case compared to the teacher's coding of the same video) or even simultaneously view teacher practices in one window while students construct and apply knowledge in the other window. A teacher participating in one of our projects actually generated this idea—she wanted to see how the students reacted to her teaching as feedback about her use of strategies facilitating learning of mathematics. Teachers use VAT to see strategies *in situ* and to code discrepancies between espoused and actual practice.

In summary, the following illustrates how VAT directly supports ERDS:

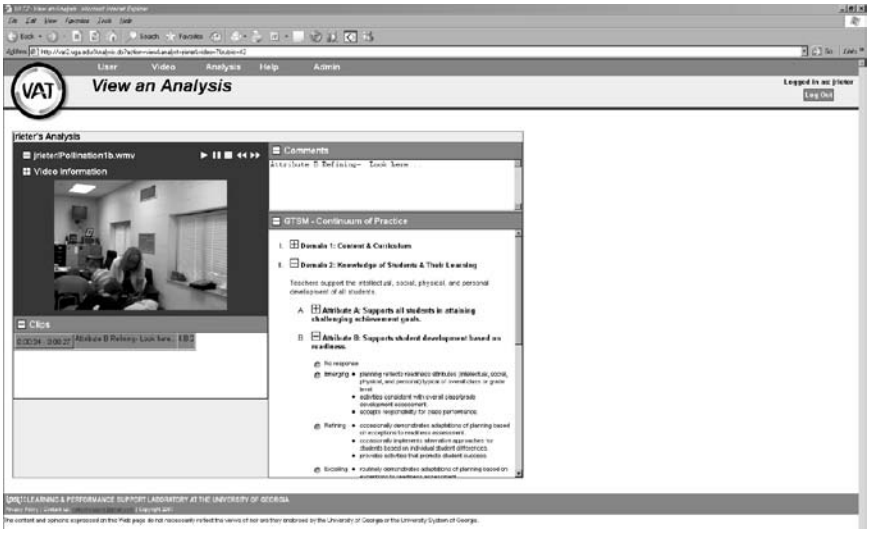

Figure 22.3 View analysis completed in VAT.

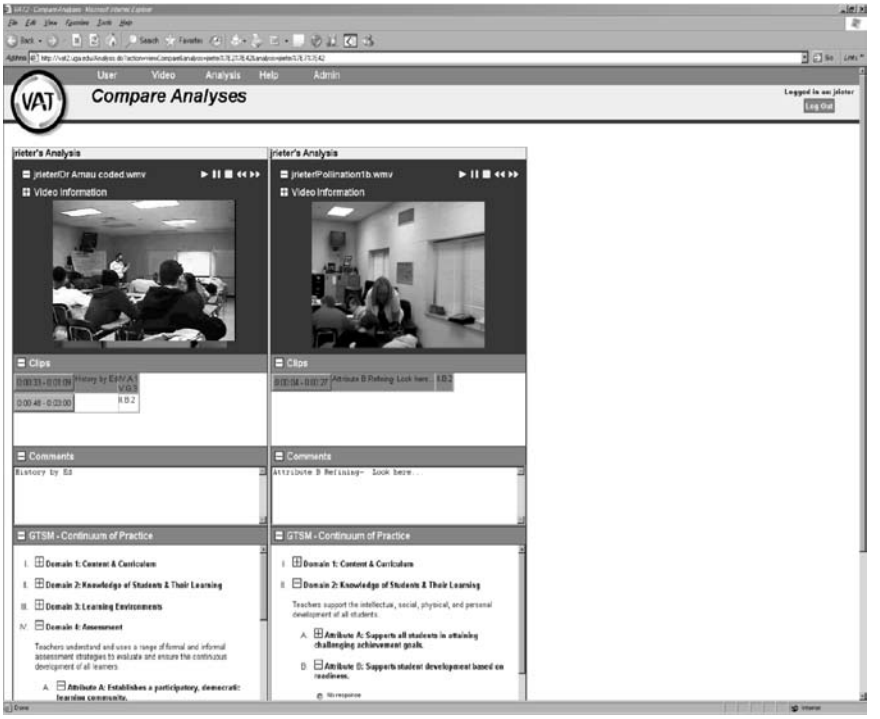

Figure 22.4 Compare analyses conducted in VAT.

VAT for Learning Triggers:

- Educators learn about triggers by watching pre-captured events (i.e., instructional strategies not aligned to expected practices) with an instructional leader who points out the specifics of an event that are important. Users also learn to gauge sensitivity to triggers and to act accordingly.

- Instruments within the VAT system assist educators' alignment of triggers to standards of practice—thus increasing granularity and clarifying the focus on improving a specific attribute of practice.

VAT for Collecting and Using Evidence:

- Live capture and post-event upload of video evidence is permitted for the user to see and share classroom practices. Users can capture evidence within close proximity of the actual event (e.g., camera focused on the student group conducting an experiment as it happens in the class). Furthermore, VAT allows for continuous replay and permanent storage for future access.

VAT for Interpretation and Explanations:

- Teaching frameworks (teaching and learning standards) and instruments (validated for measuring progress) can be uploaded into the system for subsequent codification of events. The process allows users to assess self and others' practices, gauge progress, and determine specific needs for support.

ERDS and VAT in Teacher Assessment

In this section, we present a case study in redesigning teacher assessment through the design of procedures informed by the principles and processes of ERDS. There are many complexities associated with the assessment of teaching and learning, and there are many ways to approach it (Danielson & McGreal, 2000; Wilkerson & Lang, 2007; Zepeda, 2005, 2007). Assessments vary for the purposes they serve. For example, testing for licensure at entry into the profession is a type of assessment. However, assessments are made across the career from the point of entry to the end of the career. Assessments can include standard evaluation of teaching using check-lists, classroom observations, and surveys. In this chapter, we are examining the assessment of teaching through the lens of evidential reasoning. The process is enhanced through VAT and its affordances to capture performance and enactments of teaching and student learning. Following what we know about the power of data collection through the uses of instructional supervision where a school leader conducts pre-observation conferences, observes teachers teaching, and then conducts post-observation conferences, we propose that there is a more reasoned and purposeful way to collect evidence for teachers and leaders to make sense. This approach is examined in this case.

Three pillars of the approach are anchored to what we know are critical aspects of effective instructional supervision (Zepeda 2005, 2006, 2007; Zepeda & Ponticell, 1998):

Pillar I. Measure growth
Pillar II. Provide continuous support
Pillar III. Use highly collaborative processes

We add a fourth pillar; a deliberate focus on evidential reasoning to provide explanation and insight constructed by the teacher and leader to better understand the teaching and learning events.

Pillar IV. Systematic use of evidence

To provide general parameters for the discussion, we describe the approach as it is implemented with individual induction teachers in order to make decisions using evidence to determine needs for and success of differentiated success support.

Case Context

We offer a case to illustrate how ERDS methods complement the work of the principal enacting instructional leadership and the practice of classroom observations vis-à-vis instructional supervision in a local school context. In brief, for the context of our illustration, instructional supervision is defined as the pre-observation conference, the classroom observation, and the post-observation conference. The clinical model of instructional supervision is a mainstay in assessment of teacher performance. Morton Middle School (MMS) is part of a large and fast growing school system located outside of a large metro area. MMS is implementing new statewide student performance standards in sixth-grade mathematics this year and sixth-grade science next year. The principal is new to MMS and has a successful track record in other buildings and is well respected in the school system. Thus far, the teachers have attended mandatory training about the standards provided by the state education department and the local system. Next school year, the schools will be implementing the standards. The school system's central administration has required all principals to report on implementation progress. The superintendent and Board want to know how successful teachers are in implementing new standards-based practices. Clearly, a decline in student performance will not be acceptable. The principal of MMS has added pressure of leading a school on the "needs improvement" list for two consecutive years. Furthermore, it is clear that the previous principal's teacher assessment processes were perfunctory; he used check-lists and spent little time in classrooms. The new principal began using the ERDS approach and created procedures for conducting teacher assessments.

Bringing Teacher Practice into Focus

MMS building-level performance has historically been lower than the school system and state averages. The principal has in hand the student achievement data reports from the previous years. There is a pattern of students doing poorly in sub-domain areas such as decimals and fractions. The principal has moved attention from the very global (school is underperforming) to the less ambiguous (new learning performance standards) to a more refined focus (sixth-grade mathematics: decimals and fractions). Yet, the principal knows the focus is not fine grain enough. The principal

wants the focus to be observable (e.g., using computer software to teach decimals) and measurable along a continuum of growth (no evidence to exemplary practice) to fit into the plan for revitalizing the teacher assessment model. Hence, the principal engages faculty and staff to articulate what makes something standards-based teaching and to state specifically what standards-based teaching of decimals is and looks like. Together the principal and teachers ask if this expectation of standards-based practice is different from current practices and if so, how?

The principal uses the triggers (new state-mandated sixth-grade mathematics learning standards, low school performance in sixth-grade mathematics) to communicate schoolwide needs for improvement in a meeting with faculty and staff. The principal reiterates a commitment to revising the current teacher assessment procedures to define individual needs for growth and support. At the meeting, the principal presents the issues and uses a visual aid to highlight student performance standards (with data showing "needs improvement" based on existing test scores). Next, a version of the image shows the alignment of teacher performance standards (and the specific attributes that are critical elements of the practice) with the student learning standards. A rubric for assessing progress is shown and explained. It becomes clear there is a lot of work to be done, more than could be attended to in one year. The principal opens a discussion of what attributes all teachers will focus on first (they choose two to three for the year) and asks grade level subject area teachers (sixth-grade math) to add others as part of their assessment. As a final step, the principal and teachers state the foci and how they are aligned to teacher and learner performance standards. The principal announces that in the pre-observation conference there will be further discussion about how the interpretation of evidence will inform what types of differentiated support will be provided (Zepeda, 2007).

Individual Pre-observation Conference

A pre-observation conference is arranged with each sixth-grade mathematics teacher. Collaboratively, the principal and teachers write a focus statement for each of the two or three specific attributes they agreed to address this school year. Together, they develop an assessment plan including an evidence collection component. Many teachers choose to use the Video Analysis Tool (VAT) as a way to capture practices and student discourse.

Video Evidence of Teaching and Learning Events

The teacher refers to lesson plans to see when manipulatives will be used in class to teach fractions. Through VAT, the teacher is able to schedule the date for the Internet Protocol (IP) camera in the room to turn on automatically, record video of the teaching events, and audio of the discourse taking place. The IP video camera has been pre-mounted on the wall at the rear of the classroom. It has a sensitive mic to pick up the soft voices of the children and the video is very high quality. The camera requires no tapes or other settings. It turns on, captures the video, and streams it

automatically across a secure network to mass storage. The teacher focuses on teaching and is able to view the video when she has time. In this case, the teacher permits the principal to view and code the live video through VAT.

Both the principal and the teacher access the video and code the events based on what they are viewing through VAT. They watch the video, look for important events defined by the foci, and place a digital marker at a place in time when they want to come back and visit later. The value of VAT is that the teacher can view the teaching by herself or in the company of the principal, perhaps a peer coach, or a mentor. The functions within VAT allow the principal and teacher to center their attention on the events while simply clicking a button to identify something of interest such as the use of manipulatives for teaching fractions. Later, they can return to these markers to annotate more refined feedback by using a lens that associates the events with a standard and a metric of quality of performance. In the present case, the use of manipulatives as a specific learner-centered strategy surrounding fractions is codified with NCTM standards and a 4-point growth scale of *not evident* to *mastery*. VAT allows the teacher to compare and contrast her own coding with the principal's coding through a dual-view window. Thus, feedback is very concrete and it is clear where there is success and needs for minor adjustments in the strategies that will benefit students.

Post-observation Conference: Making Sense of the Evidence

The principal and teacher have documents from the pre-observation conference, and the video capture of the teacher's classroom practices. During the post-observation conference, the teacher and the principal review the video capture and the annotations made on screen. Using the annotations and stopping at critical incidents, the teacher and the principal are able to discuss and interpret classroom practices. The principal and the teacher are able to pinpoint and assess instructional practices intended to support students' meeting content standards. Because VAT captures the totality of the room, the principal and teacher are also able to focus on how students respond to instruction. To support ongoing discussion and dialogue, the following guidelines are offered:

- Did the focus change during the process? If so, what effect did it have?
- Were there any discrepancies between your beliefs or recollection about practice and what the evidence says actually happened? If so, explain why.
- What was learned about progress toward goals and expectations established in the pre-observation conference?
- Is success or need for improvement defined in such a way that a course of action can be set?

Through these questions, the teacher and principal sit down at a computer and use VAT to watch a three-minute coded video clip and a five-minute video clip. Analyzing the evidence, they discover that the teacher did have access to resources appropriate

for the age group and ability of the students. As they transition to talking about a course of action, the principal and teacher jointly agree that the teacher will attend a workshop on using manipulatives that is being supported through a grant awarded to the school system. They also agree that the teacher should observe coded videos of other teachers who have had success with this particular standard. The principal documents the feedback and discussion, and they agree to continue revisiting certain instructional strategies related to the standards.

Discussion

Using evidence-informed methods is a transformation in the way instructional leaders conduct teacher assessment and make decisions. Using evidential reasoning and the systematic interpretation of evidence are one way to address the uncertainty with which teachers' needs are defined and aligned with resources for improvement or replication of success. Currently, we are using the methods and tools in collaboration with those who assess prospective and practicing teachers in several different contexts as they develop and refine their professional knowledge and skills for teaching and their capacity to engage in continuous growth.

The instructional leaders and teachers with whom we have worked have gained greater self-knowledge of how to enact practices and goals for teaching in their domain areas. It is not easy for instructional leaders or teachers to confront and critique teaching, particularly new building level leaders who are in a phase of their career when certainty, self-confidence, and time to thoughtfully engage in assessment of teaching practices are limited. Providing both instructional leaders and teachers with opportunities early in their career to purposefully focus on assessment of practices will help them develop personal familiarity and comfort with looking for the most powerful and appropriate evidence, and building explanations from evidence in the process of making teaching- and learning-related decisions.

We have partnered with local schools to provide continuous access to VAT. In addition, educator preparation programs in the College of Education have integrated VAT/ERDS as a tool for assessing progress during field experiences. We expect our investigation with new populations of VAT users to reveal more systematic ways to support teachers' development and refinement of practice.

Technology to Support Decision Making

Multiple support mechanisms are often in place to assist with the building level and individual teacher assessment, success replication, and improvement procedures. Technology-based Support Systems are becoming more pervasive as usability and utility improve. State agencies, national organizations, universities, and research labs continue to develop Web-based versions of processes and tools to support the assessment processes. These systems often scaffold a form of structured assistance that can disappear as users become more proficient in the system. Built-in functions allow people to build cases that can be stored and contributed to by multiple people

involved in the process. Users are able to upload and interpret evidence using a series of tools, build reports and portfolios, develop a growth plan, and enable expert review and assistance. Due to international standards set for programming it has become possible to connect multiple systems and share resources. A state agency may be developing a Web-portal, for example, to store lesson plans representing effective planning for standards-based classrooms. Another system that captures video of practices and aligns segmented clips with the same standards can be interconnected and viewed by users on a system. Such technology-based systems can only enhance and not supplant critical support mechanisms such as mentoring, coaching, critical friends, study groups, lesson study, action research, and growth plan development.

VAT Limitations/Considerations While there are many positive aspects of VAT, there are also limitations and considerations to take into account. Perhaps the most common technical concern we encounter in using VAT with groups of teachers has to do with loading video into the system. If VAT users wish to load their own video into the system, they must know how to transfer video to their computer and convert the files to .wmv files before loading onto VAT. Often, an intermediary needs to be responsible for making sure the video is converted and loaded to VAT.

Second, in our work with area schools, we have uncovered technical challenges such as schools blocking video streaming into the school (hence, no video resources for teachers can be accessed); limited bandwidth, causing schools to be concerned about streaming across their networks; and limited expertise of school technology coordinators who are ultimately responsible for managing the school's network. Within a classroom, the viewing range of the video camera has been problematic at times. Mounted video cameras may limit the viewing area unless a wide-angle lens is attached to the camera. In addition, mounted cameras sometimes do not capture conversation clearly, nor are they able to record fine grain details of classroom events such as the writing on students' paper or white boards. Clearly, users can overcome all of these issues, but they may involve more time and effort than a user is willing to devote.

Finally, VAT is currently platform specific. VAT is accessible using a PC with a Windows operating system (XP, 2000, 98), Internet Explorer 6.0+, and Windows Media Player 9 or higher. Broadband internet provides the highest quality experience when a dedicated connection is available. Although wireless networks are widely accessible in public spaces, the experience of viewing video can be diminished if the bandwidth has been throttled by the provider. Any other required software such as Java-based applets is automatically downloaded when the user accesses the system. Given these limitations, we have been able to successfully use VAT in a variety of contexts with both individuals and groups of instructional leaders and teachers. When working with a group, we provide one session of group training for the use of VAT. We have been able to introduce VAT and allow for practice time within a 30-minute (or less) block of time. During these short VAT training sessions, we provide users with a "cheat sheet" that is downloadable from the VAT website.

Expanding the Use of VAT VAT is undoubtedly a powerful tool in instructional leadership and assessment of teacher practices to encourage teachers to make purposeful and systematic inquiry a habit that will become increasingly valuable throughout their careers. In this chapter, we have focused on using VAT for the principal's assessment of teachers' practices. However, the use of VAT may be expanded to include opportunities for inquiry into one's own practice. VAT clips may be used as abbreviated video cases—"windows" into other teachers' classrooms that provide a means for learning others' practices in the process of defining and assessing one's own practice. The sixth-grade math teacher may choose clips of exemplary practices to demonstrate standards-based practices; for example, what instruction using investigations strategies looks like. By engaging in VAT analysis of others' practice, instructional leaders and teachers can clarify, confront, and revise their teaching practices in ways that impact student learning.

Conclusion

We have presented a series of processes and a technology-based tool that enhance decision making about teacher growth and development. ERDS is one conceptualization of how the principles and processes of a methodology (Recesso et al., in press) are translated into procedures for teacher assessment. Given evidence of supervisory support that influences classroom practice and student achievement, leaders can increase their understanding of the conditions that warrant various types and forms of assistance. Such an understanding of practice directly informs the next iteration of research and development on mechanisms of continuous support such as the next evidence-analysis tools to assist educational decision making.

References

Almond, R. G., Steinberg, L. S., & Mislevy, R. J. (2002). Enhancing the design and delivery of assessment systems: A four-process architecture. *Journal of Technology, Learning, and Assessment*, *1*(5). Retrieved September 10, 2007, from http://escholarship.bc.edu/jtla/vol1/5

Council of Chief State School Officers (2003). *State content standards: A 50 state resource*. Retrieved January 20, 2008, from http://www.ccsso.org/content/pdfs/StateContentStandards.pdf

Danielson, C., & McGreal, T. L. (2000). *Teacher evaluation to enhance professional practice*. Alexandria, VA: Association for Supervision and Curriculum Development.

Fuller, F. F., & Manning, B. A. (1973). Self-confrontation reviewed: A conceptualization for video playback in teacher education. *Review of Educational Research*, *43*(4), 469–528.

Hill, J., Hannafin, M., & Recesso, A. (2008). Creating a patchwork quilt for teaching and learning: The use of learning objects in teacher education. In P. Northrup (Ed.), *Learning objects for instruction: Design and evaluation* (pp. 261–280). Hershey, PA: Information Science Publishing.

Kadane, J. B., & Schum, D. A. (1996). *A probabilistic analysis of the Sacco and Vanzetti evidence*. New York: Wiley.

Kane, M. (1992). An argument-based approach to validity. *Psychological Bulletin*, *112*(33), 527–535.

Kane, M. (2004). Certification testing as an illustration of argument-based validation. *Measurement*, *2*(3), 135–170.

Leithwood, K. (1994). Leadership for restructuring. *Educational Administration Quarterly*, *30*(4), 498–518.

Messick, S. (1987). *Validity*. Princeton, NJ: Educational Testing Service.

Mislevy, R. J., & Gitomer, D. H. (1996). The role of probability-based inference in an intelligent tutoring system. *User Modeling and User-Adopted Interaction*, 5, 252–282.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives*, *1*(1), 3–62.

No Child Left Behind Act of 2001, 20 U.S.C. 6301 et. Seq. (2001).

Pea, R., & Hoffert, E. (2007). Video workflow in the learning sciences: Prospects of emerging technologies for augmenting work practices. In R. Goldman, R. Pea, B. Barron, & S. J. Derry (Eds.), *Video research in the learning sciences* (pp. 427–460). Mahwah, NJ: Lawrence Erlbaum Associates.

Recesso, A., & Hannafin, M. (2003). *Evidence-based technology enhanced alternative curriculum in higher education (ETEACH).* Project # P342A030009. Proposal funded by the U.S. Department of Education's Preparing Tomorrow's Teachers to use Technology (PT3) program office. Washington, DC: U.S. Department of Education.

Recesso, A., Hannafin, M., Wang, F., Deaton, B., Shepherd, C., & Rich, P. (in press). Direct evidence and the continuous evolution of teacher practice. In P. Adamy & N. Milman (Eds.), *Evaluating electionic portfolios in teacher education.* Greenwich, CT: Information Age Publishing, Inc.

Reilly, D. (2000). Linear or nonlinear? A metacognitive analysis of educational assumptions and reform efforts. *International Journal of Education Management*, *14*(1), 7–15.

Schum, D. A. (1994). *The evidential foundations of probabilistic reasoning.* New York: John Wiley & Sons.

Sebring, P. B., & Bryk, A. S. (2000). School leadership and the bottom line in Chicago. *Phi Delta Kappan*, *81*(6), 440–443.

Shafer, G. (1976). A theory of statistical evidence. In W. L. Harper & C. A. Hooker (Eds.), *Foundations of probability theory, statistical inference, and statistical theories of science, Vol. II* (pp. 365–436). Boston: Reidel Publishing Company.

Sherin, M. G., & van Es, E. A. (2005). Using video to support teachers' ability to notice classroom interactions. *Journal of Technology and Teacher Education*, *13*, 475–491.

Spillane, J. P., Halverson, R., & Diamond, J. B. (2001). Investigating school leadership practice: A distributed perspective. *Educational Researcher*, *30*(3), 23–28.

Stigler, J. W., Gonzales, P., Kawanaka, T., Knoll, S., & Serrano, A. (1999). *The TIMSS videotape classroom study: Methods and findings from an exploratory research project on eighth-grade mathematics instruction in Germany, Japan, and the United States.* Washington, DC: U.S. Department of Education.

Teale, W., Leu, D., Labbo, L., & Kinzer, C. (2002). The CTELL project: New ways technology can help educate tomorrow's reading teachers. *The Reading Teacher*, *55*(7), 654–659.

Twining, W. L. (1985). *Theories of evidence: Bentham and Wigmore.* Stanford, CA: Stanford University Press.

van Es, E. A., & Sherin, M. G. (2002). Learning to notice: Scaffolding new teachers' interpretations of classroom interactions. *Journal of Technology and Teacher Education*, *10*(4), 571–596.

Wilkerson, J. R., & Lang, W. S. (2007). *Assessing teacher competency: Five standards-based steps to valid measurement using the CAATS model.* Thousand Oaks, CA: Corwin Press.

Yang J. B., & Singh, M. G. (1994). An evidential reasoning approach for multiple attribute decision making with uncertainty. *IEEE Transactions on Systems, Man, and Cybernetics*, *24*(1), 1–18.

Zepeda, S. J. (2005). *The instructional leader's guide to informal classroom observations.* Larchmont, NY: Eye on Education.

Zepeda, S. J. (2006). Classroom-based assessments of teaching and learning. In J. H. Stronge (Ed.), *Evaluating teaching: A guide to current thinking and best practice* (2nd ed., pp. 101–124). Thousand Oaks, CA: Corwin Press.

Zepeda, S. J. (2007). *Instructional supervision: Applying tools and concepts* (2nd ed.). Larchmont, NY: Eye on Education.

Zepeda, S. J., & Mayers, R. S. (2004). *Supervision across the content areas.* Larchmont, NY: Eye on Education.

Zepeda, S. J., & Ponticell, J. A. (1998). At cross-purposes: What do teachers need, want, and get from supervision? *Journal of Curriculum and Supervision*, *14*(1), 68–87.