

Toward a Unified Validation Framework in Mixed Methods Research

Amy B. Dellinger

University of New Orleans, Louisiana

Nancy L. Leech

University of Colorado at Denver and Health Sciences Center

The primary purpose of this article is to further discussions of validity in mixed methods research by introducing a validation framework to guide thinking about validity in this area. To justify the use of this framework, the authors discuss traditional terminology and validity criteria for quantitative and qualitative research, as well as present recently published validity terminology for mixed methods research. The authors discuss the rationale for their framework and how it unifies thinking about validity in mixed methods research. Finally, they discuss how the framework can be used.

Keywords: *validity; construct validity; construct validation; credibility; research methods; mixed methods research; design; measurement; research quality; validity criteria*

Today, researchers can choose from which perspective to investigate phenomenon: a qualitative perspective (Denzin & Lincoln, 2005), a quantitative perspective (Shadish, Cook, & Campbell, 2002), or a mixed methods combination of the two perspectives (Tashakkori & Teddlie, 1998). With these choices of perspectives or paradigms, especially with the recent popularity of mixed methods research (Johnson & Onwuegbuzie, 2004), researchers in the social and behavioral sciences confront a large number of terms for evaluating and describing validity of a research study. The number and meanings of the available terms to describe whether research results, their interpretation, and their use are valid can be overwhelming, especially for a novice researcher.

Given the vast number of terms for validity and the multiple ways of defining validity, the primary purpose of this article is to further discussions of validity in mixed methods research. To do so, we introduce a validation framework (VF) to unify thinking about validity in this area. To justify the use of this framework, we discuss traditional terminology and validity criteria for quantitative and qualitative research, as well as present recently published validity terminology for mixed methods research. We discuss how our framework unifies thinking about validity in both qualitative and quantitative traditions. Finally, the framework is described in terms of how it can be used by mixed methods researchers.

We admit that the use of the terms *qualitative research* and *quantitative research* grossly oversimplifies what are rich and complex traditions, ideas, approaches, and techniques of research. However, we use these terms globally to indicate research that emanates from a single perspective or set of related techniques. Additionally, given our understanding

that the choice of research methodology or perspective is not a characteristic of an individual, two other terms used in this article, *qualitative researcher* and *quantitative researcher*, are defined as a researcher who chooses to use a qualitative approach to do research and a researcher who chooses to use a quantitative approach to do research, respectively.

The VF incorporates traditions of both qualitative and quantitative research as well as emerging validity terminology from mixed methods research. To assist the reader in understanding the framework, we briefly review validity in both traditions and explore works that led us to believe that a single VF is justified and useful for mixed methods research. Our discussion begins with a review of important milestones in thought about validity as it relates to single-method research.

What Was, and Is, Validity?

Definitions of validity have evolved over the years and differ for quantitative and qualitative research. Debates in both areas continue to explore what validity means and how it is determined. Presently, quantitative and qualitative researchers tend to treat issues of validity differently, even though terms used by both may be similar in name and/or function.

Validity in Quantitative Research

For quantitative research, an early definition of validity was offered by Garrett (1937) as “the extent to which it [scores] measures what it purports to measure” (quoted in Osterlind, 2006, p. 93). This definition is inadequate to reflect a modern approach to validity (Messick, 1995). Unfortunately, this definition, and its derivations, are commonly used to define construct validity, one of the names used for measurement-related validity in quantitative research. The term *construct validity* figures prominently in the discussion of validity in general and specifically in the VF discussed later in this article. Construct validity began as one of three different types of validity: content, criterion, and construct. Each of these is discussed below.

Historically, the idea of validity in quantitative research was presented as separate types of validity (Cronbach & Meehl, 1955). Cronbach and Meehl (1955) defined three types of validation procedures: (a) content validity; (b) criterion-related validity, which subsumed predictive validity and concurrent validity; and (c) construct validity. In the past, this three-part set of validity procedures could be followed to validate an instrument or measurement process. As with most techniques in quantitative methods, the procedures to achieve validity were outlined in detail to ensure standardization.

To assess content validity, a researcher was interested in making sure that a specific domain of functioning was represented by the items on a test (Crocker & Algina, 1986). Often, experts were used to judge items on a measurement instrument in terms of the specified domain being studied. Criterion-related validity was demonstrated when scores from a measure correlated appropriately in hypothesized ways with other constructs (concurrent validity) or were useful in predicting future scores representing hypothetically related constructs (predictive validity; Cronbach & Meehl, 1955). The final type of

validity, construct validity, was said to overlap with the other two types. Construct validity required researchers to demonstrate that instruments measured the constructs they were designed to measure. This was accomplished through multiple studies that focused on the internal structure of the data from the measures and their consistency with the theories they were hypothesized to represent. However, a common misconception is that the three types of validity are like three strikes; if you do not get a “hit” with one type, try one of the other three (Pedhazur & Schmelkin, 1991).

Campbell and Stanley (1963) extended the idea of validity to include design-related issues. Specifically, these researchers described two types of validity, internal and external validity. Furthermore, they defined procedures that could be incorporated into study designs to control for threats to design-related validity. Internal validity could be established by controlling for particular threats that might supply alternative explanations for studied outcomes. External validity, on the other hand, reflected the degree to which study results could generalize to different places or persons.

The triumvirate of measurement-related validity terms (i.e., *content*, *criterion*, and *construct*) can still be found in research methods and measurement textbooks. Yet most texts now recognize Cronbach and Meehl (1955) and Messick (1989a, 1989b, 1995) for explicating a view of validity that placed construct validity as the center or overarching idea of validity. Thus, construct validity encompasses all types of measurement-related validity evidence. Additionally, Messick’s model of construct validation unified all validity evidence. Therefore, construct validity includes not only measurement-related validity but also all other validity evidence, such as design-related validity evidence and statistical inference validity evidence (which assesses the appropriateness of statistically derived inferences).

According to Messick (1989a, 1989b, 1995), all traditional validity evidence accrues to the meaning of measures, or construct validity (regardless of whether test scores, observations, attitudinal assessments, etc.), not to the tests or instruments themselves. In addition, nontraditional validity evidence, such as evidence of utility and interpretation and the consequences of use, was defined as critical issues used to support the meaning of data. Although Messick used the term *test score* in the works cited above, he maintained that

The term “test score” is used generically here in its broadest sense to mean any observed consistency, not just on tests as ordinarily conceived but also on any means of observing or documenting consistent behaviors or attributes. This would include, for instance, any coding or summarization of observed consistencies on performance tests, questionnaires, observation procedures, or other assessment devices. This general usage also subsumes qualitative as well as quantitative summaries and applies, for example, to protocols, clinical interpretations, and computerized verbal score reports. (Messick, 1989a, p. 5)

Whether the intent was to be inclusive of qualitative research in its purest sense is unknown. Can the ideas associated with construct validity be applied to qualitative data and its meanings, inferences, and consequences of those inferences? Political scientists (Adcock & Collier, 2001) developed a model of measurement validity that supposedly explicated a shared standard of validity for qualitative and quantitative research. However, the reading of this article, and some of Messick’s examples, seemed to apply more to categorical data than qualitative, open-ended, narrative data. The applicability to both

quantitative and qualitative data and research will be explored further in a later section, because this discussion bears on the utility of the VF presented here.

Others (e.g., Gliner & Morgan, 2000; Krathwohl, 1993; Onwuegbuzie, 2003) have expanded on the ideas of internal and external validity, the validity of statistical inferences, and measurement-related validity. For example, Gliner and Morgan (2000) described measurement-related validity, design-related validity, and statistical inference validity as leading to judgments about the “Overall Research Validity of the Study” (p. 377). This approach certainly demonstrates consistency with the view of validity as a unified judgment, yet it does not consider construct validation as the ultimate goal of research processes and quality. We recommend several excellent reviews of validity issues, standards, and practices in the area of quantitative research to extend the reader’s knowledge (e.g., Goodwin & Leech, 2003; Hubley & Zumbo, 1996; Jonson & Plake, 1998; Messick, 1995).

Validity in Qualitative Research

Validity is a concept that traditionally has been viewed as quantitative (Campbell, 1957; Campbell & Stanley, 1963). Yet validity is also a contentious issue for qualitative researchers (Onwuegbuzie & Leech, 2007). Generally, qualitative researchers perceive validity (e.g., truth value, credibility, dependability, trustworthiness, generalizability, legitimation, authenticity) as being an unclear and ambiguous concept. In fact, as many as 17 terms for validity have been identified in the qualitative realm (Lather, 1993; Lincoln & Guba, 2005; Maxwell, 1992). To date, not only are there multiple words to describe validity in qualitative studies, there is not an agreed-on definition of validity in qualitative research. In response to the indistinct nature of validity in qualitative research, four approaches to validity in qualitative research, associated with positivists, postpositivists, poststructuralists, and postmodernists, have emerged (Onwuegbuzie & Leech, 2007).

Qualitative researchers who work from a positivist perspective interpret validity using the same criteria as are used in quantitative research. For example, Miles and Huberman (1984) referred to validity with terms such as *internal validity* and *external validity*. Similar to the positivist perspective, qualitative researchers who work from a postpositivistic viewpoint consider validity with comparable concepts as their positivistic counterparts. The difference for postpositivists is their belief that validity should be interpreted in relation to its ability to generate theory, be internally reflexive, be empirically based, and produce generalizable findings (Denzin, 1994).

Poststructuralists believe that a new set of criteria should be developed to explicate validity in qualitative research, being driven by politics and incorporating “subjectivity, emotionality, feeling, and other antifoundational criteria” (Denzin, 1994, p. 298). Furthering this, Lather (1986) presented the idea of catalytic validity, which she defined as the degree to which a research study empowers and liberates a research community. Lather also defined four other types of validity: ironic, paralogical, rhizomatic, and voluptuous legitimation. Ironic legitimation assumes that there are multiple realities of the same phenomenon; therefore, identifying the truth depends on the ability to review coexisting opposites. Paradoxes in the data are placed under paralogical legitimation. Rhizomatic legitimation focuses on going beyond describing the data to mapping the information.

Finally, voluptuous legitimation identifies how much a researcher's level of interpretation goes beyond his or her knowledge.

Some postmodernists do not believe that criteria can be developed for understanding validity. For example, Richardson (1994) believes that all methods are suspect; in fact, no method can reveal the truth. Furthermore, Wolcott (1990) questioned whether validity is appropriate for qualitative research. Others, for example critical theorists, want criteria, but want the criteria to be very different from the quantitative view of validity. Schwandt (1996) suggested three focuses for organizing our thoughts regarding validity: (a) generating knowledge to supplement "lay probing of social problems" (p. 69), (b) enhancing critical intelligence, defined as "the capacity to engage in moral critique" (p. 69), and (c) impacting philosophy.

On the basis of the controversy stemming from the multiple perspectives on validity, Lincoln and Guba (2005) concluded,

At issue here is not whether we shall have criteria, or whose criteria we as a scientific community might adopt, but rather what the nature of social inquiry ought to be, whether it ought to undergo a transformation, and what might be the basis for criteria within a projected transformation. (p. 206)

These authors argued that there is not a concrete answer for validity and how it should be assessed in qualitative research. Instead, we can have discussions regarding "what we might use to make both professional and lay judgments regarding any piece of work" (p. 207).

Over the past few decades, many researchers have participated in these discussions. Lincoln and Guba (1985) suggested the need to develop an entirely different set of criteria to assess validity than what are traditionally used by quantitative researchers. These theorists developed the concepts of trustworthiness, which corresponds with Campbell and Stanley's (1963) concepts of internal validity, and applicability, which is similar to Campbell and Stanley's definition of external validity. Eisner (1991) took this one step further by not using the word *validity* but instead used the word *credibility*. Maxwell (1992) identified five types of validity: descriptive validity, interpretive validity, theoretical validity, generalizability, and evaluative validity. Eisenhart and Howe (1992) advocated for a unitary concept of validity in which general standards for conducting qualitative research should be used as guidelines. Finally, Richardson (1994, 1997) and Richardson and St. Pierre (2005) proposed the crystalline form of validity; the text, meaning, human currents, and truth are metaphorically identified in the crystal, light, light waves, and energy, respectively.

To help further the discussion regarding validity in qualitative research, researchers have attempted to assess and define validity for qualitative studies. For example, Goetz and LeCompte (1984) used quantitative definitions of validity with qualitative vernacular (Eisenhart & Howe, 1992). Furthermore, Denzin (1989) used Campbell's (1963a, 1963b) threats to validity to interpret the robustness of seven types of studies (experiments, surveys, participants observations, unobtrusive methods, life histories, interviewing, and filming). Finally, Greenwood and Levin (2005) offered a succinct definition of validity in qualitative research:

Validity . . . [is] measured by the willingness of local stakeholders to act on the results . . . thereby risking their welfare on the “validity” of their ideas and the degree to which the outcomes meet their expectations. Thus, cogenerated contextual knowledge is deemed valid if it generates warrants for action. The core validity claim centers on the workability of the actual social change activity engaged in, and the test is whether or not the actual solution to a problem arrived at solves the problem. (p. 54)

Many definitions of the various aspects of validity in qualitative research specifically refer to the how-tos of establishing credibility, authenticity, trustworthiness, criticality, and integrity, to name a few. For example, Whitemore, Chase, and Mandle (2001) called these overarching ideas *primary criteria*. Their other term, *secondary criteria*, refers to important and flexible aspects of quality criteria that are in addition to the primary criteria. Secondary criteria include explicitness, vividness, creativity, thoroughness, congruence, and sensitivity. Together, relevant parts of the primary and secondary criteria are accomplished through the use of techniques that apply to design considerations, data generation, analysis, and presentation.

Recently, Maxwell (2004a, 2004b) identified three strategies for understanding causal relationships in qualitative studies. Maxwell (2004b) stated,

Developing causal explanations in a qualitative study is not, however, an easy or straightforward task. Furthermore, there are many potential validity threats to any causal explanation, threats that will need to be addressed in the design and conduct of a study. In this, the situation of qualitative research is no different from that of quantitative research; both approaches need to identify and deal with the plausible validity threats to any proposed causal explanation. (p. 250)

Maxwell outlined three groups of strategies, namely, those associated with (a) the quantitative field but also appropriate for qualitative research (i.e., intervention and comparison), (b) the “direct observation or indirect identification of causal processes” (p. 251) (i.e., intensive long-term involvement, rich data, and narrative and connecting analysis), and (c) developing alternative explanations (i.e., the modus operandi approach, discrepant evidence and negative cases, triangulation, and member checks). Using these methods to assist in understanding causal relationships will increase the chances of having valid data.

In summary, qualitative research has viewed validity from many perspectives. These differing views have created multiple terms for validity, as well as discussion of what criteria, if any, are needed to assess the validity of data. The vast array of choices for describing validity in qualitative studies, as well as the controversies over the types of validity, has created a difficult chasm to navigate for qualitative researchers.

Validity in Mixed Methods Research

The concept of validity has been addressed sparingly in the mixed methods literature. Unfortunately, it has not been thoroughly discussed or widely considered. Early treatments of validity for mixed methods studies tended to assess the quantitative and qualitative parts of studies separately (e.g., Krathwohl, 1993; Newman & Benz, 1998). Since then, there have been attempts to address validity issues related to the separate parts of a

mixed methods study as well as the validity issues arising from the mixing of methods in a study.

Recently, Creswell and Plano Clark (2007) suggested that issues of validity in mixed methods research still need to be addressed. Specific questions posed by these authors included the following:

How should validity be conceptualized in mixed methods research? What does mixed methods validity look like from a pragmatic viewpoint? How does this viewpoint differ from post-positivist, constructivist, and emancipatory perspectives? What special validity issues are raised by specific types of designs? (p. 190)

Thus, the concept of validity has yet to be delineated for mixed methods research.

In the *Handbook of Mixed Methods in Social & Behavioral Research* (Tashakkori & Teddlie, 2003a), there are 15 different sections listed in the index, none of which clearly presents validity information. For students and beginning researchers, this can be confusing, because there is not one section in the book devoted to validity in mixed methods research. To confuse matters even more, Teddlie and Tashakkori (2003) stated that *validity* “has become a catchall term that is increasingly losing its ability to connote anything. When a term is used with other words to connote so many meanings, it essentially has none” (p. 36). Their suggestion was to use the term *inference quality* in mixed methods studies (p. 37). According to Teddlie and Tashakkori, *inference quality* is a combination of design quality, which can be defined as whether a study adheres to best practice, and interpretive rigor, which is defined as how well the results can be trusted.

Onwuegbuzie and Johnson (2006) saw a need to compliment Tashakkori and Teddlie’s (2006) treatment of validity in mixed methods research. A typology of mixed methods legitimation was offered by Onwuegbuzie and Johnson. The term *legitimation*, as opposed to *validity*, was used because of the connection of the word *validity* to the quantitative research paradigm. The term *legitimation* has appeal because of its neutrality, having been used only to describe validity criteria for qualitative research (Onwuegbuzie & Johnson, 2006). The types of legitimation include sample integration legitimation, inside–outside legitimation, weakness minimization legitimation, sequential legitimation, conversion legitimation, paradigmatic mixing legitimation, commensurability legitimation, multiple validities legitimation, and political legitimation. Onwuegbuzie and Johnson’s model focuses on how well a researcher has integrated or blended the various design and inference aspects of a mixed methods study, therefore specifically addressing the validity of the qualitative and quantitative segments of a mixed methods study with multiple validities legitimation.

Later, Tashakkori and Teddlie (in press) expanded their treatment of inferential quality to be an overall assessment of validity in mixed methods research. This model suggests that criteria specific to the approach, quantitative or qualitative, should be used at each step of the study. In addition, design quality, the first component of inferential quality, and interpretive rigor, the second component of inferential quality, are assessed by examining specific criteria having to do with the mixing of methods. According to Tashakkori and Teddlie, design quality is indicated by the following: (a) design suitability, (b) design adequacy or fidelity, (c) within-design consistency, and (d) analytic adequacy. Interpretive

rigor consists of (a) interpretive consistency, (b) theoretical consistency, (c) interpretive agreement, (d) interpretive distinctiveness, and (e) integrative efficacy.

Each of the two models described above (Onwuegbuzie & Johnson, 2006; Tashakkori & Teddlie, in press) provides meaningful criteria for assessing issues associated with mixed methods research and the inferences associated with that research. Both models seem to defer to current norms or criteria of validity traditionally used in quantitative and qualitative research for the separate parts of mixed methods studies.

Even with the models developed by Onwuegbuzie and Johnson (2006) and Tashakkori and Teddlie (in press), mixed methods research needs further discussion of validity (Creswell & Plano Clark, 2007). In this article, we present a framework for guiding discourse about validity in mixed methods research. Our aim is to influence thinking about validity in mixed methods studies. We do not intend to introduce another set of terms for quantitative, qualitative, or mixed methods research, because there is already an abundance of terms related to validity. For more information on these terms, refer to Onwuegbuzie and Johnson and Tashakkori and Teddlie, who have made excellent strides in addressing the array of criteria and questions needed to advance a validity argument. We do intend to make progress toward unifying discussions of validity under a meaningful framework.

To do so, we depend heavily on the idea of construct validation as the all-encompassing purpose of having standards for any element of a study. To extend and apply Messick's (1995) conception of construct validity to build our framework, some background information is provided, followed by a discussion of other relevant works by different authors.

Messick's Construct Validity Revisited

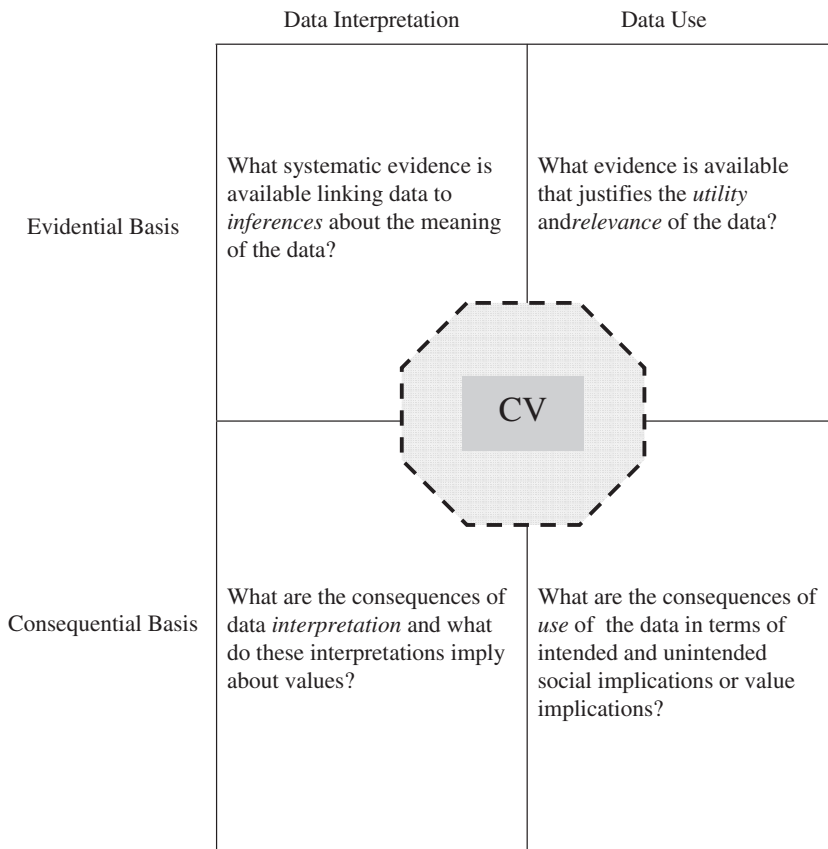
Messick's (1995) conception of construct validity as encompassing all validity evidence provides the unifying theme for our framework. According to Messick, validity is defined as

an overall evaluative judgment of the degree to which empirical evidence and theoretical rationale support the adequacy and appropriateness of interpretations and actions on the basis of test scores [data] and other modes of assessment. . . . The principles of validity apply not just to interpretive and action inferences derived from test scores as ordinarily conceived but also to inferences based on any means of observing or documenting consistent behavior or attributes. . . . This general usage subsumes qualitative as well as quantitative summaries. (p. 741)

Yet Messick's definition is too narrow when including qualitative data from the many types of qualitative research traditions. Thus, the definition of validity could be rewritten as an overall evaluative judgment of the extent to which empirical evidence and/or theoretical rationale support the adequacy and appropriateness of interpretations and actions on the basis of data generated through any means. Using *Webster's New Universal Unabridged Dictionary* (1983), a definition for construct validation can be created as an idea or perception resulting from the orderly arrangement of facts and/or impressions that is able to withstand criticism or objection through the process of argumentation.

Messick (1995) argued that assessment of the meaning of measures (whether quantitatively or qualitatively generated) is a continuous process that involves value judgments situated in social structures and varying across time. Meaning making, or the meaning of

Figure 1
Questions Asked in Construct Validation (CV)



Source: Messick (1995).

measures, which is described as construct validity, accrues from traditional sources of validity evidence and nontraditional sources of validity evidence (e.g., the use and relevance of measures, value implications of measure interpretation, consequences of the use of measures). Figure 1 is provided as a means to understand the elements of construct validation. In this figure, Messick juxtaposed data interpretation and data use with the evidential basis of data meaning and the consequential basis of data meaning. For example, the following questions become important: (a) What empirical evidence is available (perhaps via raw data, coding criteria, theoretical rationales, member checks, statistical analyses, etc.) that links data to the meaning of the data? (b) What evidence is available that justifies the utility and relevance of the data (e.g., evidence of use, value to a research community)? (c) What are the consequences of data interpretation (whether appropriate inferences were made, the consequences of the inferences, the values inherent in the choices made, and the meaning this accrues to the data)? and (d) What are the consequences of the appropriate use of the data in terms of intended and unintended social implications or value implications?

Markus (1998) expressed concern about Messick's (1995) stated understanding of construct validity. Markus pointed out that "the very idea of construct validity rests on constructivism . . . saves validity theory from obsolescence. Without it validity theory could not adapt to new problems" (p. 10). Markus's argument with Messick's (1995) theory stems from concern that if both realist and constructivist approaches are used in the validation of data meaning, then the arguments used to present the evidence can never be completed. Markus's critique equated realist and constructivist approaches with mutually exclusive value independent and value dependent arguments. Messick (1998) countered by agreeing with Markus that the synthesis was incomplete and that the argument was "rather left as a tension to be negotiated in validation practice" (p. 37). This tension and the negotiations associated with it are central to the discussion of construct validation in mixed methods research.

Construct Validity in Mixed Methods Research

Angen (2000) noted that "the same negotiation, acceptance of ambiguity, and reliance on dialogue that are required in all our myriad daily interpretations of meaning are also required for the validation of research" (p. 392). This is because, as Ercikan and Roth (2006) stated,

(a) all phenomena are quantitative and qualitative at the same time; and (b) data construction processes follow similar interpretation processes for all education research; and (c) for most constructs that education researchers are interested in, these data construction processes are based on subjective, defensible judgments. (p. 18)

This statement does not mean that qualitative approaches and quantitative approaches used to measure a similar construct are measuring the same things in the same way. Nor do the two approaches accept the same perspectives about the existence of a construct. As a result of these differences, researchers who use mixed methods approaches may experience some tensions in how a phenomenon under study is approached, described, and measured. Furthermore, there might be tension with how the ultimate understandings of those measures within their study are conceptualized.

Understanding of a phenomenon or culture in a qualitative study signifies wading in the data, whether voices, images, observations, videos, text, and so on. It also indicates finding the systematic meanings of the data and how those meanings are interwoven. The meanings developed from the data are socially constructed. A researcher may state a specific phenomenon or construct of interest at the outset of the design, but meanings of the data, as they exist for the study participants, are unknown. Usually in a qualitative study, it is not just the description of the construct that is important but the how and why of the construct. A researcher's conceptual definition, unformed or weakly formed at the outset, takes form, value, or meaning iteratively during the study as the participants share their ways of knowing or their lived experiences. From this process, a construct such as "family support" might be negotiated by the researcher (and the participants) through the meanings of the data.

Quantitative research begins differently than qualitative research. Quantitative research starts with a supposition that a specific construct or entity exists. The construct is

traditionally explicitly defined so that a measure (e.g., a test, survey items, observations, performance assessment) can be developed that is capable of capturing the essence of the defined construct. For example, if family support is the construct of interest, it must be defined explicitly, and some measurement process of generating scores must be created. Once a measuring device is created, strict criteria are used to determine if the data are reliable and that the scores are valid reflections of the defined construct.

In quantitative research, researchers base a construct definition on past research, theory, or phenomena. On the other hand, qualitative researchers may use the literature but traditionally begin with self-created, and hopefully acknowledged, beliefs about the phenomena of interest that change and develop as studies progress. Qualitative researchers' understanding of a construct emerges from immersion into the data. In qualitative research, the construct is not limited but developed and refined in a flexible, organic way.

Quantitative and qualitative research approaches measure constructs differently or create meaning from data differently, but this does not mean that they cannot measure, or are not measuring, the same constructs or meanings on the basis of the systematic arrangement of perceptions, experiences, attitudes, and so on. Yet neither are they completely different. As with quantitative research processes, according to Freeman, deMarrais, Preissle, Roulston, and St. Pierre (2007),

qualitative data and information are always already interpretations made by participants as they answer questions or by researchers as they write up their observations. Neither research participants nor researchers can be neutral, because, as emphasized earlier, they are always positioned culturally, historically, and theoretically. (p. 27)

In tying quantitative and qualitative research processes together, Ercikan and Roth (2006) stated,

The products (e.g., articles, representations) of both forms of research inherently re-present (aspects of) human experience. Both are abstractions although of different scales, rather than constituting human experience itself. Irrespective of the type of research, both require an initial observational translation—that is, a reduction or abstraction that is prelogical and preconscious and leads to data. (p. 21)

These two types of understanding of the meaning of constructs require researchers and society to negotiate, dialogically, the meaning of the data within and across studies regardless of approach (Angen, 2000). As Cherryholmes (1988) stated, “A critical approach to construct validity points out that one must take one’s eye off a reading score itself and look around it if one is to understand which theoretical constructs the reading score indicates” (p. 437).

Mixed methods research can capitalize on the juxtaposition of quantitative and qualitative ways of knowing, the data generated from these approaches, and the categories or constructs negotiated from the research. Onwuegbuzie and Leech (2005) pointed out that “as noted by Dzurec and Abraham (1993), meaning is not a function of the type of data collected (i.e. quantitative vs. qualitative). Rather, meaning results from the interpretation of data, whether represented by numbers or by words” (p. 379).

Our conceptualization of validity depends on the negotiation of data meaning. Given the nature of mixed methods approaches, our framework uses the tension inherent in the research process (Mason, 2006) to develop (or further) discourse on the meaning of data. This tension is caused when, as human beings, we think we know something one way (as in the construct we begin with) but open ourselves to other ideas about knowing, thus turning it this way and that, checking its fit with what we came to the study understanding, and developing new meanings and/or knowledge (Beach, Becker, & Kennedy, 2006). As Mason (2006) succinctly stated, “Explanations do not have to be internally consensual and neatly consistent to have meaning and to have the capacity to explain. Indeed, if the social world is multi-dimensional, then surely our explanations need be likewise?” (p. 20).

How does a mixed methods researcher do this well, working with the categories or constructs one knows and simultaneously being open to new meanings? Not easily, according to Langer (1989), who warned about traps of our own categories:

The creation of new categories . . . is a mindful activity. Mindlessness sets in when we rely too rigidly on categories and distinctions created in the past (masculine/feminine, old/young, success/failure). Once distinctions are created, they take on a life of their own. Consider: (1) First there was earth. (2) Then there was land, sea, and sky. (3) Then there were countries. (4) Then there was Germany. (5) Now there is East Germany versus West Germany. The categories we make gather momentum and are very hard to overthrow. We build our own and our shared realities and then we become victims of them—blind to the fact that they are constructs, ideas. (p. 11)

Avoiding the traps of mindlessness described by Langer requires researchers to take a mindful approach to the research process, thereby transcending the rote application of procedures and being aware of meanings that we miss when we approach a problem from only a single direction with a checklist of criteria.

Giving ourselves permission to appreciate and use multiple forms of evidence and to integrate others’ research and inferences and the varied meanings found in them allows us to make judgments about the meaning of data on the basis of its usefulness and interpretations and the consequences of these uses and interpretations. Cherryholmes (1988) put it best:

Construct validation is critical and pragmatic. It is pragmatic because it is concerned with giving persuasive (pragmatically convincing) accounts of what is going on. Whether an argument or story about construct validity and research is persuasive or not is determined against the context and background of communities of researchers, communities under investigation, and possible alternative communities. Constructs and conceptions mediate our experiences in descriptions, metaphors, models, and explanations. We do not have unmediated access to “brute” reality; our versions of things are, for better or worse, more or less adequate for our purposes. They invite criticism. Each approach to discourse and construct validity reviewed contributes information and perspective. But each is partial. From time to time one may desire or require only partial information, such as that provided by mainstream, ethnographic, or critical research, but partiality should not be mistaken for impartiality, incompleteness for completeness, bias for neutrality. (p. 451)

We propose that construct validation is the continuous process of negotiation of meaning. This is accomplished through argument as dialogue, criticism, and objection. These

must be based on empirical evidence, utility, inferential value, and inferential consequences, as to the meaning of data resulting from the orderly arrangement of facts, impressions, or perceptions. We propose that the framework we have developed, the VF, can be used to guide discourse about the meaning of measures. Therefore, it is not a definitive means of proving that data are valid or have a particular meaning or, for that matter, proving that a study is worthwhile. We believe that the application of construct validation is, as Messick (1998), Markus (1998), and Cherryholmes (1988) implied, a process that is not a closed system, a complete synthesis, or a necessarily convergent process. Instead, construct validation is an open, continuous system in which construct meaning is the product of convergent and divergent evidence, results, consequences, and arguments from all research related to that construct, whether qualitative or quantitative. As Moss and Schutz (2001) noted in their discussion of educational standards and assessments, consensus is not always desired or useful; rather, dissensus can be “a natural resource that should be acknowledged and nurtured alongside the search for consensus or agreement” (p. 258). An open system puts both qualitative and quantitative research at the same table to negotiate the meaning of data. As Maxwell (1992) noted,

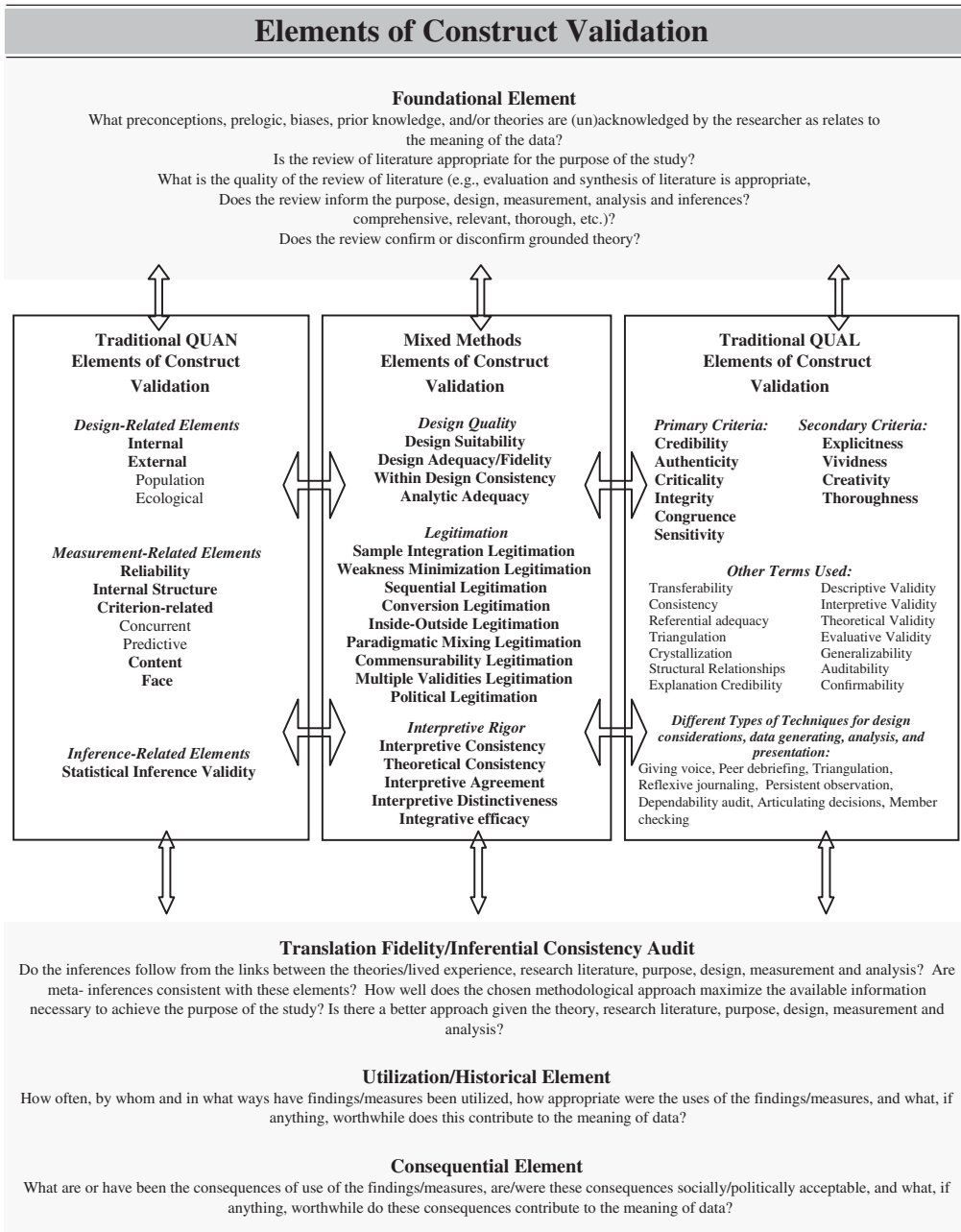
I agree with Mishler that validity is always relative to, and dependent on, some community of inquirers on whose perspective the account is based. Validity is relative in this sense because understanding is relative: as argued above, it is not always possible to challenge an account to be independent of any particular perspective. It is always possible to challenge an account from outside that community and perspective, but such a challenge amounts to *expanding* [italics added] the community that is concerned with the account and may change the nature of the validity issues in ways to be discussed below. (p. 284)

Because of the openness of this system, we believe that construct validity and the construct validation process are relevant to mixed methods research. The application of construct validation to mixed methods research is appropriate because (a) construct validation is a pragmatic process (Cherryholmes, 1988); (b) construct validation requires mindfulness (Langer, 1989), recognition, and the constant integration of relevant available information, whether qualitative(ly) or quantitative(ly); (c) construct validation assumes that there are no criteria or rulers to measure research quality except as determined by discourse and language in a community of researchers (Cherryholmes, 1988) and the subsequent weight of the research in the negotiations of data meaning; and (d) construct validation is a nonending, continuous, time- and context-specific, open process (Markus, 1998; Messick, 1995, 1998).

A VF for Mixed Methods Research

The VF presents a useful and unified method to frame the idea of validity in mixed methods research and to provide a guide for organizing the necessary evidence needed to support data meanings (see Figure 2). The framework uses traditional concepts about validity, including the terminology and criteria from the quantitative and qualitative traditions reviewed previously in this article, and the newer ideas in mixed methods research. We did, however, identify one area, and thus one term, that was missing from

Figure 2
Validation Framework: Elements of Construct Validation



Note: This is not a comprehensive list of terms. Some terms in the figure are adapted from various sources, including Messick (1995), Gliner and Morgan (2000), Tashakkori and Teddlie (1998, 2003, 2006), Dellinger (2005), Newman and Benz (1998), Whittemore et al. (2001), Onwuegbuzie and Johnson (2006), and Maxwell (1992).

the quantitative, qualitative, and mixed methods validity terminology. Our new term, *foundational element*, as well as other terms used in the framework that were not discussed previously, are included in the following description of our VF.

Traditional approaches to validity for quantitative and qualitative study elements are in the heart of the figure, at either side of the newer terminology for mixed methods research. Other elements of validity evidence, centered in the figure, apply to qualitative, quantitative, and mixed methods studies. The terms discussed here and the sources cited are by no means comprehensive but represent a few interesting and/or influential methods for thinking about validity. The reader is strongly encouraged to explore the subtleties and mechanics of thinking about validity in the various areas (quantitative, qualitative, and mixed methods) in regard to the VF.

The VF includes many terms used to describe various elements or facets of validity in the literature. Much of the figure presents elements covered in the brief reviews of traditional validity evidence under quantitative and qualitative research and in the section on validity in mixed methods research, in which new terms introduced by Tashakkori and Teddlie (in press) and Onwuegbuzie and Johnson (2006) were presented. The terms that were covered previously are in the unshaded portion of Figure 2. Mixed methods criteria are in the center of the unshaded area, while qualitative criteria are to the right and quantitative criteria are to the left. Elements of the VF that have not been mentioned previously are in the shaded areas and are described here in detail.

Foundational Element

The *foundational element* reflects researchers' prior understanding of a construct and/or phenomenon under study. This element was needed because it is not explicitly addressed in either the qualitative or quantitative literature as being important to providing evidence of construct validity (Dellinger, 2005). Additionally, it is not explicitly addressed in the newer validity criteria for mixed methods studies. Evidence of researchers' understanding of the phenomena or constructs under study includes such things as reflections on personal understanding and experiences, theoretical understanding, and understanding gained through analysis and evaluation of research related to the construct or phenomenon. Prior understandings are important because they influence a researcher's "perception (for example in observations), the choice or development of the methods used, and thereby the data collected and the understanding of the issue" (Steinke, 2004, p. 187). Furthermore, it can bias attitudes in favor of previously held theories (Lord, Ross, & Lepper, 1979). According to Beach et al. (2006),

it is impossible to construct conclusions without relying on prior knowledge. At the same time, if the meaning of research findings is dictated solely by prior knowledge, then we cannot learn from new studies. The problem that researchers face, therefore, is not one of how to put aside prior knowledge but rather one of how to capitalize on prior knowledge and use it to extract as much *new knowledge* as possible from the findings. (p. 502)

To gain understanding of research related to the construct or phenomenon under study, a researcher completes a comprehensive and critical review of the theoretical and empirical literature in the area studied (usually, but not always, at the outset of the study). This

would be true in most quantitatively oriented studies and many qualitatively oriented studies. In the case of some qualitative research and some applied quantitative research, the foundation of the study might be author-perceived or participant-perceived needs, with the literature framing and/or confirming the presentation of results and inferences (Creswell, 2003).

The review of literature is a study in itself, with its own issues of validity (see Cooper, 1998). The results of a review of literature are important to establish credibility (Neuman, 2004) and to provide evidence in the construct validation process (Dellinger, 2005). Reviews of the literature in this vein do not simply reiterate study findings in a body of literature but rather critically, meaningfully, and purposefully situate the study, its data, and the findings in relevant contexts (other research, social, or personal contexts) through reasoning (Krathwohl, 1993). Boote and Beile (2005) stated,

As the foundation of any research project, the literature review should accomplish several important objectives. It sets the broad context of the study, clearly demarcates what is and what is not within the scope of the investigation, and justifies those decisions. It also situates an existing literature in a broader scholarly and historical context. It should not only report the claims made in the existing literature but also examine critically the research methods used to better understand whether the claims are warranted. Such an examination of the literature enables the author to distinguish what has been learned and accomplished in the area of study and what still needs to be learned and accomplished. Moreover, this type of review allows the author not only to summarize the existing literature but also to synthesize it in a way that permits a new perspective. Thus a good literature review is the basis of both theoretical and methodological sophistication, thereby improving the quality and usefulness of subsequent research. (p. 4)

An adequate evaluation and synthesis of the extant literature requires using established and/or defensible criteria to evaluate the selected studies and theoretical literature. For a researcher conducting a study, this step in research development provides initial evidence of construct validity, or the meaning of the measures analyzed and interpreted in the study (Dellinger, 2005). This is particularly important when the information obtained in the review of literature has informed relevant elements of the research process. Examples include (a) confirming or disconfirming grounded theory; (b) guiding the development of the purpose or research questions in a study; (c) influencing the choice of theoretical framework, design, measurement process, and analytical methods; and (d) framing inferences and final arguments.

Inferential Consistency

Inferential consistency, one of the terms not previously discussed, refers to whether the inferences in a study are consistent given what is known from prior understandings, past research, and theory. Furthermore, inferential consistency focuses on what is appropriate given the study design, measurement, and analysis. A research evaluator would assess whether the methodology used in the study under review was adequate to maximize the amount of information available to answer the selected question and/or purpose as opposed to an alternative methodology. Although there are no set rules as to the relative

importance or usefulness of various research methodologies or perspectives (Miller, 2003), the literature review and the purpose or questions of a study can be instrumental in deciding whether the best and/or most information was obtained in a given study. Additionally, meta-inferences should be reviewed for consistency with the preliminary inferences, the extant literature, and study design and implementation (Tashakkori & Teddlie, 2003b). Inferential consistency is similar to Tashakkori and Teddlie's (in press) category of interpretive rigor under the mixed methods validity criteria.

Utilization/Historical Element

Messick (1995) cited utility and evidence of use as sources of construct validity evidence. This evidence is called *utilization/historical element*. Supposedly, the more a measure or study inferences are used to represent or provide meaning to a construct, the more evidence that the measure or study inferences are an acceptable way of representing the construct. Dellinger (2005) found that use can have a negative impact on construct validation when the use of data or inferences from a study in the relevant literature provides invalid or inaccurate evidence. Utilization or historical validity evidence accrues to a study's inferences, measures, or findings because of use (appropriate or not) in the extant literature or in other applications, such as decision making or policy development. A judgment as to the appropriateness of the use of measures and inferences can be made at this stage. Utilization or historical validity, while providing evidence of construct validity through the use of data or inferences in the extant literature, should not be accrued at face value. It must be assessed continuously over time and across contexts on the basis of current evaluative standards of the research community (Dellinger, 2005).

Consequential Element

The *consequential element* of construct validation, the final element in the VF, is determined by judging the social acceptability of consequences that occur as a result of using a study's findings, measures, or inferences. As with the utilization or historical element of construct validation, consequential validity is determined through continuing review of the relevant literature and examination of uses of data or inferences to assess the consequences of use. Consequential validity cannot be assessed by a researcher conducting a study or by subsequent researchers examining only the original study's results. Subsequent use of the measures, results, and inferences of a study should be evaluated for adequacy and social acceptability by the user. For example, high-stakes consequences that result from the use of data or inferences from data need to be evaluated for acceptability by stakeholders.

Using the VF

The purpose of the VF is to influence thinking about validity in mixed methods research. As such, it provides a framework to guide thought in ongoing construct validation processes that consist of well-developed and complex procedures and techniques within quantitative, qualitative, and mixed methods research traditions. The VF does not represent a checklist of validity types or criteria, nor does it reflect a linear, top-to-bottom

process. The arrows in the VF indicate a multipath approach that is dependent on the design of a study.

Using the VF to its full potential would reflect a deep understanding of the complex ideas and procedures previously developed by many thoughtful scholars in the areas of measurement, research design, and methodology, as well as a high level of expertise in any substantive area under investigation. Some practical uses of the VF might include assisting researchers in organizing thoughts about what elements of validity evidence are important for their particular studies during study development, assisting researchers in organizing evidence of construct validation during studies in progress, assisting researchers in reading and making evaluative judgments about construct validation evidence in a single study or a body of literature, and assisting researchers interested in using the findings from a study or body of literature. We will attempt to explain some uses of the VF during the many phases of research.

When designing a study, use the VF to carefully plan to accrue evidence that will support data meanings. A researcher developing a study must make a number of decisions, which have the potential to influence or provide evidence to support construct validation. First, the terms, criteria, and/or related techniques included in the traditional ways of approaching validity for quantitative and qualitative research are not applicable to every quantitative or qualitative study or element of a study. The particular rationale, purpose, research question, design, measurement process, analytical choices, and so on, determine the validation evidence that is important to a study's inferences regarding the meaning of the data. This is a complex process for a unidimensional study and becomes substantially more complex in mixed methods studies. As such, beginning researchers may experience difficulty wading through the many choices and need instruction in selecting the most appropriate elements or techniques that will provide the best construct validation evidence.

For example, a researcher planning a sequential mixed methods study (QUAL → quan) in which the QUAL segment of the study is most prominent would determine which primary and secondary criteria, and subsequent techniques, are necessary to establish credibility or trustworthiness in the qualitative inferences. The researcher would also plan for criteria important to measurement and design-related validity elements in the quantitative portion of the study to support score meaning. Critical methodological issues related to the mixed methods design and inferences should also be considered. If the qualitative portion of the study leads to the development of a measurement process in the quantitative study, the meanings that will emerge from the qualitative segment must be supported with adequate evidence and become part of the evidence required for the quantitative results and inferences.

A researcher planning a study has ample opportunity to gather and organize construct validation evidence. The researcher could document the foundational elements of the study and link these elements with other aspects of the study, including any rationale, purpose, significance, choices made regarding measurement, design, and methodology, as well as inferences that are justified.

In conducting a study, use the VF to analyze the gravity of the validation evidence and make adjustments as needed. A researcher cannot guarantee that the data and inferences in a study being conducted will be viewed by others as important to the meaning of the construct under study or that the data or inferences will be used or valued in the future.

Designing and conducting a study that meets agreed-on standards of quality in a field increases the chances that the study's data and inferences will be influential in defining a construct.

However, a mixed methods researcher would remain flexible, because during the conduct of a study, it may become necessary for the researcher to adjust the type of validity evidence or combine types of evidence in novel, organic ways. In mixed methods and mixed model studies, it may be necessary to blur the dichotomy of quantitative and qualitative validity criteria when assessing design- and measurement-related validity. In other words, depending on the structure of the mixed methods or mixed model study, the researcher may need to use the quantitative criteria and/or the qualitative criteria separately, as in a sequential mixed methods study. In some cases in which there is blending in the various elements of the study, perhaps in a mixed model study, both sets of criteria appropriate to the phase of the study may be essential to adequately address the need for validity evidence.

For example, suppose a researcher is conducting a mixed methods study in which qualitative data were collected and analyzed and the results were then quantitized for use in a quantitative analysis. The inferences and meanings based on these data are dependent on the quality and characteristics of the measurement and analytical processes; therefore, any meanings must be justified by the researcher with construct validation evidence that best demonstrates both the trustworthiness of the qualitative data and analytical results and the credibility of the quantifying process. Evidence of reliability and other validity criteria could assist the researcher in making arguments for the value of the inferences or meaning of the data. Additionally, the appropriate mixed methods criteria outlined in the VF and appropriate to the study design would provide adequate means of organizing additional critical information for construct validation.

In reading and evaluating a study or body of literature, use the VF to assist in evaluating individual study's inferences and to organize thoughts about a body of literature. Not all elements of construct validation are under the control of the original researcher. For example, historical and consequential elements of construct validation are important elements for the consumer of research. A researcher reading and evaluating literature could use the VF to organize information needed to make a judgment about the current value of a study. This evaluation could be in terms of its importance to construct meaning or determine how the study's data and inferences contribute to the meaning of the construct being examined. However, a researcher evaluating studies can use both evidence of use and consequences by iteratively synthesizing evaluative evidence from a body of literature related to a construct.

Traditional validity criteria, including the terms used for design-related and measurement-related validity, were developed specific to either quantitative or qualitative studies. When a primarily quantitative or primarily qualitative study is evaluated using the VF, one would use the appropriate aspect of construct validity that is specific to the perspective of the study. However, mixed methods research is more complex, because the reader must decide whether adequate construct validation evidence is provided for both the quantitative and qualitative elements of the study being evaluated as well as whether the study provides evidence of meeting the criteria set forth by the mixed methods community. In addition to the design- and measurement-related validity evidence, an evaluator would

assess the validity evidence relevant to the foundation element, inferential consistency, the utilization/historical element, and the consequential element.

An overall evaluative statement about the validity of the inferences in the study would follow the analysis of all the aspects of the study. This evaluation is a time-sensitive, interpretive, value-laden judgment, as are individual narratives addressing each element of construct validation. Individuals evaluating a single study may come to different conclusions on the basis of the information and literature available or known to them at the time and the methodological standards valued by the reviewers.

Design-related issues or design elements that affect inferential quality can be assessed through traditional means. However, it is important at this step, and each step thereafter, to assess whether each research element is consistent with the foundation established through the review of literature, through the researchers' lived experiences, or through the participants' needs and is the best choice compared with other design possibilities. This is true as well for measurement-related validity evidence or data quality. One would ask whether the best choice in terms of consistency with the foundational aspect and purpose of the study and with the study context has been made in the measurement of the study phenomenon and in the data-analytic and inferential procedures. This question may be linear in a quantitative study but iterative in a qualitative study and would need to be addressed explicitly in a mixed methods context.

A researcher can also examine historical validity evidence of other studies' measures and inferences during the reading process by determining how they are used in the primary study. The historical validity of the study being evaluated can be assessed by continued reading in the area of study. In this way, a researcher or evaluator determines how an individual study's findings or measures have been used and how often they have been used in the extant literature or in applied settings. Additional insight into the adequacy of the study's findings may be found in later reviews of the literature. Historical validity evidence must include an evaluation of the adequacy and appropriateness of the use of the study's findings or measures as elements in subsequent research studies.

Researchers rarely review single pieces of research literature but rather engage in more comprehensive systematic reviews of the literature, such as in meta-analyses or meta-syntheses. The VF can be useful to researchers interested in doing mixed research synthesis studies (Onwuegbuzie, Collins, Leech, Dellinger, & Jiao, in press; Sandelowski, Volls, & Barroso, 2006). Using the VF as a means to focus quantitative effects and qualitative process information may be useful to development of this newer area of research.

In examining social consequences of the use of a study's inferences, the VF organizes convergent and divergent elements of data meaning to justify the appropriateness (or inappropriateness) of the consequences. Often there are unintended consequences to the use of a study's methods, findings, measures, or inferences. These consequences may be valued or not by stakeholders in the current social structure or research community. These are the criteria used to evaluate the consequential element of construct validation. A study could be well developed and sound, only to have its findings used in such a way (appropriately or inappropriately) that results in negative consequences. These negative consequences in connection with the particular use can accrue validity evidence to the original study's resultant meanings.

Conclusion

The VF takes us beyond the idea that validity is equated with “goodness,” quality, or credibility. Although these are all critically important ideals and are desirable characteristics of research, we would ask, “Goodness to what purpose?” Validity as goodness ignores the purpose of research; quality serves a purpose: to make us pay attention to or value the meanings set forth by the study. Quality is important only in that it supports construct validation or the meaning of our data. It is researchers’ desires to produce meaningful data and inferences through negotiation that makes it natural, practical, and useful, or pragmatic, to use mixed methods approaches.

As Markus (1998) noted, “even the theory of validity must be grounded in particulars of its perspective on validity” (p. 12). The VF presented here obviously attempts to unify thinking about validity and demonstrate its usefulness in mixed methods research, a pragmatic approach to research. It is unclear whether this framework will appeal to those who use only quantitative approaches or those who use only qualitative approaches or even to those who use both, separately and/or together. However, it continues some of the discourse, started by both the quantitative tradition and the qualitative tradition and recently extended by mixed methods researchers. This continued dialogue indicates the possible vision of unifying discussions and debates about the value of divergent and convergent results of measured phenomena or constructs under a large and flexible umbrella of construct validation. The differences, resultant tensions, and dialogue on both sides of the table are what make this possible.

References

- Adcock, R., & Collier, D. (2001). Measurement validity: A shared standard for qualitative and quantitative research. *American Political Science Review*, *95*(3), 529-546.
- Angen, M. J. (2000). Evaluating interpretive inquiry: Reviewing the validity debate and opening the dialogue. *Qualitative Health Research*, *10*(3), 378-395.
- Beach, K. D., Becker, B. J., & Kennedy, M. M. (2006). Constructing conclusions. In C. F. Conrad & R. C. Serlin (Eds.), *The Sage handbook for research in education* (pp. 493-509). Thousand Oaks, CA: Sage.
- Boote, D. N., & Beile, P. (2005). Scholars before researchers: On the centrality of the dissertation literature review in research preparation. *Educational Researcher*, *34*, 3-15.
- Campbell, D. T. (1957). Factors relevant to the validity of experiments in social settings. *Psychological Bulletin*, *54*, 297-312.
- Campbell, D. T. (1963a). From description to experimentation: Interpreting trends as quasi-experiments. In C. W. Harris (Ed.), *Problems in measuring change* (pp. 212-242). Chicago: University of Chicago Press.
- Campbell, D. T. (1963b). Social attitudes and other acquired behavioral dispositions. In S. Koch (Ed.), *Psychology: A study of science: Investigations of man as socius* (Vol. 6, pp. 94-172). New York: Rand McNally.
- Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.
- Cherryholmes, C. H. (1988). Construct validity and the discourses of research. *American Journal of Education*, *96*(3), 421-457.
- Cooper, H. (1998). *Synthesizing research* (3rd ed.). Thousand Oaks, CA: Sage.
- Creswell, J. W. (2003). *Research design: Qualitative, quantitative, and mixed methods approaches*. Thousand Oaks, CA: Sage.

- Creswell, J. W., & Plano Clark, V. L. (2007). *Designing and conducting mixed methods research*. Thousand Oaks, CA: Sage.
- Crocker, L., & Algina, J. (1986). *Introduction to classical & modern test theory*. Fort Worth, TX: Harcourt Brace Jovanovich.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281-302.
- Dellinger, A. B. (2005). Validity and the review of literature. *Research in the Schools*, 12(2), 41-54.
- Denzin, N. K. (1989). *The research act: A theoretical introduction to sociological methods* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.
- Denzin, N. K. (1994). Evaluating qualitative research in the poststructural moment: The lessons James Joyce teaches us. *Qualitative Studies in Education*, 7, 295-308.
- Denzin, N. K., & Lincoln, Y. S. (Eds.). (2005). *The Sage handbook of qualitative research* (3rd ed.). Thousand Oaks, CA: Sage.
- Eisenhart, M. A., & Howe, K. R. (1992). Validity in educational research. In M. D. LeCompte, W. L. Millroy, & J. Preissle (Eds.), *The handbook of qualitative research in education* (pp. 643-680). San Diego, CA: Academic Press.
- Eisner, E. W. (1991). *The enlightened eye: Qualitative inquiry and the enhancement of educational practice*. Upper Saddle River, NJ: Prentice Hall.
- Ercikan, K., & Roth, W. M. (2006). What good is polarizing research into qualitative and quantitative? *Educational Researcher*, 35(5), 14-23.
- Freeman, M., deMarrais, K., Preissle, J., Roulston, K., & St. Pierre, E. A. (2007). Standards of evidence in qualitative research: An incitement to discourse. *Educational Researcher*, 36(1), 25-32.
- Gliner, J. A., & Morgan, G. A. (2000). *Research methods in applied settings: An integrated approach to design and analysis*. Mahwah, NJ: Lawrence Erlbaum.
- Goetz, J. P., & LeCompte, M. D. (1984). *Ethnography and the qualitative design in educational research*. New York: Academic Press.
- Goodwin, L. D., & Leech, N. L. (2003). The meaning of validity in the new standards for educational and psychological testing: Implications for measurement courses. *Measurement and Evaluation in Counseling and Development*, 36, 181-191.
- Greenwood, D. J., & Levin, M. (2005). Reform of the social sciences and of universities through action research. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* (3rd ed., pp. 43-64). Thousand Oaks, CA: Sage.
- Hubley, A. M., & Zumbo, B. D. (1996). A dialectic on validity: Where we have been and where we are going. *Journal of General Psychology*, 123(3), 207-215.
- Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33(7), 14-26.
- Jonson, J. L., & Plake, B. S. (1998). A historical comparison of validity standards and validity practices. *Educational and Psychological Measurement*, 58(5), 736-753.
- Krathwohl, D. R. (1993). *Methods of educational and social science research: An integrated approach*. New York: Longman.
- Langer, E. J. (1989). *Mindfulness*. Reading, MA: Perseus.
- Lather, P. (1986). Issues of validity in openly ideological research: Between a rock and a soft place. *Interchange*, 17, 63-84.
- Lather, P. (1993). Fertile obsession: Validity after poststructuralism. *Sociological Quarterly*, 34, 673-693.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.
- Lincoln, Y. S., & Guba, E. G. (2005). Paradigmatic controversies, contradictions, and emerging confluences. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* (3rd ed., pp. 191-215). Thousand Oaks, CA: Sage.
- Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, 37(11), 2098-2109.
- Markus, K. A. (1998). Science, measurement, and validity: Is completion of Samuel Messick's synthesis possible? *Social Indicators Research*, 45, 7-34.

- Mason, J. (2006). Mixing methods in a qualitatively driven way. *Qualitative Research*, 6(1), 9-25.
- Maxwell, J. A. (1992). Understanding and validity in qualitative research. *Harvard Educational Review*, 62, 279-300.
- Maxwell, J. A. (2004a). Causal explanation, qualitative research, and scientific inquiry in education. *Educational Researcher*, 33, 3-11.
- Maxwell, J. A. (2004b). Using qualitative methods for causal explanation. *Field Methods*, 16(3), 243-264.
- Messick, S. (1989a). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, 18(2), 5-11.
- Messick (1989b). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York: Macmillan.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741-749.
- Messick, S. (1998). Test validity: A matter of consequence. *Social Indicators Research*, 45, 35-44.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage.
- Miller, S. (2003). Impact of mixed methods and design on inference quality. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social & behavioral research* (pp. 423-455). Thousand Oaks, CA: Sage.
- Moss, P. A., & Schutz, A. (2001). Educational standards, assessment, and the search for consensus. *American Educational Research Journal*, 38(1), 37-70.
- Neuman, W. L. (2004). *Basics of social research: Qualitative and quantitative approaches*. Boston: Pearson Education.
- Newman, I., & Benz, C. R. (1998). *Qualitative-quantitative research methodology: Exploring the interactive continuum*. Carbondale: Southern Illinois University Press.
- Onwuegbuzie, A. J. (2003). Expanding the framework of internal and external validity in quantitative research. *Research in the Schools*, 10(1), 71-89.
- Onwuegbuzie, A. J., Collins, K.M.T., Leech, N. L., Dellinger, A. B., & Jiao, Q. G. (in press). A meta framework for conducting and writing rigorous, comprehensive, and insightful literature reviews. In K.M.T. Collins, A. J. Onwuegbuzie, & Q. G. Jiao (Eds.), *Toward a broader understanding of stress and coping: Mixed methods approaches*. Greenwich, CT: Information Age.
- Onwuegbuzie, A. J., & Johnson, R. B. (2006). The validity issue in mixed research. *Research in the Schools*, 13(1), 48-63.
- Onwuegbuzie, A. J., & Leech, N. L. (2005). On becoming a pragmatic researcher: The importance of combining quantitative and qualitative research methodologies. *International Journal of Social Research Methodology: Theory & Practice*, 8(5), 375-387.
- Onwuegbuzie, A. J., & Leech, N. L. (2007). Validity and qualitative research: An oxymoron? *Quality and Quantity: International Journal of Methodology*, 41, 233-249.
- Osterlind, S. J. (2006). *Modern measurement: Theory, principles, and applications of mental appraisal*. Upper Saddle River, NJ: Pearson-Merrill Prentice Hall.
- Pedhazur, E. J., & Schmelkin, L. P. (1991). *Measurement, design, and analysis: An integrated approach*. Hillsdale, NJ: Lawrence Erlbaum.
- Richardson, L. (1994). *Writing: A method of inquiry*. In N. K. Denzin & Y. L. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 516-529). Thousand Oaks, CA: Sage.
- Richardson, L. (1997). *Fields of play: Constructing an academic life*. New Brunswick, NJ: Rutgers University Press.
- Richardson, L., & St. Pierre, E. A. (2005). Writing: A method of inquiry. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* (3rd ed., pp. 959-978). Thousand Oaks, CA: Sage.
- Sandelowski, M., Volls, C. I., & Barraso, J. (2006). Defining and designing mixed research synthesis studies. *Research in the Schools*, 13(1), 29-40.
- Schwandt, T. A. (1996). Farewell to criteriology. *Qualitative Inquiry*, 2, 58-72.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.

- Steinke, I. (2004). Quality criteria in qualitative research. In U. Flick, E. von Kardorff, & I. Steinke (Eds.), *A companion to qualitative research* (pp. 184-190). London: Sage Ltd.
- Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches*. Thousand Oaks, CA: Sage.
- Tashakkori, A., & Teddlie, C. (Eds.). (2003a). *Handbook of mixed methods in social & behavioral research*. Thousand Oaks, CA: Sage.
- Tashakkori, A., & Teddlie, C. (2003b). The past and future of mixed methods research: From data triangulation to mixed model designs. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social & behavioral research* (pp. 671-702). Thousand Oaks, CA: Sage.
- Tashakkori, A., & Teddlie, C. (2006, April). *Validity issues in mixed methods research: Calling for an integrative framework*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.
- Tashakkori, A., & Teddlie, C. (in press). Quality of inferences in mixed methods research. In M. Bergman. (Ed.) *Advances in mixed methods research: Theories and applications*. London: Sage Ltd.
- Teddlie, C., & Tashakkori, A. (2003). Major issues and controversies in the use of mixed methods in the social and behavioral sciences. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social & behavioral research* (pp. 3-50). Thousand Oaks, CA: Sage.
- Webster's new universal unabridged dictionary (2nd ed.). (1983). New York: New World Dictionaries/Simon & Schuster.
- Whittemore, R., Chase, S. K., & Mandle, C. L. (2001). Validity in qualitative research. *Qualitative Health Research, 11*(4), 522-537.
- Wolcott, H. (1990). On seeking—and rejecting—validity in qualitative research. In E. Eisner & A. Peshkin (Eds.), *Qualitative inquiry in education: The continuing debate* (pp. 121-152). New York: Teachers College Press.