Reframing Portfolio Evidence
Empowering Teachers through Single-Case Frameworks

Craig E. Shepherd
The University of Wyoming
& Michael J. Hannafin
The University of Georgia

Introduction

Bransford, Darling-Hammond, and LePage (2005) indicate that quality instruction requires teachers to navigate learner characteristics, content knowledge, and pedagogy through social and political climates. In the past 30 years, every state has increased requirements for teachers to enter the profession (Zumwalt & Craig, 2005). Yet assessing effective teachers remains difficult; although teacher inputs play a role, so too do student prior achievement and family and community norms (Fallon, 2006; Hanushek & Rivken, 2006, 2010). Traditionally, assessment of teacher quality relied on standardized testing, peer and administrator observations, student achievement, and related activities. Yet many aspects of effective teaching remain elusive with these assessments (Cochran-Smith & the Boston College Evidence Team, 2009; Nazier, 1997).

Standardized test scores aligned to state and national standards are common measures of teacher and school quality in the United States. They allow administrators and policy makers to track and compare student achievement over time. They also certify that teachers have sufficient background knowledge in specified subject areas. However, standardized tests measure content knowledge while minimally considering classroom practice, student background, or individual needs and circumstances. To complement standardized subject assessment exams, teacher education programs provide multiple field experiences for purposes of guidance, methods exploration, and student accountability.

Shaping and measuring the success of these programs is difficult and costly. Placements are often made at schools located several miles from university settings. Cooperating teachers and mentors are sometimes selected on availability rather than quality. Periodic observations require large time commitments, may interrupt classroom dynamics, and are limited in number. In-service teachers face similar challenges. Although support personnel reside within local schools, they often have their own teaching or administrative responsibilities and may lack content knowledge to judge instructional quality (Kelley, 2004; West, Rich, Shepherd, Recesso, & Hannafin, 2009).

To better measure teacher effectiveness, Fallon (2006) and the National Research Council (2005) suggest that researchers turn to longitudinal, quantitative studies focused on student achievement. However, others question the type and quality of evidence these studies reveal. Cochran-Smith and the Boston College Evidence Team (2009) claim that value questions embedded within educational research may not be addressed with traditional quantitative methods and recommend using multiple approaches to explore teacher effectiveness. Cochran-Smith (2006) suggested that researchers must consider the purpose of evidence collection, who collected it, and under what contexts as they examine teacher effectiveness. These circumstances guide researchers on the quality and usefulness of evidence collections.

In addition to these concerns, teachers often lack proper training to conduct rigorous, scientific research and lack access to effective sample sizes (Fallon, 2006; Rippon & Martin, 2006). A teacher's primary responsibility is instruction. It is doubtful that teachers can conduct advanced quantitative research while simultaneously meeting students' learning needs. Researchers who possess these credentials are often distanced from classroom settings and may fail to capture relevant and adequate measures of classroom instruction. While not minimizing the importance of evidence identified through complex quantitative studies, these approaches are unrealistic for most teachers, who must rely on their own devices for classroom evaluation and development. Alternative evidence collection, organization, and analysis methods are needed. The purpose of this article is to propose a framework of evidence selection and organization through portfolio development and provide guidance for evidence selection that accommodates valid formative practices in classroom settings.

Portfolios to Capture and Evaluate Teaching

Portfolios are collections of purposefully selected materials, organized to depict and examine professional practice (Grossman, 2005; Hartmann,

2004). Unlike experimental and quasi-experimental studies that require numerous participants to draw generalized conclusions, portfolios may allow practicing teachers to focus on themselves and their classrooms for purposes of assessment and professional development. Portfolios are believed to unobtrusively document events through classroom artifacts gathered by instructors and support personnel (Habib & Wittek, 2007; Zepeda, 2002). Since the mid 1980s many teacher education programs turned to portfolios to document teacher practice and growth; promote inquiry, reflection, and skill development; and assess competency (Dhonau & McApline, 2005; Hallman, 2007; Rickards et al., 2008; Wetzel & Strudler, 2005).

Land and Zembal-Saul (2003) found that portfolios documented the thought processes of 20 preservice teachers who were exploring properties of light (e.g., reflection, refraction). Using factor analysis to compare four portfolio implementations across pre- and in-service teachers, Beck, Livne, and Bear (2005) found that professional development portfolios heightened perceived knowledge of teacher roles, reflective practices, and peer collaboration. Portfolios may also structure mentoring relationships and help novice teachers examine their own and others' practices (Kelley, 2004; Orland-Barak, 2005; Redish, Webb, & Jiang, 2006). Rolheiser and Schwartz (2001) reported that teachers who incorporated portfolios to assess student performance clarified their own teaching philosophies and developed structured arguments regarding student achievement to share with parents and administration. The act of identifying, collecting, organizing, and examining teaching artifacts to recreate events may help teachers reconstruct and improve practice.

Teachers completing National Board for Professional Teaching Standards (NBPTS) certification must create portfolios that include video recordings, student work samples, and other commentaries to depict classroom practice and aid summative assessment (NBPTS, 2010; Silver, Mesa, Morris, Star, & Benken, 2009). In these instances pre- and in-service teachers gather evidence of classroom practice, organize them in portfolio entries, and examine and evaluate their work through reflective entries, standards-based practices, and program expectations. Consistent with claims by Cochran-Smith (2009), these portfolios may allow for collections of evidence to depict multi-faceted elements of practice as opposed to relying on single assessments. They may also foster longitudinal examinations through sustained collection and analysis (Anderson & Friesen, 2004; Heinrich, Bhattacharya, & Rayuda, 2007).

Limitations of Portfolio Evidence

Yet portfolios have limitations. Researchers describe needs for

teacher training and support to complete these assessments (Fallon & Watts, 2001; Shepherd & Hannafin, 2008). Additionally, little research has examined the validity of portfolio findings or students' ability to collect and organize evidence that justifies conclusions.

Practicality. Effective evidence collection and examination require dedicated time and planning (Fallon & Watts, 2001; Hadley, 2007; Kjaer, Maagaard, & Wied, 2006). Prior to evidence collection, teachers must identify portfolio purposes as well as tangible artifacts that accurately represent those purposes. Teachers gather these artifacts while simultaneously attending to lesson objectives, student needs, and other teaching responsibilities. Once artifacts are gathered, they must be organized and presented. This often requires written descriptions identifying included artifacts and articulating purposes. Next, teachers examine evidence, draw conclusions, and develop action plans for improvement (Recesso et al., 2009). To facilitate evidence collection and analysis, many programs turn to direct coaching supplemented with portfolio question prompts (Strudler & Wetzel, 2005; Wray, 2007; Zepeda, 2002). Others develop detailed guidelines for artifact collection and examination (NBPTS, 2010; Silver et al., 2009).

Despite personal coaching and detailed guidelines, little research focuses on the practicality of sustained evidence collection among teachers. Indeed, Burroughs (2001) questioned the validity of National Board Certification because teachers were asked to reflect and write in ways that were not taught or expected of the profession. Although summative methods exist to capture and examine teaching practice, more research is needed on the feasibility of methods for sustained, formative inquiry among teachers.

Validity and reliability. Ultimately portfolios are containers that house and disseminate evidence collections. Thus, individual and programmatic purposes influence what artifact collection and examination methods are deemed relevant. Large variability in portfolio practice exists. Wade and Yarbrough (1996), Zeichner and Wray (2001), and Zepeda (2002) argue that researchers must explain portfolio purposes, artifact selection and organization procedures, degree of creative freedom, provided supports, and assessment criteria to clarify outcomes and compare approaches. Identifying procedures for artifact selection and examination helps researchers to compare intentions, rules, and standards applied to individual practice and their influence on portfolio outcomes (Darling, 2001). Yet few studies focus on evidence selection and examination.

The extent that artifacts justify portfolio outcomes is questionable. While examining six measures of validity among 128 preservice teacher

portfolios, Yao et al. (2008) found that artifacts matched portfolio purposes and standards but were under-represented and based primarily on reflective entries. Reflective statements, as opposed to tangible classroom evidence, drove inquiry and development. Deladshere and Arens (2003) reported similar findings. Portfolio evaluators from three universities were so unclear as to how preservice teacher artifacts documented portfolio purposes that they relied on personal recollections to "fill in the blanks…and confirm that the students [were] 'meeting the standards'" (p. 69). Ultimately retrospective reflection, as opposed to tangible evidence, directed assessment decisions.

Although researchers ask for greater details regarding artifact selection and examination, few studies provide guidance to obtain these details. When teachers and mentors reflect on memories to examine practice, they may misrepresent events, misinterpret included artifacts, and misdirect inquiry and assessment. Although Naizer (1997) and Burns and Haight (2005) found that portfolio success was a good indicator of course grades, strategic and pedagogical knowledge, and prior teaching experience, Tillema and Smith (2007) concluded that portfolio assessment was often confounded by assessor expectations regardless of program objectives and grading criteria. It is questionable whether principles of evidence credibility and relevance, instrumental to event reconstruction and argument formation, are adequately taught to teachers during portfolio practices.

Training. Training in evidence selection is necessary to facilitate artifact collection and assessment. Several researchers indicate that training is required to ensure program purposes and coordinate faculty, support, and assessment needs (Delandshere & Arens, 2003; Strudler & Wetzel, 2005; Zepeda, 2002). Wray (2007) stated that regular face-to-face communication was needed to mentor preservice teachers on portfolio purposes, artifact collection, and event reconstructions. West et al. (2009) made similar claims regarding in-service teachers and mentors. As teachers examined and discussed portfolio entries together and with trained faculty, they clarified evidence composition, focused reflective statements, and explored outcomes in greater detail. Portfolio implementers cannot assume that teachers in any stage of their career will know how to select appropriate evidence to examine and build argument chains and develop plans for professional growth. These practices must be developed over time through consistent training, particularly if teachers are to assume evidence collection, organization, and analysis roles in their own classrooms for purposes of professional development.

Although portfolios may allow teachers to depict and examine profes-

sional practice through formative and summative assessments, guidance is needed regarding the quantity and type of evidence required to form valid and reliable depictions that meet quantitative rigor advocated by policy makers while retaining rich descriptions of teacher practice. We believe that these requirements may be met by turning to single-case design methods and legal evidence collection techniques.

Single-Case Methods

Single-case methods were developed for health and behavioral scientists working with small populations that cannot be randomly assigned to treatment and control groups or meet sufficient sample sizes to use common statistical analyses. Regardless of approach, all single-case methods require users to establish and compare a baseline with subsequent interventions while ruling out alternative hypotheses (Kratochwill et al., 2010; Morgan & Morgan, 2009). Because evidence is collected to establish both baseline and treatment conditions, guidelines exist regarding evidence quantity required to demonstrate these patterns. This section briefly describes three single-case methods (reversal, multiple baseline, and changing criteria designs) and articulates how they may be used to guide portfolio practices.

Reversal design. Reversal designs begin by examining baseline behavior. Once a baseline pattern is established through evidence collection, a treatment is introduced and more data collected. The intervention is then removed and reintroduced for comparison purposes to complete at minimum four phases (e.g., baseline, intervention, withdrawal, intervention). Patterns are established by collecting evidence at least three times per phase (Kratochwill et al., 2010).

If teachers are interested in examining how question prompts influence discussion participation through portfolio development, for example, they may collect baseline data over three or more discussions that identify discussion length, number of student participants, extent of student-teacher versus student-student participation, and so forth. Once a pattern is identified, teachers can introduce an intervention believed to improve discussion quality (e.g., asking more open-ended questions, providing greater wait time to allow thought organization) and again collect data on discussion length and participation. To determine if discussion changes are based on the intervention, they remove it and collect additional data in a withdrawal phase. Teachers finally reintroduce the intervention and compare discussion length and participation across phases by visually comparing data across phases.

Multiple baseline. However, inquiry-based portfolios may focus on interventions believed to yield permanent outcomes (e.g., teaching algorithms to simplify mathematics principles, identifying states and capitals, describing the citric acid cycle). These activities may not allow for intervention withdrawal because students continue to use principles previously learned. Multiple baseline studies are useful in these situations. Multiple baseline studies introduce a single change but stagger it across time, settings, and units (Morgan & Morgan, 2009, p. 127).

Thus, a teacher interested in whether an intervention, like teaching an algorithm to solve a math problem, affects student behavior or achievement might identify multiple students or classes with similar demographics and behaviors, gather baseline data from each unit, and stagger the introduction of the intervention among them to compare the change in each group. Kratochwill et al. (2010) explain that multiple baseline studies must include a minimum of six staggered phases with at least three evidence collection points per phase.

Changing criterion. Changing criterion models are also used when behavior is expected to change permanently but at a slow rate (Hartmann & Hall, 1977; Morgan & Morgan, 2009). A baseline is established through evidence collection, but the intervention is given in a series of incremental phases. Thus, the previous treatment phase becomes a new baseline. According to Kratochwill et al. (2010) these designs must have a minimum of four phases (including the baseline) with three evidence collection points per phase.

Because single-case methods rely on small sample sizes, they are easier to enact among practicing teachers examining classroom phenomena. These methods also allow teachers to define the unit of analysis. They may focus one inquiry on individual students and another on groups or classrooms, providing flexibility in scope and analysis of professional practice. Additionally, single-case methods introduce artifact selection and organization guidance that may be missing from portfolio implementations. Guidelines for evidence quantities are provided as well as organization approaches to evaluate them.

Single-case methods are also easier to analyze than many other statistical procedures. Data are plotted on graphs and visually inspected (Morgan & Morgan, 2009). Kratochwill et al. (2010) described four steps that occur during visual inspection. First, reviewers graph results of each phase and examine baseline data to determine if the a pattern is present. Next, baseline and intervention phases are visually compared on mean scores within phases, linear trends, and data fluctuations. Third, the degree of overlap between phases and immediacy of intervention changes

is examined. Finally, data from all phases are combined to determine if three demonstrations of an intervention effect are visually documented. Because single-case methods rely on visual inspection, they may ease training requirements among teachers who are often not proficient in advanced research methods.

Although single-case methods describe evidence collection and organization strategies to yield potentially causal claims, teachers must still recognize and select quality evidence represented in the same metric as other sources for comparison purposes. Additionally, causal claims with single-case methods can only be made when alternative hypotheses are ruled out. Thus, teachers must be taught how to examine and refute alternative claims as they collect and examine evidence. Turning to legal frameworks on evidence selection may help teachers make these decisions.

Legal Evidence Characteristics

According to Schum (1994), evidence cannot completely reconstruct events. Regardless of source (e.g., proficiency testing, video recordings, witness testimony, statements of intent), there are always facets that remain ambiguous, unexplored, or unattainable. Video recordings, for example, are lauded for their ability to depict teaching moments. However, they may fail to capture details, context, and student and teacher decisions (Hatch & Grossman, 2009; Sherin & van Es, 2009). Despite these limitations, evidence collection is fundamental to court-based practices. Using chains of evidence to support or refute claims, experts attempt to recreate past events that are analyzed and interpreted by laypersons to pass judgment. Because teachers use artifacts to reconstruct and examine practice, similar selection and organization criteria may help them to construct arguments and refute alternative claims with evidence.

Evidence Organization

When a possible crime is committed, evidence is sought to recreate and examine events surrounding it. Prosecutors and defense attorneys select and organize evidence to argue particular claims and refute alternative hypotheses. Generally, main claims cannot be established through singular pieces of evidence. Rather, legal experts use chains of evidence to support assumptions that lead to main claims (Schum, 1994). Assumptions leading to main arguments and the evidence supporting or refuting these assumptions can be organized into chains that depict intentions and claims.

Similar argument and assumption chains may clarify and drive

artifact collection and organization. A teacher inquiring into a student's mathematics ability, for example, may examine a previous assessment. This action includes many assumptions about the assessment that may not be accurate. Examining these assumptions (e.g., score accuracy, assessment representation, instructional quality, etc.) requires additional evidence. Once assumptions are identified, evidence regarding their accuracy and alternative perspectives can be sought and explored, resulting in robust arguments (see Figure 1). While these examinations appear to increase the evidence quantity, according to Schum (1994), argument strength is inversely related to argument length. Argument chains that require fewer pieces of evidence are stronger than those that require larger quantities. Thus, teachers not only focus on assumptions but also on selecting non-redundant evidence to justify claims.

Evidence Quality

Regardless of organization approaches, individuals must also consider evidence quality. Schum (1994) suggests that evidence quality is based on its credibility. Aspects of source, proximity, and structure influence credibility. Additionally, relevance influences evidence quality and subsequent inclusion in portfolio entries and evaluations.

Source. Evidence source must be examined to ascertain intent, authority, reliability, and bias (Cochran-Smith, 2006; Schum, 1994). Are evidence sources authoritative and trustworthy? What agendas do these sources have in relation to intended assessments? Might these agendas distort or bias supplied information? Are claims verifiable, accurate, valid, and reliable? When collecting and reviewing teacher evidence, individuals should establish whether included items represent typical work or best practice. Evidence given most weight accurately describes events, comes from authoritative sources, and articulates or reduces bias.

Proximity. Evidence proximity (physical and temporal distance from the event being assessed) also influences credibility. Physical proximity to events in question may differentiate evidence quality and relevance. When examining classroom practice, standardized exam scores that demonstrate content mastery may be less credible than observation notes from subject matter experts because the latter are physically present, may identify how content knowledge is represented in the learning environment, and may consider how other intervening variables influence classroom practice.

Temporal proximity also influences evidence credibility. Schum (1994, p. 75) articulates temporal proximity using three classifications: prospectant, concomitant, and retrospectant. Prospectant evidence

Figure 1
Organizational Chains of Assumptions and Evidence
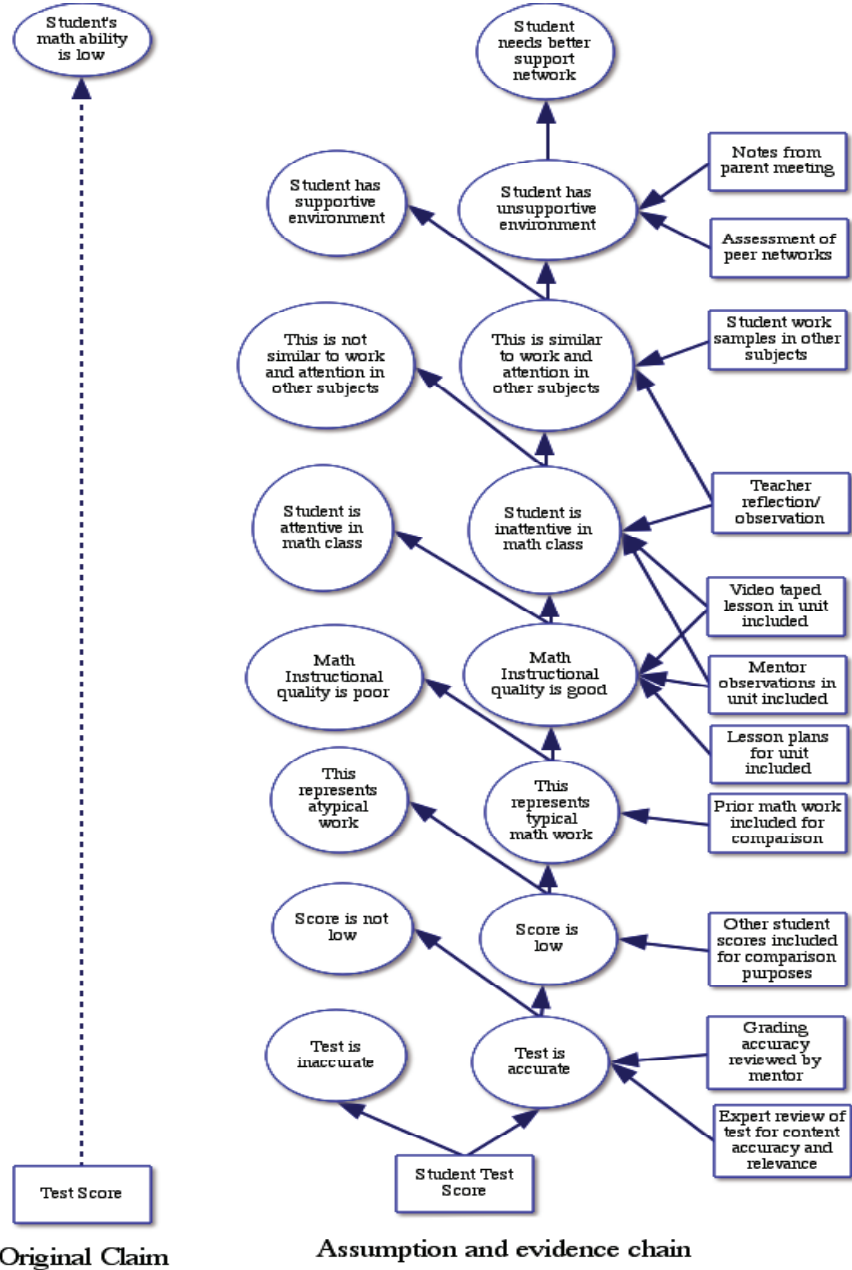
documents incidents prior to events in question. Concomitant evidence occurs during events, and retrospectant evidence occurs after events. Holding other variables constant, concomitant evidence caries the greatest weight in event reconstruction. Thus, lesson plans hold less weight in classroom reconstruction than a video recording of the event in question because the video partially captures the event, whereas the lesson plan only articulates what was meant to happen. Even exam scores may fail to represent classroom events as directly as observer notes because they are more distanced from lesson implementation and may fail to account for relevant variables including student fatigue, absence, lack of engagement, test anxiety, and so forth.

However, prospectant and retrospectant evidence may still be instrumental in establishing arguments regarding intent, preparation, prior and subsequent content knowledge, methods, instructional outcomes, retention, comparison, and perspective—particularly when they occur temporally close to the event in question. Comparing artifacts from previous lessons (e.g., lesson plans, student work samples, observation notes), may facilitate arguments regarding classroom norms and student ability, whereas exams may document student understanding and retention. Although policy makers often focus on student achievement as the ultimate measure of teacher effectiveness, others counter that the art and science of teaching are more complicated (Cochran-Smith, 2006). Articulating events of interest determines what evidence will capture them best. If student achievement is deemed the event for examination, exam scores may become concomitant, with lesson implementations, student characteristics, and other evidence sources playing a supportive role.

Structure. Evidence structure also affects credibility. People experience events through varied means. Some events are experienced directly; others are expressed through testimony, reenactment, models, visual representations, and symbols. Dale (1969) placed direct, firsthand experience as the foundation from which all other experiences are interpreted. Above direct experience are increasing levels of abstraction, including observation, graphic depiction, and symbolic communication (Dale, 1969). Greater levels of abstraction require increased interpretation that is subject to personal, social, political, and other influences. These factors increase potential for miscommunication, misrepresentation, and bias. Schum (1994) articulated levels of abstraction in evidence type. "Real evidence" (p. 94) is made up of physical objects from the event in question. This evidence is tangible; it can be observed, handled, directly experienced, and interpreted. Demonstrative evidence (e.g., models, maps, diagrams) is also observable but represents a degree of abstraction

because it depicts an individual's interpretation of the event as opposed to the event itself. Testimonial evidence is also abstracted because event reconstructions rely on selective experience interpreted, reconstructed, and communicated via language systems.

Direct classroom experience is difficult to capture. Although video recordings and photographs show portions of events, they are limited to operator focus and may miss fine classroom details, student and teacher behaviors, and instructional context (Rich & Hannafin, 2008; Shepherd & Hannafin, 2009; Sherin & van Es, 2009). Teachers and students directly experience events of interest but articulate them through written feedback, assignments, and reflective practices. These sources are distanced from actual events and abstracted in presentation format. Written language is used to interpret events that can no longer be reviewed directly. Users must ascertain the veracity of this evidence, their sources, and interpretations. If work samples created during events are collected, individuals must establish why these samples were chosen, what kind of work they represent (typical, advanced, etc.), whether they can be generalized to other students in the class, and how they support or refute claims.

Relevance. Regardless of credibility, evidence must relate to events in question. If selected evidence is not relevant to assessment purposes, its credibility is moot. Schum (1994, pp. 87, 88) indicated that relevance is also established by indicating how evidence supports asserted arguments and refutes alternative claims. The unit of analysis is important to determine artifact relevance. If the unit in question is an entire class, for example, providing student work samples from one or two students may not represent that unit. Mean scores or class scores may provide more relevant evidence of practice. If, however, the unit of analysis is an individual student or group of students, individual homework examples and test scores may be relevant.

Considering evidence by its relevance and credibility to the events in question may empower teachers as they reconstruct practice for professional development and assessment purposes. Using evidence policies established through single-case methods and legal frameworks may allow teachers to examine and take control of their public presentation. Sweeping claims regarding teacher quality or student achievement based on national mean scores and public assumption can be examined through local intervening variables supported with evidence chains. Additionally, these practices may be more approachable to practicing teachers with primary interests in student instruction and learning.

Discussion

Although portfolios are developed for several purposes, they all include elements of assessment. Despite claims of increased skills development, facilitated reflection, and professional development, few implementations document how portfolios influence these outcomes. After decades of portfolio use among pre and in-service teachers and years of research criticizing lack of validity and reliability in portfolio claims (Burns & Haight, 2005; Tillema & Smith, 2007; Zeichner & Wray, 2001), few researchers have considered how to select, organize, and examine artifacts to provide valid measures of practice or document them in such a way that they may be replicated by others. This article provides a framework for artifact collection that attempts to bridge previous limitations by relying on flexible foci, validated methods, data-driven decisions, and documentation of problem solving processes for purposes of professional development. However, these methods have limitations. Portfolio processes already require extensive coaching and support. Additional training to learn these methods may be necessary. Additionally, research is needed regarding effective evidence chains to support given purposes.

Validity

Few studies articulate how teachers collect, organize, and examine artifacts for assessment purposes (Wade & Yarbrough, 1996; Zeichner & Wray, 2001; Zepeda, 2002). One benefit of this framework is that it relies on validated methods frequently used in social and behavior sciences to support causal claims while furthering evidence examination to consider alternative hypotheses. Thus, researchers can focus on the degree of portfolio adherence to these methods and their subsequent outcomes of practice as opposed to detailed artifact collection strategies and rationales. This approach may also mitigate concerns that turning to longitudinal, quantitative methods will omit important characteristics and variables associated with the teaching profession (Cochran-Smith & the Boston College Evidence Team, 2009; Nazier, 1997). Indeed, argument chains formed through evidence collections may show nuanced representations of teacher practice that consider individual, family, classroom, community, and local norms in addition to content and pedagogy approaches for individual students, groups, and classes.

However, research is needed to establish guidelines regarding evidence chains that best represent teaching activities. Because evidence sources vary in regards to credibility and relevance, guidance is needed to help teachers at all stages of their careers select and organize appropriate sources to best represent classroom practice. According to the

Family Educational Rights and Privacy Act, schools must obtain permission from legal guardians or students to disclose information from their educational records. Additionally, copyright laws protect student rights to distribute their work. Clarification about what can and cannot be included within teacher-based portfolios needs further clarification.

Flexibility

In addition to providing validated methods for artifact collection that can be linked to causal outcomes, single-case methods also provide flexibility in data collection. Not only are multiple designs available to examine teaching events, but units of analysis can vary between portfolio entries. Similar methods are useful for examining single individuals as well as small groups, whole classrooms, fifth grade teachers in the building, and entire schools.

This flexibility allows teachers to focus on diverse aspects of their practice without having to master multiple research methods. Additionally, the visual nature of data analysis may be more approachable and manageable to practicing teachers than more complex quantitative and qualitative methods. However, few portfolio studies have examined the use of single-case methods in practice. More research is needed regarding teachers' uses of these methods during portfolio tasks and their attitudes towards them.

Empowerment

Policy leaders advocate the use of data to drive teacher decisions and to provide individualized supports. Although current measures of teacher quality are often tied to standardized testing and student achievement (Cochran-Smith, 2005; Hanushek & Rivken, 2010), portfolios may provide additional nuances into teacher and student practice. One of the original purposes of portfolio development was to provide more holistic representations of teacher practice (Bird, 1990; Wolf, 1991). Advocates have long believed that portfolio entries have the potential to depict teacher practice better than standardized tests because they capture variables in addition to student achievement and teacher content knowledge.

Using single-case methods to examine classroom practices and alternative hypotheses not only allows for more descriptive classroom depictions but also for depictions of external variables that may influence learning. Ancillary evidence gathered to examine a student's home life could not only articulate influences affecting student progress but also become the impetus for developing and examining targeted teaching strategies tailored to individual needs and circumstances. This dossier of information could also be used to direct dialogue between parents,

administration, teachers, and students; support classroom decisions and inquiry; and justify actions taken.

Training

Of course, using single-case methods also present many uncertainties to portfolio users. Some researchers question novice teachers' abilities to examine evidence of practice, stating they lack sufficient experience to focus inquiries (Carver & Katz, 2004; Rippon & Martin, 2006). Direct mentoring with experienced others is often advocated by researchers because it helps teachers select and interpret artifacts, consider alternative viewpoints, and justify portfolio claims (Land & Zembal-Saul, 2003; Wolf, Whinery, & Hagerty, 1995; Wray, 2007). Despite simplified data analysis methods when compared with other quantitative approaches, establishing and analyzing single-case studies is difficult.

Kratochwill et al. (2010) stress that clearly defined outcome and treatment variables must be specified prior to data collection as well as alternative hypotheses that may affect design implications. Teachers must also articulate their unit of analysis, ensure that collected data matches that unit, and quantify evidence relating to treatment outcomes for analysis purposes. Although single-case methods provide guidance on artifact inclusion and organization, they do not articulate what chains of evidence reconstruct classroom events nor how those events influence outcomes. Thus, teachers and reviewers must rely on evidence quality characteristics (e.g., relevance, credibility, proximity, abstractedness) when examining portfolio arguments and assumptions. Certainly training is needed to master these tasks. However, research suggests that direct training can be at least partly replaced with electronic tutorials, question prompts, and feedback mechanisms (Bolliger & Shepherd, 2010; Land & Zembal-Saul, 2003; Shepherd & Hannafin, 2011; Whipp, 2003).

Time

Time spent developing portfolio events is another potential limitation. Many researchers discuss the large time commitments portfolio entries require (Fallon & Watts, 2001; Hadley, 2007; Kjaer, Maagaard, & Wied, 2006). It is likely that these time commitments will increase when teachers develop evidence chains through single-case methods because they force users to consider baseline and treatment situations over time, examine evidence quality, and consider alternative hypotheses. Some researchers suggest that teachers ultimately abandon portfolio practices in part because of the time commitments associated with their use (Milman, 2009; Rolheiser & Schwartz, 2001; Shepherd & Hannafin, 2008).

Few schools value teacher portfolio practices by giving them time

and support to examine practice. However, assessment always takes time. Having the potential to directly construct assessment experiences as opposed to relying on others to set national agendas may help the profession to focus on formative and summative assessment approaches as they articulate teacher quality. The current debate about teacher effectiveness will not end any time soon; portfolios may help teachers better control the dialogue.

References

Anderson, T., & Friesen, N. (2004). Interaction for lifelong learning. *British Journal of Educational Technology, 35*, 679-687.

Beck, R., Livne, N. L., & Bear, S. L. (2005). Teachers' self-assessment of the effects of formative and summative electronic portfolios on professional development. *European Journal of Teacher Education, 28*(3), 221-244.

Bird, T. (1990). The schoolteachers portfolio: An essay on possibilities. In J. Millman & L. Darling-Hammond (Eds.), *The new handbook of teacher evaluation assessing elementary and secondary school teachers* (pp. 241-256). Newbury Park, CA: Sage.

Bolliger, D. U., & Shepherd, C. E. (2010). Student perceptions of eportfolio integration in online courses. *Distance Education, 31*(3), 295-314.

Bransford, J., Darling-Hammond, L., & LePage, P. (2005). Introduction. In L. Darling-Hammond & J. Bransford (Eds.), *Preparing teachers for a changing world* (pp. 1-39). San Francisco: Jossey-Bass.

Burns, M. K., & Haight, S. L. (2005). Psychometric properties and instructional utility of assessing special education teacher candidate knowledge with portfolios. *Teacher Education and Special Education, 28*, 185-194.

Carver, C. L., & Katz, D. S. (2004). Teaching at the boundary of acceptable practice what is a new teacher mentor to do? *Journal of Teacher Education, 55*, 449-462.

Cochran-Smith, M., & Boston College Evidence Team. (2009). "Re-culturing" teacher education: Inquiry, evidence, and action. *Journal of Teacher Education, 60*(5), 458-468.

Cochran-Smith, M. (2006). Taking stock in 2006: Evidence, evidence everywhere. *Journal of Teacher Education, 57*(1), 6-12.

Cochran-Smith, M. (2005). The new teacher education: For better or for worse. *Educational Researcher, 34*(7), 3-17.

Dale, E. (1969). *Audiovisual methods in teaching* (3rd ed.). New York: Dryden Press.

Darling, L. F. (2001). Portfolio as practice: The narratives of emerging teachers. *Teaching and Teacher Education, 17*, 107-121.

Delandshere, G., & Arens, S. A. (2003). Examining the quality of the evidence in preservice teacher portfolios. *Journal of Teacher Education, 54*, 57-73.

Dhonau, S., & McAlpine, D. (2005). An electronic portfolio for the ACTFL/NCATE teacher education program standards in the second language methods course. *Foreign Language Annals, 38*(1), 69-76.

Fallon, D. (2006). The buffalo upon the chimneypiece: The value of evidence. *Journal of Teacher Education, 57*(2), 139-154.

Fallon, M. A., & Watts, E. (2001). Portfolio assessment and use: Navigating uncharted territory. *Teacher Education and Special Education, 24*(1), 50-57.

Grossman, P. (2005). Research on pedagogical approaches in teacher education. In M. Cochran-Smith & K. M. Keichner (Eds.), *Studying teacher education the report of the AERA panel on research and teacher education* (pp. 425-476). Mahwah, NJ: Lawrence Erlbaum Associates.

Habib, L., & Wittek, L. (2007). The portfolio as artifact and actor. *Mind, Culture, and Activity, 14*(2), 266-282.

Hadley, N. J. (2007). The portfolio forum: Power in reflection. *Journal of Educational Technology Systems, 35*(4), 449-455.

Hallman, H. L. (2007). Negotiating teacher identity: Exploring the use of electronic teaching portfolios with preservice English teachers. *Journal of Adolescent and Adult Literacy, 50*(6), 474-485.

Hanushek, E. A., & Rivkin, S. G. (2006). Teacher quality. In E. A. Hanushek & F. Welch (Eds.), *Handbook of the economics of education* (Vol. 2, pp. 1051-1078). Amsterdam, North Holland: Elsevier.

Hanushek, E. A., & Rivkin, S. G. (2010). Using value-added measures of teacher quality.

Hartmann, C. (2004). Using teacher portfolios to enrich the methods course experiences of prospective mathematics teachers. *School Science and Mathematics, 104*, 392-407.

Hartmann, D. P., & Hall, R. V. (1976). The changing criterion design. *Journal of Applied Behavior Analysis, 9*(4), 527-532.

Hatch, T., & Grossman, P. (2009). Learning to look beyond the boundaries of representation: Using technology to examine teaching. *Journal of Teacher Education, 60*(1), 70-86.

Heinrich, E., Bhattacharya, M., & Rayuda, R. (2007). Preparation for lifelong learning using eportfolios. *Journal of Engineering Education, 32*, 653-663.

Kelley, L. M. (2004). Why induction matters. *Journal of Teacher Education, 55*, 438-448.

Kjaer, N.K., Maargaard, R., & Wied, S. (2006). Using an online portfolio in postgraduate training. *Medical Teacher, 28*(8), 708-712.

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). *Single-case designs technical documentation*. Retrieved from What Works Clearinghouse website: http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf.

Land, S. M., & Zembal-Saul, C. (2003). Scaffolding reflection and articulation of scientific explanations in a data-rich, project-based learning environment: An investigation of progress portfolio. *Educational Technology Research and Development, 51*(4), 65-84.

Milman, N. B. (2009). Web-based digital teaching portfolios: What happens after they graduate? In P. Adamy & N. B. Milman (Eds.), *Evaluating electronic portfolios in teacher education* (pp. 91-109). Charlotte, NC: Information Age.

Morgan, D. L., & Morgan, R. K. (2009). *Single-case research methods for the behavioral and health sciences*. Los Angeles, CA: Sage.

National Board for Professional Teaching Standards. (2010). *The portfolio*. Retrieved from http://www.nbpts.org/for_candidates/the_portfolio

Naizer, G. L. (1997). Validity and reliability issues of performance-portfolio assessment. *Action in Teacher Education, 18*(4), 1-9.

Orland-Barak, L. (2005). Portfolios as evidence of reflective practice: What remains 'untold.' *Educational Research, 47*(1), 25-44.

Recesso, A., Hannafin, M., Wang, F., Deaton, B., Rich, P., & Shepherd, C. E. (2009). Direct evidence and the continuous evolution of teacher practice. In P. Adamy & N. B. Milman (Eds.), *Evaluating electronic portfolios in teacher education* (pp. 23-47). Charlotte, NC: Information Age.

Redish, T., Webb, L., & Jiang, B. (2006). Design and implementation of a web-based portfolio for aspiring educational leaders: A comprehensive, evidence-based model. *Journal of Educational Technology Systems, 34*(3), 283-295.

Rich, P. J., & Hannafin, M. J. (2008). Capturing and assessing evidence of student teacher inquiry: A case study. *Teaching and Teacher Education, 24*(6), 1426-1440.

Rickards, W. H., Diez, M. E., Ehley, L., Guilbault, L. F., Loacher, G., Hart, J. R., & Smith, P. C. (2008). Learning, reflection, and electronic portfolios: Stepping toward an assessment practice. *The Journal of General Education, 57*(1), 31-50.

Rippon, J. H., & Martin, M. (2006). What makes a good induction supporter? *Teaching and Teacher Education, 22*(1), 84-99.

Rolheiser, C., & Schwartz, S. (2001). Pre-service portfolios: A base for professional growth. *Canadian Journal of Education, 26*(3), 283-300.

Schum, D. A. (1994). *The evidential foundations of probabilistic reasoning*. Evanston, IL: Northwestern University Press.

Shepherd, C. E., & Hannafin, M. (2011). Supporting preservice teacher inquiry with electronic portfolios. *Journal of Technology and Teacher Education, 19*(2), 189-207.

Shepherd, C. E., & Hannafin, M. (2009). Beyond recollection: Reexamining preservice teacher practices using structured evidence, analysis, and reflection. *Journal of Technology and Teacher Education, 17*(2), 229-251.

Shepherd, C. E., & Hannafin, M. (2008). Examining preservice teacher inquiry through video-based, formative assessment e-portfolios. *Journal of Computing in Teacher Education, 25*(1), 31-37.

Sherin, M. G., & van Es, E. A. (2009). Effects of video club participation on teachers' professional vision. *Journal of Teacher Education, 60*(1), 20-37.

Silver, E. A., Mesa, V. M., Morris, K. A., Star, J. R., & Benken, B. M. (2009). Teaching mathematics for understanding: An analysis of lessons submitted by teachers seeking NBPTS certification. *American Educational Research Journal, 46*(2), 501-531.

Strudler, N., & Wetzel, K. (2005). The diffusion of electronic portfolios in teacher education: Issues of initiation and practice. *Journal of Research on Technology in Education, 37*, 411-433.

Tillema, H., & Smith, K. (2007). Portfolio appraisal: In search of criteria. *Teaching and Teacher Education, 23*, 442-456.

Wade, R. C., & Yarbrough, D. B. (1996). Portfolios: A tool for reflective thinking

in teacher education? *Teaching and Teacher Education, 12*(1), 63-79.

West, R. E., Rich, P., Shepherd, C. E., Recesso, A., & Hannafin, M. (2009). Supporting induction teachers' development using performance-based video evidence. *Journal of Technology and Teacher Education, 17*(3), 369-391.

Wetzel, K., & Strudler, N. (2005). The diffusion of electronic portfolios in teacher education: Next steps and recommendations from accomplished users. *Journal of Research on Technology in Education, 38*, 231-243.

Whipp, J. L. (2003). Scaffolding critical reflection in online discussions helping prospective teachers think deeply about field experiences in urban schools. *Journal of Teacher Education, 54*, 321-333.

Wolf, K. (1991, October). The schoolteacher's portfolio: Issues in design, implementation, and evaluation. *Phi Delta Kappan, 73*, 129-136.

Wolf, K., Whinery, B., & Hagerty, P. (1995). Teaching portfolios and portfolio conversations for teacher educators and teachers. *Action in Teacher Education, 17*(1), 30-39.

Wray, S. (2007). Teaching portfolios, community, and pre-service teachers' professional development. *Teaching and Teacher Education, 23*, 1139-1152.

Yao, Y., Thomas, M., Nickens, N., Downing, J. A., Burkett, R. S., & Lamson, S. (2008). Validity evidence of an electronic portfolio for preservice teachers. *Educational Measurement: Issues and Practice, 27*(1), 10-24.

Zeichner, K., & Wray, S. (2001). The teaching portfolio in US teacher education programs: What we know and what we need to know. *Teaching and Teacher Education, 17*, 613-621.

Zepeda, S. J. (2002). Linking portfolio development to clinical supervision: A case study. *Journal of Curriculum and Supervision, 18*(1), 83-102.

Zumwalt, K., & Craig, E. (2005). Teachers' characteristics: Research on the indicators of quality. In M. Cochran-Smith & K. M. Keichner (Eds.), *Studying teacher education: The report of the AERA panel on research and teacher education* (pp. 157-260). Mahwah, NJ: Lawrence Erlbaum Associates.