6

Data Collection Considerations: Validity, Reliability, and Generalizability

After reading this chapter you should be able to:

- **6.1** Understand the concept of validity as it applies to quantitative research.
- **6.2** Understand the concept of validity as it applies to qualitative research.
- **6.3** Understand the concepts of reliability and generalizability as they apply to action research.

This chapter addresses the importance of validity, reliability, and generalizability as ways to ensure the quality of qualitatively oriented action research.

Improving Student Understanding and Motivation of Multiplication Facts

Alyson Marland

Alyson Marland, a student teacher in a fourth-grade classroom, was a participant in an action research class. Like many elementary teachers, Alyson was challenged about how best to teach elementary students basic number facts while at the same time keeping the children motivated to learn. Alyson's story illustrates the importance of using multiple data sources (qualitative and quantitative) to address issues of validity and reliability that ensured the quality—and robustness—of her action research findings.

This action research project focuses on promoting student understanding and motivation while teaching the basic multiplication facts (0–9). The study examined the effectiveness of teaching methods with an emphasis on rote memorization, compared to those focusing on problem solving. Research advocates the use of problem

solving when introducing and teaching basic facts and holding off on drill and pra tice methods until after students have developed an understanding. The participar in the study consist of 35 fourth-grade students. The students participated in lesso on arrays and multiplication games, and they discussed efficient versus inefficie counting strategies. Data were collected from their old timed tests, state math score interviews, and work sheets. The results of the study suggest a positive relationsh between balancing conceptual understanding and procedural skills and student suc cess with basic facts. The results also show a positive relationship between playin games and student motivation for studying the basic facts.

Introduction

Research suggests that for every time you do something wrong, you have to do i right 17 times before your brain gets used to doing it correctly. These findings are startling when you consider the vast amount of drill and practice methods used in the classroom to teach students their basic math facts. The purpose of my action research project was to seek out and examine effective and efficient methods for teaching the basic facts. My second objective for this project was to develop effective strategies for increasing student motivation in terms of studying their multiplication facts. Students are often frustrated and bored when studying things they don't understand or when using tedious study methods.

Background

According to the 2004–2005 Oregon State Standards, by the end of fourth grade, students are expected to have developed efficient strategies for solving multiplication problems. They should be fluent with these strategies and able to solve all basic fact problems mentally within three seconds. Unfortunately, researchers have found that an alarming number of eighth-grade students still resort to finger counting and other inefficient strategies when solving simple problems (Isaacs, Carroll, & Bell, 2001). These strategies are inefficient because they take too much time and are not done mentally.

One promising practice is to reduce the use of drill and practice methods (Jones, 1995). These methods include timed tests, flash cards, and work sheets with rows of basic facts. One reason researchers advise educators to stay away from drill and practice methods is because they don't aid in students' conceptual understanding of multiplication. These methods provide students with procedural skills that they are taught to mimic. Teaching in this way makes it hard for students to apply multiplication concepts to word problems or real-life situations.

Effective strategies for teaching basic facts include balancing procedural skills and conceptual knowledge. Teaching with balance includes the use of word problems, arrays, and open-ended problem-solving assignments. Isaacs et al. (2001) recommend practicing multiplication in a variety of contexts and situations to increase the students' ability to transfer the skill/concept.

Other promising practices suggest three components to use when teaching basic facts. The first component is developing a strong understanding of the operations of number relationships. The second focuses on trading inefficient strategies for efficient ones, and the third component is providing students with drill and practice assignments. Van de Walle (2003) stresses the importance of not moving to the third step until after the students have developed efficient strategies. If you introduce drill before the students have mastered one or more efficient strategies, the practice will only reinforce the inefficient strategies (counting on fingers, adding the numbers instead of multiplying them, using manipulatives, drawing pictures, etc.). Inefficient methods are inefficient because they take a long time and cannot be done mentally. Efficient methods include skip counting, simplifying the problem, and, ultimately, memorization.

Intervention

My planned intervention to address students' understanding of and motivation to learn basic number facts included the following:

- *Lessons focused on building students' conceptual understanding of the multiplication process*
- *Teaching students efficient strategies for solving basic facts*
- *Introducing students to fun games they could play while studying their basic number facts*

Data Collection

The data collection tools I used were the following:

1. *Students' scores on state math tests*
2. *Results of timed tests*
3. *Informal interviews*
4. *Students' written work*

Data Analysis

The data I collected from the students' state math test was surprising. I was shocked to find that 11 students (about 35 percent of the class) did not meet the state standards in math. The students who did not meet state standards were consistent with the students who were not passing their timed tests. I decided I would need to spend some extra time working with these 11 students.

The data I collected from students' timed tests were by far the most helpful in planning my intervention and understanding where students were having the most trouble. The data I collected from the timed tests focused on the types of errors

students were making. Edelman, Abdit, and Valentin (1995) describe four types of errors humans make when multiplying—operand, table, operation, and nontable errors. Operand errors occur when the incorrect answer given is correct for another problem that shares an operand (e.g., $4 \times 2 = 16$, when 16 is the correct answer for 4×4). Table errors occur when the incorrect answer given does not share an operand with the correct answer but the answer given does reside in the multiplication table (e.g., $6 \times 9 = 56$). Humans make operation errors when they perform a different operation, such as adding or subtracting when solving a multiplication problem (e.g., $9 \times 0 = 9$). The final error is a nontable error, and it occurs when the incorrect answer is not an answer to any problem in the multiplication table (e.g., $5 \times 6 = 31$). Operand errors were by far the most common error made by the students. They accounted for 55 percent of the error total. These data told me that students were associating the incorrect answer with one of the operands, and the early emphasis on drill and practice has reinforced these wrong answers. Nontable errors accounted for 23 percent of the errors. I believe students are making a high amount of nontable errors because they are miscounting on their fingers. Solving 6×7 on your fingers is both hard and confusing, so I assumed students who rely on their fingers make the majority of the table errors. Operation errors occur mostly in problems containing an operand of zero or one. The students often switch to addition and solve the problem by adding zero or one to the other number. This error occurs mostly when the zero or one is on the bottom. I didn't feel the error occurred consistently enough to be considered a problem, so I decided not to focus on it during my intervention.

I also used the students' timed tests to identify which multiplication facts they were having the most trouble with. The problems I found that gave students the most trouble were 6×7, 7×6, 7×4, 7×7, 8×8, 8×7, 6×6, 8×6, and 6×8. In general, though, any problem including an operand of 6, 7, 8, and/or 9 was answered incorrectly by the majority of students. These data were especially helpful in adjusting my interventions because they allowed me to focus on the problems with which students were having the most trouble.

The informal interviews provided me with insight into what the students thought about timed tests, what strategies they used to solve basic fact problems, and how well they could transfer their skills to real-world situations. I was surprised to find that every student answered "yes" to the question about whether they felt timed tests were helpful. The majority of students supported their answer, stating that timed tests gave them a chance to practice their multiplication facts. The interviews also allowed me to discover which strategy each student relied on when solving multiplication facts.

Discussion

The purpose of my action research project was to seek out and examine effective and efficient methods for teaching basic math facts. On completion of my action research project, I feel confident that I will be able to effectively and efficiently teach basic facts to any grade, first through sixth. I am now aware of the common misconceptions regarding rote memorization and the premature use of timed tests.

My results tell me that timed tests can be beneficial when used as a form of practice. Students should be able to correct their own tests, allowing for immediate feedback on which problems they got wrong and what the correct answer should have been. Teachers should also consider not grading the tests so that students can focus on improving their skills and not on a grade.

From my research, experience, and results, I was also able to infer that effective teaching strategies center around balancing conceptual understanding and procedural skills. To provide this balance, educators should emphasize problem-solving strategies at an early age. Problem-solving methods provide students with flexibility in their learning, making it easier for them to transfer their knowledge to various math problems and real-life situations. Also, teachers should teach one concept at a time and teach it to mastery before moving on.

To motivate their students, educators must first develop their understanding of the multiplication procedure and concept. This understanding will boost the students' self-esteem, giving them confidence and motivating them intrinsically. The second step in motivating students is introducing fun ways for them to learn, study, and memorize the basic facts. In my experience, students preferred competitive games, especially when playing against the teacher. The participants in my study were really excited about having a "multiplication bowl," where the two fourth-grade classes would face off in a multiplication competition. I am excited to continue to test these teaching strategies in the next cycle of my action research journey.

References

- *Edelman, B., Abdit, H., & Valentin, D. (1995). Multiplication number facts: Modeling human performance with connectionist networks.* Psychologica Belgica, 36, *31–63.*
- *Isaacs, A., Carroll, W., & Bell, M. (2001). UCSMP Everyday Mathematics Curriculum.* Journal of Mathematical Behavior, 19, *49–62.*
- *Jones, S. C. (1995). Review of cognitive research.* Educational Memory Aids, 2, *57–60.*
- *Van de Walle, J. A. (2003).* Elementary and middle school mathematics: Teaching developmentally. *Upper Saddle River, NJ: Pearson, 156–176.*

Attention to the three important concepts of validity, reliability, and generalizability will help teacher researchers ensure the quality of their work. These concepts are also important for teacher researchers who are reviewing published and unpublished research. That is, to both do action research and measure the quality of the action research you're reading about, you need a basic understanding of the concepts of validity, reliability, and generalizability.

Validity

The word *validity* is common in our everyday professional language. For example, teachers will ask, "Are the results of the California Achievement Test really valid?," or my preservice teachers will often comment, "My students did poorly on the history test I gave them, but I'tm not sure it's an accurate representation of what they really know." Recently, I have also heard teachers discuss whether open-ended assessment strategies really measure their students' ability, and with the movement in the United States to the Common Core State Standards (CCSS) and the associated assessments (e.g., Smarter Balanced), there will no doubt be continued debate about whether the assessments accurately reflect what children know and are able to use as it relates to college readiness. The reauthorization of the Elementary and Secondary Education Act (ESEA) and its move from No Child Left Behind (NCLB) to the Every Student Succeeds Act (ESSA) in the United States signals yet another shift in how states will grapple with issues of teacher quality and student achievement. Specifically, ESSA allows for nonacademic measures as part of school scores of "school climate" and "social-emotional learning." All these trends will raise questions about **validity**, or how we know that the data we collect (e.g., test scores) accurately gauge what we are trying to measure. To put it technically, "validity refers to the degree to which scientific observations actually measure or record what they purport to measure" (Pelto & Pelto, 1978, p. 33).

Historically, validity was linked to numerically based research conducted in the positivistic tradition. For example, Cronbach and Meehl (1955) developed criteria for four different types of validity. These types of validity served to convince the researcher and the researchee that the "results" of the research were "right" and "accurate" and could withstand scrutiny from other researchers. In quantitative research, **internal validity** refers to the degree to which results are true for the participants in the study (Hendricks, 2017). For example, Alyson Marland's vignette at the start of this chapter attempted to determine whether the planned intervention of using students' conceptual understanding of the multiplication process along with efficient basic facts strategies affected students' scores on state math tests. The degree to which Marland's research results are attributable to the students' conceptual understanding of the multiplication process along with efficient basic facts strategies and not to another explanation is the degree to which the study is internally valid. In her study, Marland discovered that in fact 35 percent of her students were not meeting state standards in math in spite of her planned intervention. However, this discovery led her to a deeper understanding of the impact of her teaching strategies on student achievement on the statewide assessment test.

In quantitative research, we must also consider the external validity of the research. **External validity** is the degree to which study results are generalizable, or applicable, to groups and environments outside of the research setting. In other

words, an examination of external validity focuses on threats or rival explanations that disallow the results of a study to be generalized to other settings or groups. Marland made no claims to external validity in her study of 35 fourth-grade students in her class. To make claims to findings that would have external validity, Marland would need to satisfy the tenets of quantitative research, including the random assignment of her students to a control group and an experimental group. This was clearly not the goal of her research, which focused on promoting student understanding and motivation while learning basic number facts.

As many types of qualitative research became more popular in classroom settings in the late 1970s and early 1980s, it became common for qualitative researchers to begin to justify and defend the validity of their studies according to the criteria that had previously been applied to quantitative studies. For example, as a graduate student completing a research-based master's thesis on the effects of high geographic mobility on the children of low-income families, I was required by my advisors to dedicate considerable time and effort to justifying and defending the accuracy of my account. They confronted me with the question, "How will your readers know that your case studies accurately portray the lives of these children?" (Mills, 1985).

In the early days of my career, this seemed like an overwhelming task because there was a paucity of literature that specifically dealt with the issue. Since then, individuals have been experimenting with a new vocabulary that captures the essence of the term *validity* in a way that applies specifically to the methods of qualitative research. Kincheloe (1991) asks, "Is *trustworthiness* a more appropriate word to use?" (p. 135), whereas Wolcott (1994) suggests that "*understanding* seems to encapsulate the idea as well as any other everyday term" (p. 367). Greenwood and Levin (2000) argue that because action researchers do not make claims to context-free knowledge (i.e., action research by its very nature is based in the context of our own classrooms and schools), issues of credibility, validity, and reliability in action research are measured by the willingness of teacher researchers (and the stakeholders in our studies) "to act on the results of the action research, thereby risking their welfare on the 'validity' of their ideas and the degree to which the outcomes meet their expectations" (p. 98). In short, the validity of our action research depends on whether the solution to a problem (our planned intervention) actually solves our problem!

Let's look at criteria for measuring the quality of qualitative research based on these two terms—*trustworthiness* and *understanding*—and then look at strategies for increasing the validity of your action research.

Guba's Criteria for Validity of Qualitative Research

Guba's article "Criteria for Assessing the Trustworthiness of Naturalistic Inquiries" (Guba, 1981) speaks directly to qualitative researchers. Guba argued that the

trustworthiness of qualitative inquiry could be established by addressing the following characteristics of a study: *credibility*, *transferability*, *dependability*, and *confirmability*.

Credibility

The **credibility** of the study refers to the researcher's ability to take into account the complexities that present themselves in a study and to deal with patterns that are not easily explained. To do this, Guba (1981) suggested that the following methods be used:

- *Do prolonged participation at the study site* to overcome distortions produced by the presence of researchers and to provide researchers with the opportunity to test biases and perceptions. By virtue of studying your own school, classroom, and students, you will be immersed in the setting and spend a prolonged amount of time at the site—probably close to 180 days per year!
- *Do persistent observation* to identify pervasive qualities as well as atypical characteristics.
- *Do peer debriefing* to provide researchers with the opportunity to test their growing insights through interactions with other professionals. For example, most of us will be able to identify a "critical friend," a colleague, or a "significant other"—somebody who is willing and able to help us reflect on our own situations by listening, prompting, and recording our insights throughout the process.
- *Practice triangulation* (discussed in Chapter 5) to compare a variety of data sources and different methods with one another in order to cross-check data.
- *Collect documents, films, video recordings, audio recordings, artifacts, and other "raw" or "slice-of-life" data items.*
- *Do member checks* to test the overall report with the study's participants before sharing it in final form.
- *Establish structural corroboration or coherence* to ensure that there are no internal conflicts or contradictions.
- *Establish referential adequacy*; that is, test analyses and interpretations against documents, recordings, films, and the like that were collected as part of the study.

Transferability

Guba's (1981) second criteria of **transferability** refers to qualitative researchers' beliefs that everything they study is context bound and that the goal of their work is not to develop "truth" statements that can be generalized to larger groups of people. To facilitate the development of descriptive, context-relevant statements, Guba proposed that the researcher should do the following:

- *Collect detailed descriptive data* that will permit comparison of a given context (classroom/school) to other possible contexts to which transfer might be contemplated.
- *Develop detailed descriptions of the context* to make judgments about fittingness with other contexts possible.

The transferability of an action research account depends largely on whether the consumer of the research can identify with the setting. Include as much detail as possible to allow the recipients of your work to "see" the setting for themselves.

Dependability

According to Guba (1981), **dependability** refers to the stability of the data. To address issues related to the dependability of the data we collect, Guba recommended the following steps:

- *Overlap methods* (similar to a triangulation process). Use two or more methods in such a way that the weakness of one is compensated by the strength of another. For example, interviews with students may be used to contribute to your understanding of what you observed happening during a lesson.
- *Establish an audit trail.* This process makes it possible for an external "auditor" (maybe a critical friend, principal, or graduate student) to examine the processes of data collection, analysis, and interpretation. This audit trail may take the form of a written description of each process and perhaps even access to original field notes, artifacts, video recordings, pictures, archival data, and so on.

Confirmability

The final characteristic that Guba (1981) addresses is the **confirmability** of the data, or the neutrality or objectivity of the data that have been collected. Guba argues that the following two steps can be taken to address this issue:

- *Practice triangulation* (discussed in Chapter 4), whereby a variety of data sources and different methods are compared with one another to cross-check data.
- *Practice reflexivity*; that is, to intentionally reveal underlying assumptions or biases that cause the researcher to formulate a set of questions in a particular way and to present findings in a particular way. One technique for doing this is to keep a journal in which reflections/musings are recorded on a regular basis. Key Concepts Box 6–1 lists Guba's criteria for validity of qualitative research.

KEY CONCEPTS BOX 6–1

Guba's Criteria for Validity of Qualitative Research

Criteria	Definition	Strategies
Credibility	The researcher's ability to take into account the complexities that present themselves in a study and to deal with patterns that are not easily explained.	Do prolonged participation at study site. Do persistent observation. Do peer debriefing. Practice triangulation. Collect "slice-of-life" data items. Do member checks. Establish structural corroboration or coherence. Establish referential adequacy.
Transferability	The researcher's belief that everything is context bound.	Collect detailed descriptive data. Develop detailed descriptions of the context.
Dependability	The stability of the data.	Overlap methods. Establish an audit trail.
Confirmability	The neutrality or objectivity of the data collected.	Practice triangulation. Practice reflexivity.

Wolcott's Strategies for Ensuring the Validity of Qualitative Research

Taken in concert with the previous discussion about validity criteria, the following strategies provide teacher researchers with practical options for making sure their research is the best it can be (adapted from Wolcott, 1994).

Talk Little; Listen a Lot

This strategy suggests that teacher researchers who are conducting interviews, asking questions, or engaging children, parents, and colleagues in discussions about the problem being studied ought to carefully monitor the ratio of listening to talking. For example, interviewing children can be difficult work—our best-thought-out questions elicit painfully brief replies, and we are left wondering what to do next. As teachers, we are in the business of talking for a living, so it comes quite naturally to us to jump in with our own answer for the child. The trustworthiness of our inquiries will be enhanced if we can bite our tongue, think of some other probing questions, and wait patiently (one thousand . . . two thousand . . . three thousand . . .). As a teacher, I have never been very comfortable with silence in my

classroom, particularly when I thought that I had asked an engaging question. My advice is to be patient and allow the respondents time to respond. Avoid being your own best informant.

Record Observations Accurately

When conducting classroom research, recording observations while you are teaching is nearly impossible. However, you should record observations as soon as possible following a teaching episode to accurately capture the essence of what transpired. Although audio and video recordings can assist with our efforts to record accurately, there will still be many occasions when, as participant observers, we have to rely on our field notes, our journals, or our memories; and for me, relying on my memory is becoming an increasingly scary thing!

Begin Writing Early

In a workday already crunched by the pressures of time, finding time to write in journals is often difficult. However, if we rely solely on our memories of what has been happening in our classrooms over an extended period of time, we are likely to fall victim to writing romanticized versions of classroom and school life. Make time to write down your reflections. The act of writing down your recollections of a teaching episode or observation will make evident to you what blanks need to be filled in—for example, what questions need to be asked the next day or what should be the focus of your observations.

Let Readers "See" for Themselves

Include primary data in any account to let the readers of your action research accounts (colleagues, principals, university professors) see the data for themselves. As Wolcott (1994) suggests, "In striking the delicate balance between providing too much detail and too little, I would rather err on the side of too much; conversely, between overanalyzing and underanalyzing data, I would rather say too little" (p. 350). This is particularly true in a schoolwide action research effort in which you are seeking support for possible change based on data that you must present to colleagues who may not have had a central role in the conduct of the study. When sharing your research reports with colleagues, let them see the data. This may mean using charts, graphs, photographs, film—whatever you have collected. In doing so, you will bring the recipient of your work along in the process and perhaps earn his or her buy-in to the next action research cycle. Showing can be more persuasive than telling.

Report Fully

In our quest to find neat answers and solutions to our problems, it is often easy to avoid keeping track of discrepant events and data. Just when we think we know the answer, some data come along to shatter the illusion of having neatly resolved the problem! We do not need to be fearful of discrepant data. After all, it is all grist

for the research mill, and although we do not need to report everything, it is helpful to keep track of the discrepant data and to seek further explanation to understand what is happening in our classrooms/schools.

Be Candid

Teacher researchers should be candid about their work, and if writing a narrative that they hope to publish or share with a broader audience, they should make explicit any biases that they may have about the inquiry they have undertaken. Teacher researchers should also make explicit the things about which they have made judgments because it is easy to slip into a narrative that seeks to validate one's position. Being candid may also provide an opportunity to be explicit about events that occurred during the study and that may have affected the outcomes. For example, high student turnover rates may provide an explanation for fluctuating test scores.

Seek Feedback

It is always a good idea to seek feedback from colleagues (and perhaps even students, parents, volunteers, and administrators) on your written study. Other readers will help raise questions about what you as the writer will have taken for granted. They will raise questions about the accuracy of the account and help you to go back to your classroom in your quest to get the story right (or, at least, not all wrong).

Write Accurately

Once you have written a description of your action research, it is a good idea to read the account aloud or to ask a close colleague to read the account carefully to look for contradictions in the text. Often we are too close to the investigation to really see the contradictions that may be blatantly obvious to an outsider. Nevertheless, the accuracy of the account (whether written or "performed") is critical to the validity of the study. (For further discussion of these points and a discussion of "When It Really Matters, Does Validity Really Matter?" see Wolcott, 1994, pp. 348–370.) See Research in Action Checklist 6–1 for Wolcott's strategies for ensuring the validity of qualitative research.

Reliability

In everyday English, *reliability* means "dependability" or "trustworthiness." The term means essentially the same thing with respect to measurement. Basically, **reliability** is the degree to which a test consistently measures whatever it measures. The more reliable a test is, the more confidence we can have that the scores obtained from the administration of the test are essentially the same scores that would be obtained if the test were readministered. An unreliable test is essentially useless; if a test is unreliable, then scores for a given sample would be expected to

RESEARCH IN ACTION CHECKLIST 6–1

Wolcott's Strategies for Ensuring the Validity of Qualitative Research

_____ Talk little; listen a lot.

_____ Record accurately.

_____ Begin writing early.

_____ Let readers "see" for themselves.

_____ Report fully.

_____ Be candid.

_____ Seek feedback.

_____ Write accurately.

be different every time the test was administered. If an intelligence test was unreliable, for example, then a student scoring an IQ of 120 today might score an IQ of 140 tomorrow and a 95 the day after tomorrow. If the test was reliable and if the student's IQ was 110, then we would not expect his or her score to fluctuate too greatly from testing to testing; a score of 105 would not be unusual, but a score of 145 would be very unlikely. If you have ever administered standardized tests to students, you will be familiar with the reliability coefficients that are presented in the administration manuals. The numbers are meant to convey to the test user the peace of mind that, if the test were administered on a future occasion, individual students would score roughly the same.

Reliability is expressed numerically, usually as a coefficient; a high coefficient indicates high reliability. If a test were perfectly reliable, the coefficient would be 1.00; this would mean that a student's score perfectly reflected his or her true status with respect to the variable being measured. However, no test is perfectly reliable. Scores are invariably affected by errors of measurement resulting from a variety of causes. High reliability indicates minimum error variance; if a test has high reliability, then the effect of errors of measurement has been reduced. Errors of measurement affect scores in a random fashion; some scores may be increased, while others are decreased. Errors of measurement can be caused by characteristics of the test itself (e.g., ambiguous test items that some students just happen to interpret correctly), by conditions of administration (e.g., directions not properly followed), by the current status of the persons taking the test (some may be tired, others unmotivated), or by a combination of any

of the above. High reliability indicates that these sources of error have been eliminated as much as possible.

Errors of measurement that affect reliability are random errors; systematic or constant errors affect validity. If an achievement test was too difficult for a given group of students, all scores would be systematically lowered; the test would have low validity for that group (remember, "valid for whom?"). The test might, however, yield consistent scores (i.e., might be reliable); in other words, the scores might be systematically lowered in the same way every time. A given student whose "true" achievement score was 80 and who scored 60 on the test (invalidity) might score 60 every time he or she took the test (reliability). This illustrates an interesting relationship between validity and reliability: A valid test is always reliable, but a reliable test is not necessarily valid. In other words, if a test is measuring what it is supposed to be measuring, it will be reliable and do so every time, but a reliable test can consistently measure the wrong thing and be invalid!

For qualitatively oriented action researchers, the message is simple: As you think about the results of your inquiry, consider whether you think that your data would be consistently collected if the same techniques were utilized over time, or, if you are working as a member of a team that is collecting data, work out how to resolve any differences among observers so you can agree on the descriptive accuracy of an account.

Voices from the Field

Reliability

The teacher researcher in this vignette provides a compelling example of the challenges facing action researchers attempting to measure constructs such as "tenderness" as it relates to gender bias in basal readers. Jeanette explains that she relied on her own worldview based on her undergraduate education (and life) to "observe" specific traits and behaviors. Without connecting to the literature related to these traits and behaviors and without any attempt to establish inter-rater reliability, it raises questions about the dependability and trustworthiness of Jeanette's observations.

ENHANCEDetext video example 6–1

In this video, Jeannette, a teacher researcher, notes that the data collection instruments and procedures she used in her action research project required subjective judgment and admits that she did not assess the reliability of her data collection and analysis. Her comments make it clear that, in the future, she plans to pay more attention to this important aspect of action research.

The Difference Between Reliability and Validity

To review, reliability "is the degree to which a test consistently measures whatever it is measuring" (Mills & Gay, 2016, p. 168), and validity is "the degree to which a test measures what it is supposed to measure" (Mills & Gay, 2016, p. 117). Reliability, however, is not the same thing as validity. Remember, a valid test that measures what it purports to measure will do so consistently over time. A reliable test may consistently measure the wrong thing.

Generalizability

Historically, research in education concerned itself with **generalizability**, a term that refers to the applicability of findings to settings and contexts different from the one in which they were obtained; that is, based on the behavior of a small group of individuals, researchers try to explain the behavior of a wider group of people. This view of generalizability, however, is not directly applicable to teacher action research—even though there is still a mind-set among some teachers, administrators, and policymakers that the findings of action research studies should be transferable. Many of these people believe that we should be able to generalize from the outcomes of a study in one classroom, one school, and one district to all similar classrooms in the state or country. This is not the nature of the research in which we are engaged.

The goal of action research is to understand what is happening in your school or classroom and to determine what might improve things in that context (Sagor, 1992). Therefore, action researchers don't need to worry about the generalizability of data because they are not seeking to define ultimate truths. However, one reviewer had the following reaction to this dismissal of generalizability:

> I fear, however, that this approach lends credence to many of my colleagues' beliefs that action research is unscientific, biased, and not generalizable. Some go so far as to call it "garbage research." The question that they often pose is what good is research that is not generalizable? (Anonymous reviewer)

Indeed, action research has faced a self-esteem problem among many "academics" who question the worthiness of the activity as "scientific" inquiry. Confronted with a similar argument, Stringer (1996) offered the following response:

> Whether or not action research is accepted as "scientific" depends on the way in which science is defined. Certainly it is, in one sense, rigorously empirical, insofar as it requires people to define clearly and observe the phenomenon under investigation. What is also evident, however, is that action research does not follow the carefully prescribed procedures that have become inscribed as scientific method. (p. 145)

Stringer goes on to argue that in spite of the success of the scientific method in advancing our knowledge in the "hard sciences," the applicability of this

Voices from the Field

Generalizability

The teacher researcher in this vignette makes a compelling case for the difference between external validity/generalizability and the goal of action research to understand what is happening in your school or classroom and to determine what might improve things in that context. Jureen is passionate in her belief that what she learned in her study can lead to changes in her teaching in order to better meet her students' needs.

ENHANCEDetext
video example 6–2
Jureen, the teacher researcher shown in this video, explains why, in her view, action research does not necessarily need to lead to generalizable results.

method to inquiries of human behavior has met with little success in increasing the predictability of human behavior. Other texts on educational research agree that action research is a different type of inquiry entirely and as such should not be focused on generalizability. For example, Vockell and Asher (1996) state,

> Action research refers to the practical application of the scientific method or other forms of disciplined inquiry to the process of dealing with everyday problems. It is particularly focused on teachers and other educators doing action research in order to make their particular educational activities more productive. It is more concerned with specific classes and programs and less concerned with generalized conclusions about other classes and programs. (p. 10)

Action research is not "garbage research" at the classroom/school level. As teacher researchers, we are challenging the experimental researcher's view that the only credible research is that which can be generalized to a larger population. Many examples of teacher research are generalizable to other classroom settings, but the power of action research is not in its generalizability. It is in the relevance of the findings to the researcher or the audience of the research.

Personal Bias in the Conduct of Action Research

Related to the issue of generalizability of research is the issue of personal bias. If we conduct our research in a systematic, disciplined manner, we will go a long way toward minimizing personal bias in our findings. However, in an intimate activity

such as action research, it is a challenge to remain "objective" and open, to look into the mirror of our findings and reflect on what we see. It is relatively easy in any research, should we so choose, to collect data that simply validate our existing practices, to maintain the status quo, to pat ourselves on our collective backs, and to ignore discrepant data or discredit research results. The same can be said for reviewing related literature—we may choose to review only the literature that supports a particular thesis that we wish to promote. None of these are acceptable approaches for reconciling the biased collection of data.

Propositions

One way for teacher researchers to get in touch with their biases about the subject they are investigating is to develop a list of propositions about what they think they will find during the course of their investigations. These propositions provide a window into the belief system and personal biases that can—and often do—creep into the investigation. These statements also provide a good starting point for examining teacher researchers' theories about teaching and learning and for understanding where those theories came from.

For example, a teacher who wishes to investigate the effects of manipulatives on student achievement in mathematics may generate propositions such as the following:

1. The use of manipulatives when teaching mathematics will increase students' conceptual knowledge of mathematics.

Voices from the Field

Personal Bias in the Conduct of Action Research (Video 1)

The teacher researcher in this vignette makes a compelling case of action researchers to "embrace" their personal biases as a strategy for making explicit what they expect to find in their quests to understand causal relationships in teaching and learning. In so doing, teacher researchers can better understand the theories of teaching and learning that have impacted how they view their practices and to design systematic data collection strategies that will in all likelihood support and challenge their preconceived notions about what works in classrooms.

ENHANCEDetext video example 6–3
Watch this video as Doug presents his views on some of the personal biases that affect action research.

Voices from the Field

Personal Bias in the Conduct of Action Research (Video 2)

The teacher researcher in this vignette provides a compelling example of the importance of action researchers to be aware of their preconceived notions and beliefs (perhaps through completing a "Propositions" activity) and how these beliefs potentially impact the research process, from the review of literature to data collection and data analysis and interpretation. Jureen recommends using the review of literature as a kind of instructional "check for understanding" to help teacher researchers link their findings to the existing body of knowledge of the phenomenon under investigation (in her case, the relationships between technology implementation and student achievement). Throughout the action research process, teacher researchers must remain vigilant to potential biases that emerge through the very nature of being engaged in an intimate and open-ended process.

ENHANCEDetext video example 6–4
Watch this video as Jureen presents her views on some of the personal biases that affect action research.

2. The use of manipulatives will help overcome math anxiety because the children will have more fun doing math.
3. The use of manipulatives will improve students' basic number facts skills.

Teacher researchers will find it useful to examine their propositions closely so that they can explore what they believe they will find before they start their investigations and what they might do to ensure that they remain vigilant in the fidelity with which they collect their data (thus addressing the concerns of researcher bias). Similarly, this activity helps to clarify teacher researchers' conceptual frameworks for their investigations by making explicit the theories that affect what they do before, during, and after the research.

SUMMARY

1. Attention to the three important concepts of validity, reliability, and generalizability will help teacher researchers ensure the quality of their work. These concepts are also important for teacher researchers who are reviewing published and unpublished research.

2. Qualitative researchers, action researchers, and quantitative researchers disagree about the value of applying these concepts of validity, reliability, and generalizability to qualitatively oriented action research.
3. Teacher researchers must understand the meanings of validity, reliability, and generalizability to be knowledgeable consumers of research as well as producers of research that we hope will be trustworthy and persuasive in their own eyes and in the eyes of their audience.

Validity

4. Validity is "the degree to which scientific observations actually measure or record what they purport to measure" (Pelto & Pelto, 1978, p. 33).
5. Validity in quantitative research can be thought of in terms of internal validity and external validity.
6. Internal validity is the degree to which observed differences on the dependent variable are a direct result of manipulation of the independent variable, not some other variable.
7. External validity is the degree to which study results are generalizable, or applicable, to groups and environments outside of the research setting.
8. Action researchers do not make claims to context-free knowledge (i.e., action research by its very nature is based in the context of our own classrooms and schools); issues of credibility, validity, and reliability in action research are measured by the willingness of teacher researchers (and the stakeholders in our studies) "to act on the results of the action research" (Greenwood & Levin, 2000, p. 98).
9. The validity of our action research depends on whether the solution to a problem (our planned intervention) actually solves our problem.
10. Guba (1981) argued that the trustworthiness of qualitative inquiry could be established by addressing the following characteristics of a study: credibility, transferability, dependability, and confirmability.
11. The credibility of a study refers to the researcher's ability to take into account the complexities that present themselves in a study and to deal with patterns that are not easily explained.
12. Transferability refers to qualitative researchers' beliefs that everything they study is context bound and that the goal of their work is not to develop "truth" statements that can be generalized to larger groups of people.
13. Dependability refers to the stability of the data.
14. Confirmability of the data refers to the neutrality or objectivity of the data that have been collected.
15. Wolcott's (1994) strategies for ensuring validity of qualitative research are talk little, listen a lot; record observations accurately; begin writing early; let readers "see" for themselves; report fully; be candid; seek feedback; and write accurately.

16. Talk Little; Listen a Lot. This strategy suggests that teacher researchers who are conducting interviews ought to carefully monitor the ratio of listening to talking. The trustworthiness of our inquiries will be enhanced if we can bite our tongue, think of some other probing questions, and wait patiently. Avoid being your own best informant.
17. Record Observations Accurately. When conducting research in your own classroom, recording observations while you are teaching is nearly impossible. However, you should record observations as soon as possible following a teaching episode to accurately capture the essence of what transpired.
18. Begin Writing Early. The act of writing down your recollections of a teaching episode or observation will make evident to you what blanks need to be filled in—for example, what questions need to be asked the next day or what should be the focus of your observations.
19. Let Readers "See" for Themselves. Include primary data in any account to let the readers of your action research accounts "see" the data for themselves. This may mean using charts, graphs, and photographs—whatever you have collected—to provide the reader with a window into your research. Showing can be more persuasive than telling.
20. Report Fully. In our quest to find neat answers and solutions to our problems, it is often easy to avoid keeping track of discrepant events and data. We do not need to be fearful of discrepant data that may help us to seek further explanation to understand what is happening in our classrooms and schools.
21. Be Candid. Teacher researchers should be candid about their work, and if writing a narrative that they hope to publish or share with a broader audience, they should make explicit any biases that they may have about the inquiry they have undertaken.
22. Seek Feedback. It is always a good idea to seek feedback from colleagues on your written study. Other readers will raise questions about what you as the writer will have taken for granted and about the accuracy of the account.
23. Write Accurately. Once you have written a description of your action research, it is a good idea to read the account aloud or to ask a close colleague to read the account carefully to look for contradictions in the text.

Reliability

24. Reliability is the degree to which a test consistently measures whatever it measures. The more reliable a test is, the more confidence we can have that the scores obtained from the administration of the test are essentially the same scores that would be obtained if the test were readministered.
25. Remember, a valid test is always reliable, but a reliable test is not necessarily valid. In other words, if a test is measuring what it is supposed to be measuring, it will be reliable and do so every time, but a reliable test can consistently measure the wrong thing and be invalid!

26. For qualitatively oriented action researchers, the message is simple: As you think about the results of your inquiry, consider whether you think that your data would be consistently collected if the same techniques were utilized over time.

Generalizability

27. Historically, research in education concerned itself with generalizability, a term that refers to the applicability of findings to settings and contexts different from the one in which they were obtained; that is, based on the behavior of a small group of individuals, researchers try to explain the behavior of a wider group of people. This view of generalizability, however, is not directly applicable to teacher action research.
28. The goal of action research is to understand what is happening in your school or classroom and to determine what might improve things in that context (Sagor, 1992). Therefore, action researchers don't need to worry about the generalizability of findings because they are not seeking to define ultimate truths.

Personal Bias in the Conduct of Action Research

29. Related to the issue of generalizability of research is the issue of personal bias. If we conduct our research in a systematic, disciplined manner, we will go a long way toward minimizing personal bias in our findings.
30. Propositions. One way for teacher researchers to get in touch with their biases about the subject they are investigating is to develop a list of propositions about what they think they will find during the course of their investigations. These propositions provide a window into the belief system and personal biases that can—and often do—creep into the investigation. These statements also provide a good starting point for examining teacher researchers' theories about teaching and learning and for understanding where those theories came from.

TASKS

1. Describe how you have addressed the issues of validity, reliability, and generalizability in your action research inquiry. Specifically, how will you know that your planned intervention actually solves your problem?

2. Develop a list of propositions about the subject you are investigating (see the section "Personal Bias in the Conduct of Action Research"). What do each of these propositions tell you about your belief system and biases?