

# A Model for Presenting Threats to Legitimation at the Planning and Interpretation Phases in the Quantitative, Qualitative, and Mixed Research Components of a Dissertation

Cindy L. Bengé<sup>1</sup>, Anthony J. Onwuegbuzie<sup>2,\*</sup>, & Mary E. Robbins<sup>1</sup>

<sup>1</sup>Department of Language, Literacy and Special Populations, Sam Houston State University, Huntsville, Texas 77341-2119, USA

<sup>2</sup>Department of Educational Leadership and Counseling, Sam Houston State University, Huntsville, Texas 77341-2119, USA

\*Corresponding author: Department of Educational Leadership and Counseling, Box 2119, Sam Houston State University, Huntsville, Texas 77341-2119, USA

E-mail: [tonyonwuegbuzie@aol.com](mailto:tonyonwuegbuzie@aol.com)

Received: September 10, 2012 Accepted: October 26, 2012 Published: November 25, 2012

doi:10.5296/ije.v4i4.2360 URL: <http://dx.doi.org/10.5296/ije.v4i4.2360>

## Abstract

The assessment of validity—or more appropriately known as legitimation—is the most important step in all research studies, whether the research study represents a quantitative research study, a qualitative research study, or a mixed research study. Despite its importance, a significant proportion of authors do not discuss to any degree the legitimation issues that emerged in their studies. It is likely that the prevalence rate among authors of dissertations is even higher, not only because they represent beginning researchers but also because many of their advisors/supervisors themselves do not discuss (adequately) the limitations of findings/interpretations. Thus, in this article we provide a model for presenting threats to legitimation both at the planning and interpretation phases in the quantitative, qualitative, and mixed research components of a dissertation. In so doing, we provide an exemplar of the legitimation process, which we believe provides evidence of a comprehensive technique for identifying and addressing threats to legitimation for researchers in general and dissertation researchers in particular.

**Keywords:** *legitimation; validity; quantitative research; qualitative research; mixed research; mixed methods research*

The assessment of validity—or more appropriately known as legitimation (Note 1)—is the most important step in all research studies, whether the research study represents a quantitative research study, a qualitative research study, or a mixed research study (Note 2). In fact, it does not matter how appropriate or important the research question is, how appropriate the sample size or sampling scheme is, how appropriate the research design is, how much data are collected, or how complex an analysis is undertaken, if the findings and/or interpretations lack legitimation, then the study has little or no merit or value (i.e., no beneficence), and, moreover, actually can invoke harm depending on the extent to which the interpretations that ensue from the findings are misleading (i.e., maleficence). In other words, a study is limited by the legitimation of its findings and/or interpretations. As such, legitimation issues—or what can be referred to as quality criteria—can make or break a study.

Despite the importance of legitimation in quantitative, qualitative, and mixed research studies, a significant proportion of authors do not discuss to any degree the legitimation issues that emerged in their studies. For example, Onwuegbuzie and Daniel (2005), who examined 52 manuscripts submitted to a peer reviewed journal—namely, *Research in the Schools*—during the course of a 2-year period, documented that 65% of authors do not discuss the limitations of their findings/interpretations at any point. More recently, Onwuegbuzie (2012) reported a similar prevalence rate of 66.7%. Even more disturbingly, some specific limitations were not discussed by an even higher proportion of authors. For example, between 83.3% (Onwuegbuzie, 2012) and 91.0% (Onwuegbuzie & Daniel, 2005) of authors of quantitative research articles do not discuss the assumptions associated with their statistical models (e.g., normality, independence, homogeneity of variance).

Unfortunately, not discussing the limitations of findings/interpretations contradicts the reporting criteria stipulated by the American Educational Research Association (AERA) in the document entitled, “Standards for Reporting on Empirical Social Science Research in AERA Publications.” The AERA (2006) standards are driven by two general principles:

First, reports of empirical research should be *warranted*; that is, adequate evidence should be provided to justify the results and conclusions. Second, reports of empirical research should be *transparent*; that is, reporting should make explicit the logic of inquiry and activities that led from the development of the initial interest, topic, problem, or research question; through the definition, collection, and analysis of data or empirical evidence; to the articulated outcomes of the study. Reporting that takes these principles into account permits scholars to understand one another’s work, prepares that work for public scrutiny, and enables others to use that work. (AERA, 2006, p. 33)

According to AERA (2006), “Reporting that takes these principles into account permits scholars to understand one another’s work, prepares that work for public scrutiny, and enables others to use that work” (p. 33). Thus, not discussing the limitations of findings/interpretations culminates in empirical reports that are neither adequately warranted nor transparent.

Most importantly, reflecting on the limitations of findings/interpretations is beneficial at all various points of the research process—namely, at the research conceptualization, research planning, research implementation, and research utilization stages. Specifically, at the research conceptualization and research planning stages, reflecting on the potential legitimation issues can help researchers optimize their research designs by making necessary adjustments. For example, in quantitative, qualitative, and mixed research studies, if the researcher anticipates that a significant proportion of participants might drop out of the study (i.e., mortality), then he/she might compensate by oversampling. Such oversampling involves selecting additional participants—either randomly or purposively—above and beyond the number of participants needed for the sample to achieve the desired statistical power (i.e., “how likely it is that the researcher will find a relationship or difference that really prevails” Onwuegbuzie & Leech, 2004, p. 205) for a quantitative study/phase and/or some form of saturation (cf. Flick, 1998; Lincoln & Guba, 1985; Morse, 1995; Onwuegbuzie, Dickinson, Leech, & Zoran, 2009; Strauss & Corbin, 1990) for a qualitative study/phase—whether data saturation (i.e., which arises when information emerges so repeatedly that the researcher can expect it and wherein the collection of additional data appears to yield no additional interpretive value; Sandelowski, 2008; Saumure & Given, 2008) or theoretical saturation (i.e., which arises when the researcher can conclude that her/his emergent theory is adequately developed to fit any future data that are collected; Sandelowski, 2008).

At the research implementation stage, reflecting on the potential legitimation issues can help researchers place their findings and interpretations in a more appropriate context. For example, being aware of the actual sample size limitations in quantitative, qualitative, and mixed research studies can prevent a researcher from over-generalizing findings, whether the generalization represents one of the types conceptualized by Onwuegbuzie, Slate, Leech, and Collins (2009), which include the following: (a) external (statistical) generalizations (i.e., which involves making generalizations, judgments, inferences, or predictions on data extracted from a representative statistical [i.e., optimally large and random] sample to the *population* from which the sample was selected [i.e., universalistic generalizability]); (b) internal (statistical) generalizations (i.e., involves making generalizations, judgments, inferences, or predictions on data extracted from one or more representative or elite research participants [e.g., sub-sample members, key informants] to the *sample* from which the participant[s] was drawn [i.e., particularistic generalizability]); (c) analytic generalizations (i.e., wherein “the investigator is striving to generalize a particular set of [case study] results to some broader theory” [Yin, 2009, p. 43] and are “applied to wider theory on the basis of how selected cases ‘fit’ with general constructs” [Curtis, Gesler, Smith, & Washburn, 2000, p. 1002]); or (d) case-to-case transfer (i.e., which involves making generalizations, judgments, or inferences from one case to another [similar] case; Miles & Huberman, 1994). Finally, at the research utilization stage, identifying and presenting potential legitimation issues can provide directions for future research.

As high as is the prevalence rate of authors who do not discuss the limitations of findings/interpretations in their articles, it is likely that the prevalence rate among authors of dissertations is even higher, not only because they represent beginning researchers but also

because many of their advisors/supervisors themselves do not discuss (adequately) the limitations of findings/interpretations. Thus, it is likely that a cycle of non-discussion or inadequate discussion of limitations is occurring within the academic community. As such, the most effective way to end this cycle is by instilling good reporting practices at the doctoral level. As such, the purpose of the remainder of this article is to provide a model for presenting threats to legitimation both at the planning and interpretation phases in the quantitative, qualitative, and mixed research components of a dissertation. However, before we provide this model, we will present the conceptual frameworks that drove this model.

## **Conceptual Framework**

### *Quantitative Research Dissertations/Quantitative Components of Mixed Research Reports*

For quantitative research, discussion of limitations takes the form of a presentation of threats to internal validity and external validity. Interestingly, and consistent with our assertion of the importance of identifying and discussing limitations in quantitative research, instructors of research methodology, statistics, and measurement courses regard internal validity and external validity as representing the most important issues in their fields, as well as receiving the most extensive coverage in their classes (Mundfrom, Shaw, Thomas, Young, & Moore, 1998). Donald T. Campbell and Julian C. Stanley and their colleagues (Campbell, 1957; Campbell & Stanley, 1963; Cook & Campbell, 1979) promoted the concepts of internal validity and external validity. According to Cook and Campbell (1979), internal validity is the “approximate validity with which we infer that a relationship between two variables is causal” (p. 37). Johnson and Christensen (2010) provide a simpler definition wherein internal validity is “the ability to infer that a causal relationship exists between two variables” (p. 247). Moreover, as noted by Onwuegbuzie (2003), internal validity is threatened to the extent that plausible rival explanations cannot be eliminated. In contrast, external validity is defined as “the extent to which the study results can be generalized to and across populations of persons, settings, times, outcomes, and treatment variations” (Johnson & Christensen, 2010, p. 585). Accordingly, even if a finding represents high internal validity, this attribute does not necessarily imply that this finding can be generalized outside the study context.

Campbell and Stanley (1963) conceptualized eight threats to internal validity, as follows: history, maturation, testing, instrumentation, statistical regression, differential selection of participants, mortality, and interaction effects (e.g., selection-mortality interaction). Further, Campbell and Stanley (1963) also conceptualized the following four threats to external validity: the reactive or interaction effect of testing, the reactive effects of experimental arrangements, the interaction effects of selection biases and the experimental variable, and multiple treatment interference. Although Campbell and Stanley’s (1963) groundbreaking work represents the most authoritative source regarding threats to internal validity and external validity to date—with 16,660 citations identified at the time of writing this article using Harzing’s (2009) *Publish or Perish* software and Google Scholar—their framework for internal and external validity contains two important limitations. First, as noted by Onwuegbuzie (2003), Campbell and Stanley’s (1963) framework was developed exclusively

for experimental studies; thus, it is insufficient for the other types of quantitative research designs (e.g., descriptive, correlational)—essentially meaning that the Campbell and Stanley's (1963) validity threats lack external validity inasmuch as they cannot necessarily be generalized to other research designs. Second, Campbell and Stanley's (1963) threats to internal validity and external validity are presented as a static entity; yet, as conceptualized by Onwuegbuzie (2003), threats to internal validity and external validity occur at one or more of the three major stages of the inquiry process, namely: (a) research design/data collection, (b) data analysis, and (c) data interpretation—which “typically represent three distinct time points in the research process” (p. 74).

To this end, building on the works of Campbell and Stanley, Onwuegbuzie (2003) conceptualized a framework for internal and external validity threats that represented a 3 (stage of research process) x 2 (internal vs. external validity) matrix comprising 50 unique dimensions of internal and external validity threats, with many of the dimensions containing sub-dimensions (cf. Figure 1). This framework is called the Quantitative Legitimation Model. Each of these threats is defined in Figure 2.

#### *Qualitative Research Dissertations/Qualitative Components of Mixed Research Reports*

As noted by Onwuegbuzie and Leech (2007), for qualitative research, discussion of limitations takes the form of discussion of threats to internal credibility and external credibility. According to Onwuegbuzie and Leech (2007), internal credibility represents “the truth value, applicability, consistency, neutrality, dependability, and/or credibility of interpretations and conclusions within the underlying setting or group” (p. 234). In contrast, external credibility represents “the confirmability and transferability of findings and conclusions” (p. 235). As such, to a degree, internal credibility in qualitative research is parallel to internal validity in quantitative research, whereas external credibility in qualitative research is parallel to external validity in quantitative research.

In an attempt to obtain a comprehensive framework of legitimation in qualitative research, and building on the work of Onwuegbuzie (2003), Onwuegbuzie and Leech (2007) conceptualized what they labeled as the *Qualitative Legitimation Model*, which comprises 29 elements of legitimation for qualitative research at the following three interactive and recursive stages of the research process: research design/data collection, data analysis, and data interpretation (cf. Figure 3). Each of these threats is defined in Figure 4. A major difference between the Quantitative Legitimation Model and the Qualitative Legitimation Model is that the model developed for quantitative research studies is based on the assumption that the design/data collection, data analysis, and data interpretation stages of the quantitative research process generally represent three distinct and (mostly) linear phases, whereas the model developed for qualitative research operates under the premise that, in qualitative research, these three stages are both interactive and iterative.

#### *Mixed Research Reports*

When conducting mixed research, researchers not only have to assess the limitations of the findings and interpretations yielded by both the quantitative research components/phases

and qualitative research components/phases, but also they have to assess the limitations of the findings and interpretations stemming from the mixed research components/phases. As outlined earlier, mixed researchers can use Figure 1 and Figure 2 to assess the limitations of the findings and interpretations stemming from the quantitative components/phases and Figure 3 and Figure 4 to assess the limitations of the findings and interpretations stemming from the qualitative components/phases. For the mixed research components/phases, mixed researchers can use Onwuegbuzie and Johnson's (2006) typology of mixed research legitimation, which comprises the following nine types of legitimation: sample integration, insider-outsider, weakness minimization, sequential, conversion, paradigmatic mixing, commensurability, multiple validities, and political. Each of these types of legitimation is defined in Table 1.

### **An Exemplar for Presenting Threats to Legitimation at the Planning and Interpretation Phases in the Quantitative, Qualitative, and Mixed Research Components of a Dissertation**

The example we provide was written by Benge (2012), who conducted a mixed research study for two purposes: (a) to replicate and to extend previous research involving the examination of both student success and perceptions concerning the use of cartoon mnemonics in combination with traditional definitions (e.g., Atkinson, 1975; Mastropieri & Scruggs, 1998), the use of dictionary definitions (McKeown, 1993; Miller & Gildea, 1987), and the use of revised definitions (McKeown, 1993) alone as tools in vocabulary acquisition; and (b) to explore the use of cartoon mnemonics in combination with revised definitions as an instructional technique for introducing individual vocabulary words and their meanings to fifth-grade students in an attempt to ascertain whether the combination of methods affects student understanding and attitudes toward learning new vocabulary. The research questions were as follows:

#### **Quantitative Research Question**

The following quantitative research question was addressed in this study: What is the difference in definition understanding of second-tier vocabulary words among the use of traditional dictionary definitions, revised dictionary definitions, traditional dictionary definitions plus cartoon mnemonics, and revised dictionary definitions plus cartoon mnemonics for fifth-grade students?

#### **Qualitative Research Questions**

The following qualitative research questions were addressed in this study: (a) What are select fifth-grade students' perceptions about the effect of traditional dictionary definitions, revised dictionary definitions, traditional dictionary definitions plus cartoon mnemonics, and revised dictionary definitions plus cartoon mnemonics on their understanding of vocabulary words? (b) What are select fifth-grade teachers' perceptions about the effect of cartoon mnemonics and revised definitions on their students' learning and understanding of vocabulary concepts? and (c) To what extent are perceptions of select fifth-grade students

regarding the effect of the concurrent use of cartoon mnemonics and revised definitions on their ability to learn vocabulary concepts similar to those of their teachers?

### **Mixed Methods Research Questions**

The following mixed methods research questions were addressed in this research study: (a) What is the relationship between fifth-grade students' self-reported vocabulary learning strategies and attitudes and their acquisition of second-tier vocabulary words? and (b) What is the relationship between fifth-grade students' perceptions of the use of cartoon mnemonics and revised definitions and their acquisition of second-tier vocabulary words?

#### *Quantitative Hypothesis*

The following hypothesis was tested in this study: There is a difference in definition understanding among the effects of traditional definitions, revised definitions, traditional definitions plus cartoon mnemonics, and revised definitions plus cartoon mnemonics for fifth-grade students.

#### *Research Design*

Benge (2012) used a mixed research design that Leech and Onwuegbuzie (2009) referred to as a fully mixed sequential equal status mixed research design, wherein both quantitative and qualitative data were collected and analyzed sequentially, and the quantitative and qualitative data were afforded approximately equal weight. In the quantitative phase, Benge (2012) utilized a counterbalanced (i.e., quasi-experimental) research design due to the use of pre-formed groups. In her counter-balanced design, each student was exposed to every condition but students in each school took part in each condition in a different order, thereby counterbalancing the conditions to control for a possible sequencing effect (Johnson & Christensen, 2010). More specifically, Benge (2012) used a Latin square as the form of counterbalanced design—specifically, she used a standard square that eliminated systematic patterns that arise when one intervention always follows another intervention in the Latin square. Table 2 presents the Latin square used by Benge (2012).

#### *Data Collection*

**Participants and setting.** For the quantitative research phase, the student participants were divided according to school and teacher. All students in each school who were enrolled in the classes of the teachers randomly selected for either Study 1 or Study 2 were eligible to participate in the study. In total, 87 students in Study 1 and 133 in Study 2 were included in the final data analysis. For the qualitative research phase, eight participating teachers and 12 randomly selected students who were identified as representing extreme cases with respect to their reading and writing achievement participated in follow-up interviews.

*Instruments and Procedures*

**Vocabulary Knowledge Check Sheet (VKCS).** The Vocabulary Knowledge Check Sheet (VKCS), a 60-word checklist, was administered by the participating teachers prior to the quantitative phase. The VKCS determined students' prior knowledge and understanding of 48 second-tier words (i.e., high-frequency words for "mature language users" that lie somewhere between what is generally considered basic vocabulary and content-specific vocabulary; Beck, McKeown, & Kucan, 2002, p. 7). Students selected from four response options based on Dale's (1965) "continuum" (p. 898) of vocabulary knowledge that provided for levels of word knowledge. The response items comprised the following: (a) do not know the word, have never seen it; (b) have seen or heard the word but don't know what it means; (c) know something about it, can relate it to something; and (d) know it well, can define it, and use it in speaking or writing (Dale, 1965).

Benge (2012) calculated descriptive statistics and used the results during a second iterative process to select the 24 words introduced over the course of the 4 weeks and the eight words used as examples over the course of both studies (i.e., the sentence completion task study and the questioning task study). Selection of words was based on the students' self-reported knowledge of the words, the type of dictionary definition, and the quality of the cartoon provided for the word. A quality cartoon included a good context for word understanding along with a corresponding caption that provided a connection to the meaning of the word and the cartoon. Each of the 24 experiment words and each of the eight example words were randomly assigned to one of the 4 weeks. The order of presentation of words also was selected randomly.

**Revised Definitions and Cartoons.** For both the sentence and questioning activities, revised definitions were created utilizing the four principles for writing definitions for young learners proposed by McKeown (1993), namely, (a) provide the subtleties of a word's role in language; (b) let the reader know how the word is used; (c) limit the number of steps the learner has to go through to access the meaning; and (d) craft the definition in such a way that the learner pays attention to the whole definition, not a fragment. One knowledgeable colleague not involved in the study and two researchers created the definitions and provided feedback on the fidelity of the definitions to the four principles for writing definitions. In addition, a panel of reading doctoral students read all definitions to provide feedback and to assess the definitions as to their student-friendliness. Cartoons included in the keyword mnemonic activities were generated commercially and utilized with the permission of the publisher (Burchers, Burchers, & Burchers, 1997, 2000).

**Sentence and Questioning Tasks.** Benge (2012) produced packets of six words for each of the four conditions (i.e., traditional definitions, revised definitions, traditional definitions plus keyword mnemonic, and revised definitions plus keyword mnemonic) in each study. Over the course of 4 weeks, participants, who were assigned to one of two tasks (i.e., sentence task or questioning task), were introduced to 24 new words (six words each week) with either a dictionary definition, a revised definition, a dictionary definition plus a cartoon mnemonic, or a revised definition plus a cartoon mnemonic. For the first study, Benge (2012)



created a sentence task activity for each of the 24 second-tier words that was administered by the participating teachers (McKeown, 1993). This activity required that students write sentences using the vocabulary words being introduced by either a traditional definition, a revised definition, a traditional definition plus cartoon mnemonic, or a revised definition plus cartoon mnemonic.

For the second study, for each 6-word activity packet, Benge (2012) created two questions for each word for all 24 second-tier words and eight example words (McKeown, 1993). Packets were administered each week over the course of the 4-week intervention. This activity required students to answer two questions concerning the concept involved, with each vocabulary word being introduced by either a traditional definition, a revised definition, a traditional definition plus cartoon mnemonic, or a revised definition plus cartoon mnemonic. Responses were assessed and coded nominally and ordinally by two of the authors.

## **Limitations of Findings Pertaining to Benge's (2012) Dissertation Study**

### *Pre-Study Considerations*

In the introduction chapter of her dissertation (i.e., Chapter 1)—which was written prior to the start of her study (i.e., as part of her dissertation research proposal), Benge (2012) provided an extremely extensive discussion of the potential limitations of her findings. Specifically, she discussed the potential limitations of her findings that might stem from her quantitative research phase/component (i.e., internal validity and external validity; cf. Tables 3-4, Appendix A), qualitative research phase/component (i.e., internal credibility and external credibility; cf. Table 5, Appendix B), and mixed research phase/component (mixed legitimacy; cf. Table 6, Appendix C). Although in many dissertations, this discussion often takes place in the introduction chapter (i.e., Chapter 1)—as was the case in Benge's (2012) dissertation, in some dissertations, such a discussion might take place at the dissertation proposal stage in the Methodology (or Methods) section of the dissertation (circa Chapter 3 in traditional dissertations). The important point here is that such deep reflection and extensive discussion needs to take place during the proposal creation stage of the dissertation process, prior to the student's defense of the dissertation and before applying for Institutional Review Board (IRB) approval to conduct the study and before beginning the study. In this way, the student can make adjustments to the research design in order to address as many of the identified limitations as possible. For example, in Benge's (2012) dissertation, because all of the student participants were exposed to all four experimental conditions, order bias (cf. Figure 2) was a potential threat to the internal validity of the findings. Thus, to reduce this threat, Benge (2012) redesigned her study by using a counter-balanced design, namely, a Latin square design (Maxwell & Delaney, 1990)—a design that is underutilized in educational research. Use of a Latin square design made her mixed research design in general and her quantitative research design in particular much more rigorous.

### *Post-Analysis Considerations*

In the final chapter of her dissertation (i.e., Chapter 5)—which was written after she had analyzed all of her quantitative and qualitative data and conducted mixed analyses (Onwuegbuzie & Combs, 2010; Onwuegbuzie & Teddlie, 2003)—Benge (2012) provided an extremely extensive discussion of the *actual* limitations that ended up manifesting themselves in her dissertation research study. Specifically, she discussed the actual limitations of her findings that pertained to her quantitative research phase/component (i.e., internal validity and external validity; cf. Tables 7-8), qualitative research phase/component (i.e., internal credibility and external credibility; cf. Table 9, Appendix E), and mixed research phase/component (mixed legitimation; cf. Table 10, Appendix F). The most useful component of these post-analysis legitimation tables (i.e., Tables 7-10, Appendices D-F) is the last column, which summarizes the steps that were taken to reduce the impact of each threat, thereby increasing validity/credibility/legitimation. For example, in Benge's (2012) dissertation, to minimize researcher bias pertaining to the qualitative phase of her study, she engaged in written self-reflection prior to and throughout the interviewing process; she was continually debriefed by one of her committee co-chairpersons; she included her self-reflection as part of her findings; and she utilized member checking with the teachers in order to confirm her impressions. The important point here is that such deep reflection and extensive discussion needs to take place during the data interpretation stage of the dissertation process so that the researcher can place her/his findings in the most appropriate context.

### **Conclusions**

As posited by Collins, Onwuegbuzie, and Johnson (2012, p. 855), the assessment of legitimation should represent “a continuous, iterative, holistic, and synergistic” process. By continuous, we mean that the assessment of legitimation should occur throughout the (dissertation) research process. By iterative, we mean that the process for assessing legitimation is recursive. That is, any or even all of the legitimation phases can be repeated as many times as is deemed appropriate. By holistic, we mean that legitimation criteria used should be comprehensive—such as those conceptualized by Onwuegbuzie (2003), Onwuegbuzie and Leech (2007), and Onwuegbuzie and Johnson (2006). Finally, by synergistic, we mean that the legitimation approach should reflect collaboration and cooperation among all researchers, between the researcher(s) and participant(s), and between the researcher(s) and other people who have an influential role in the student's dissertation research process (e.g., mentors, advisors, dissertation chairs/supervisors, gatekeepers, stakeholders). In addition, as mentioned earlier, the legitimation process should promote research that is both warranted and transparent.

In order to achieve a legitimation process that is continuous, iterative, holistic, synergistic, warranted, and transparent, we suggest that researchers in general and dissertation researchers in particular undergo a series of what Onwuegbuzie, Leech, and Collins (2008) refer to as debriefing interviews. Specifically, debriefing interviews involve a

peer debriefer formally interviewing the researcher at multiple phases of the study. The debriefer should be a person who is knowledgeable about the research process, who possesses good interviewing skills, and who understands the research topic.

As a guide, Onwuegbuzie et al. (2008) developed an array of open-ended debriefing interview questions that the debriefer could ask the researcher.

Benge (2012) underwent a series of interviews at several phases of her dissertation process. In addition to asking many questions presented by Onwuegbuzie et al. (2008), the debriefer, who was one of her co-chairs, asked several questions that helped her reflect on legitimation issues, which added rich and thick descriptions of Benge's (2012) reflective process and the legitimation strategies that came to the fore in her dissertation study. Moreover, the debriefing interviews helped inform the tables and discussions that appear in the appendices below. Thus, we recommend that chair/supervisors, mentors, research methodologists, and the like help students reflect on legitimation issues that occur throughout their dissertation studies by interviewing them at multiple stages of the research process. Frels and Onwuegbuzie (2012) also documented the utility of the debriefing interviews, declaring that "The [debriefing] interviews were helpful in increasing her [the dissertation student's] awareness of researcher responsibility and provided her the opportunity to share her thoughts and feelings as the study progressed" (p. 23).

In any case, we hope that Benge's (2012) exemplar is helpful for students conducting and writing up their dissertation research, whether it represents a quantitative research-based, qualitative research-based, or mixed research-based dissertation. Also, we hope that this exemplar is useful for instructors of research methodology courses, chairs/supervisors, mentors, and others who play a role in helping students negotiate the dissertation process by providing them with a comprehensive framework for identifying and addressing legitimation issues.

## References

- American Educational Research Association. (2006). Standards for reporting on empirical social science research in AERA publications. *Educational Researcher*, 35(6), 33–40.
- Atkinson, R. C. (1975). Mnemotechnics in second-language learning. *American Psychologist*, 30, 821-828. <http://dx.doi.org/10.1037/h0077029>
- Beck, I. L., McKeown, M. G., & Kucan, L. (2002). *Bringing words to life: Robust vocabulary instruction*. New York, NY: Guilford.
- Benge, C. L. (2012). *Effect of cartoon mnemonics and revised definitions on the acquisition of tier-two vocabulary words among selected fifth-grade students*. Unpublished doctoral dissertation, Sam Houston State University, Huntsville, TX.

- Burchers, S., Burchers, M., & Burchers, B. (1997). *Vocabulary cartoons: Learn hundreds of new SAT words fast with easy memory techniques* (2nd ed.). Punta Gorda, FL: New Monic Books.
- Burchers, S., Burchers, M., & Burchers, B. (2000). *Vocabulary cartoons II: Learn hundreds more new SAT words fast with easy memory techniques*. Punta Gorda, FL: New Monic Books.
- Campbell, D. T. (1957). Factors relevant to the validity of experiments in social settings. *Psychological Bulletin*, *54*, 297-312. <http://dx.doi.org/10.1037/h0040950>
- Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Chicago, IL: Rand McNally.
- Collins, K. M. T., Onwuegbuzie, A. J., & Johnson, R. B. (2012). Securing a place at the table: A review and extension of legitimation criteria for the conduct of mixed research. *American Behavioral Scientist*, *56*, 849-865. <http://dx.doi.org/10.1177/0002764211433799>
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation. Design and analysis issues for field settings*. Chicago, IL: Rand McNally.
- Curtis, S., Gesler, W., Smith, G., & Washburn, S. (2000). Approaches to sampling and case selection in qualitative research: Examples in the geography of health. *Social Science and Medicine*, *50*, 1001-1014. [http://dx.doi.org/10.1016/S0277-9536\(99\)00350-0](http://dx.doi.org/10.1016/S0277-9536(99)00350-0)
- Dale, E. (1965). Vocabulary measurement: Techniques and major findings. *Elementary English*, *42*, 895-901; 498.
- Daniel, L. G., & Onwuegbuzie, A. J. (2000, November). *Toward an extended typology of research errors*. Paper presented at the annual conference of the Mid-South Educational Research Association, Bowling Green, KY.
- Eisner, E. W. (1991). *The enlightened eye: Qualitative inquiry and the enhancement of educational practice*. New York, NY: Macmillan.
- Flick, U. (1998). *An introduction to qualitative research: Theory, method and applications*. London, England: Sage.
- Frels, R. K., & Onwuegbuzie, A. J. (2012). Interviewing the interpretive researcher: An impressionist tale. *The Qualitative Report*, *17*(Art. 60), 1-27. Retrieved from <http://www.nova.edu/ssss/QR/QR17/frels.pdf>
- Harzing, A. W. K. (2009, January). *Publish or perish*. Retrieved from [www.harzing.com/pop.htm](http://www.harzing.com/pop.htm)
- Johnson, R. B., & Christensen, L. (2010). *Educational research: Quantitative, qualitative, and mixed approaches* (4th ed.). Thousand Oaks, CA: Sage.

- Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33(7), 14-26. <http://dx.doi.org/10.1177/1558689806298224>
- Kvale, S. (1995). The social construction of validity. *Qualitative Inquiry*, 1, 19-40. <http://dx.doi.org/10.1177/107780049500100103>
- Lather, P. (1986). Issues of validity in openly ideological research: Between a rock and a soft place. *Interchange*, 17, 63-84. <http://dx.doi.org/10.1007/BF01807017>
- Lather, P. (1993). Fertile obsession: Validity after poststructuralism. *Sociological Quarterly*, 34, 673-693. <http://dx.doi.org/10.1111/j.1533-8525.1993.tb00112.x>
- Leech, N. L., & Onwuegbuzie, A. J. (2009). A typology of mixed methods research designs. *Quality & Quantity: International Journal of Methodology*, 43, 265-275. <http://dx.doi.org/10.1007/s11135-007-9105-3>
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.
- Mastropieri, M. A., & Scruggs, T. E. (1998). Enhancing school success with mnemonic strategies. *Intervention in School & Clinic*, 33, 201-208. <http://dx.doi.org/10.1177/105345129803300402>
- Maxwell, J. A. (1992). Understanding and validity in qualitative research. *Harvard Educational Review*, 62, 279-299.
- Maxwell, J. A. (2005). *Qualitative research design: An interactive approach* (2nd. ed.). Newbury Park, CA: Sage.
- Maxwell, S. E., & Delaney, H. D. (1990). *Designing experiments and analyzing data: A model comparison approach*. Belmont, CA: Wadsworth.
- McKeown, M. G. (1993). Creating effective definitions for young word learners. *Reading Research Quarterly*, 28, 16-31. <http://dx.doi.org/10.2307/747814>
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage.
- Miller, G. A., & Gildea, P. M. (1987). How children learn words. *Scientific American*, 257, 94-99. <http://dx.doi.org/10.1038/scientificamerican0987-94>
- Morse, J. M. (1995). The significance of saturation. *Qualitative Health Research*, 5, 147-149. <http://dx.doi.org/10.1177/104973239500500201>
- Mundfrom, D. J., Shaw, D. G., Thomas, A., Young, S., & Moore, A. D. (1998, April). *Introductory graduate research courses: An examination of the knowledge base*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.
- Onwuegbuzie, A. J. (2003). Expanding the framework of internal and external validity in quantitative research. *Research in the Schools*, 10(1), 71-90.

- Onwuegbuzie, A. J. (2012, February). *Common analytical and interpretational errors in quantitative, qualitative, and mixed research*. Paper presented at the annual meeting of the Southwest Educational Research Association, New Orleans, LA.
- Onwuegbuzie, A. J., & Combs, J. P. (2010). Emergent data analysis techniques in mixed methods research: A synthesis. In A. Tashakkori & C. Teddlie (Eds.), *Sage Handbook of mixed methods in social and behavioral research* (2nd ed., pp. 397-430). Thousand Oaks, CA: Sage.
- Onwuegbuzie, A. J., & Daniel, L. G. (2005). Editorial: Evidence-based guidelines for publishing articles in *Research in the Schools* and beyond. *Research in the Schools*, 12(2), 1-11.
- Onwuegbuzie, A. J., Dickinson, W. B., Leech, N. L., & Zoran, A. G. (2009). Toward more rigor in focus group research: A new framework for collecting and analyzing focus group data. *International Journal of Qualitative Methods*, 8, 1-21.
- 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. (2004). Post-hoc power: A concept whose time has come. *Understanding Statistics*, 3, 201-230. [http://dx.doi.org/10.1207/s15328031us0304\\_1](http://dx.doi.org/10.1207/s15328031us0304_1)
- Onwuegbuzie, A. J., & Leech, N. L. (2007). Validity and qualitative research: An oxymoron? *Quality & Quantity: International Journal of Methodology*, 41, 233-249. <http://dx.doi.org/10.1007/s11135-006-9000-3>
- Onwuegbuzie, A. J., & Leech, N. L., & Collins, K. M. T. (2008). Interviewing the interpretive researcher: A method for addressing the crises of representation, legitimation, and praxis. *International Institute for Qualitative Methodology*, 7, 1-17.
- Onwuegbuzie, A. J., Slate, J. R., Leech, N. L., & Collins, K. M. T. (2009). Mixed data analysis: Advanced integration techniques. *International Journal of Multiple Research Approaches*, 3, 13-33. <http://dx.doi.org/10.5172/mra.455.3.1.13>
- Onwuegbuzie, A. J., & Teddlie, C. (2003). A framework for analyzing data in mixed methods research. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 351-383). Thousand Oaks, CA: Sage.
- Sandelowski, M. (2008). Theoretical saturation. In L. M. Given (Ed.), *The Sage encyclopedia of qualitative methods* (Vol. 1, pp. 875-876). Thousand Oaks, CA: Sage.
- Saumure, K., & Given, L. M. (2008). Data saturation. In L. M. Given (Ed.), *The Sage encyclopedia of qualitative methods* (Vol. 1, pp. 195-196). Thousand Oaks, CA: Sage.
- Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Newbury Park, CA: Sage.
- Yin, R. K. (2009). *Case study research: Design and methods* (4th ed.). Thousand Oaks, CA: Sage.

---

## Notes

Note 1. As recommended by Onwuegbuzie and Johnson (2006), we believe that use of the term *legitimation* is much more appropriate because this term is acceptable to both quantitative and qualitative researchers—unlike the term *validity*, which, although used by virtually all quantitative researchers, is criticized by many researchers as representing a quantitative concept. Thus, as noted by Onwuegbuzie and Johnson (2006), the term *legitimation* represents “bilingual nomenclature” (p. 48).

Note 2. By mixed research, we are referring to what is more commonly known as mixed methods research. However, as recommended by several authors (e.g., Johnson & Onwuegbuzie, 2004), we prefer use of the term *mixed research* instead of the term *mixed methods research* because the latter term might suggest that this research approach only involves the mixing of *methods*, when, indeed, this research approach involves mixing methodologies, which represent a broad approach to scientific inquiry, including potentially the mixing of mental models in general and philosophical assumptions and stances in particular.

Table 1: Typology of Mixed Methods Legitimation Types

Legitimation Type	Description
Sample Integration	The extent to which the relationship between the quantitative and qualitative sampling designs yields quality meta-inferences.
Inside-Outside	The extent to which the researcher accurately presents and appropriately utilizes the insider's view and the observer's view for purposes such as description and explanation.
Weakness Minimization	The extent to which the weakness from one approach is compensated by the strengths from the other approach.
Sequential	The extent to which one has minimized the potential problem wherein the meta-inferences could be affected by reversing the sequence of the quantitative and qualitative phases.
Conversion	The extent to which the quantitizing or qualitzing yields quality meta-inferences.
Paradigmatic mixing	The degree to which the mixed researcher reflects on, understands, and documents his or her 'integrated' mixed research philosophical and methodological paradigm, including his or her epistemological, ontological, axiological, methodological, and rhetorical beliefs about mixed research.
Commensurability	The extent to which the meta-inferences made in a mixed research study reflect a mixed worldview. It is based on the cognitive process of Gestalt switching (to "see" fully and understand the different perspectives) and integration into a new "mixed" or "multi-lens" worldview. This new viewpoint is not possible for some individuals without extensive training, but can usually be obtained via an open-minded "mixed team" of researchers.
Multiple Validities	The extent to which addressing legitimation of the quantitative and qualitative components of the study results from the use of quantitative, qualitative, <i>and</i> mixed validity types, yielding high quality meta-inferences.
Political	The extent to which a mixed researcher appropriately addresses the interests, values, and standpoints of multiple stakeholders in the research process. One should be extra sensitive to the needs of stakeholders with minimal power and voice.

Table 1 was adapted from Onwuegbuzie and Johnson (2006). Reprinted with kind permission of the Mid-South Educational Research Association and the Editors of *Research in the Schools*.



Table 2: Latin Square (i.e., Standard Squares) for Four Vocabulary Conditions and Four Groups

Week	Groups			
	School 1	School 2	School 3	School 4
1	Traditional definition only	Revised definition only	Traditional definition plus cartoon	Revised definition plus cartoon
2	Traditional definition plus cartoon	Traditional definition only	Revised definition plus cartoon	Revised definition only
3	Revised definition only	Revised definition plus cartoon	Traditional definition only	Traditional definition plus cartoon
4	Revised definition plus cartoon	Traditional definition plus cartoon	Revised definition only	Traditional definition only

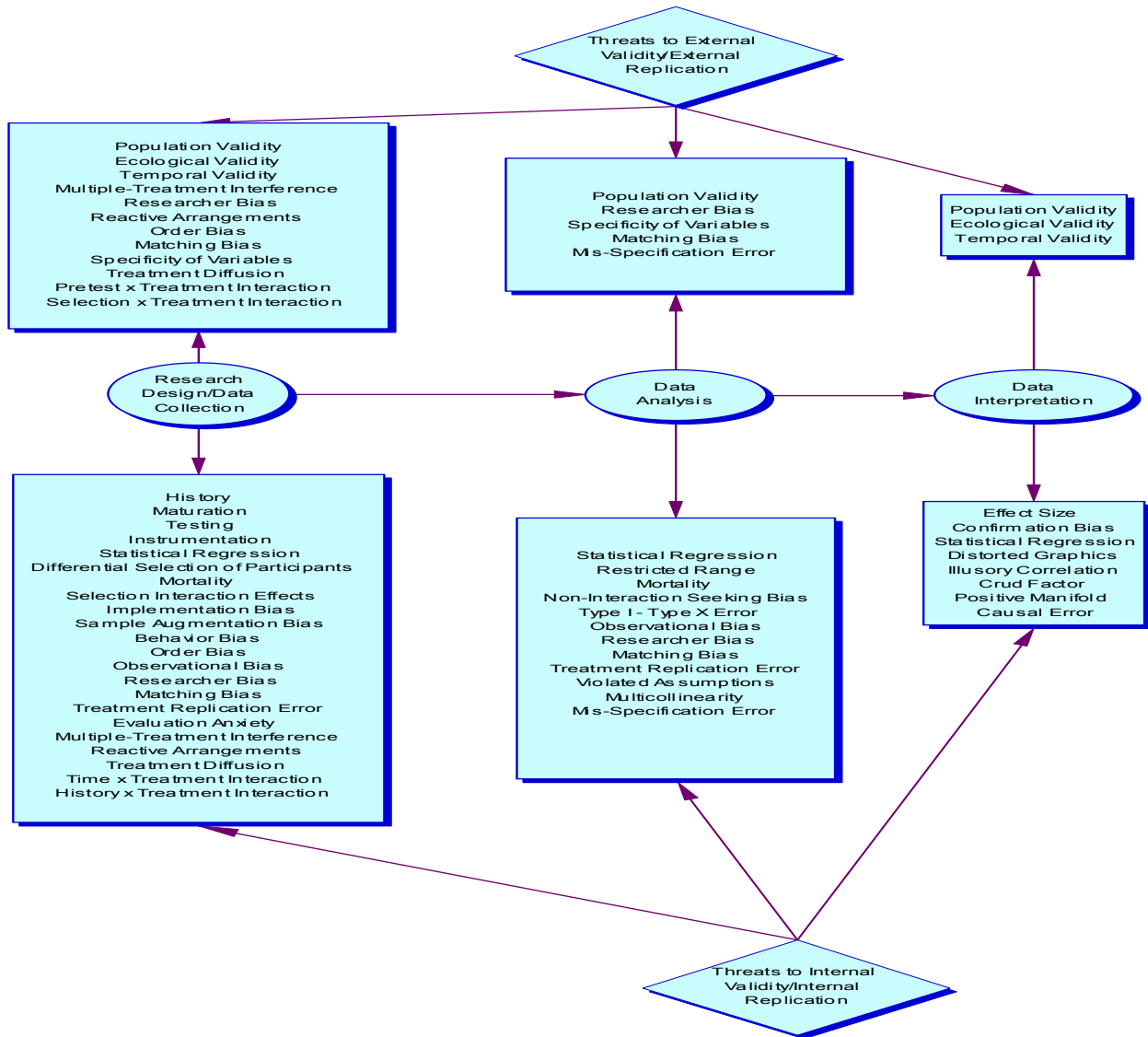


Figure 1: Onwuegbuzie's (2003) Quantitative Legitimation Model. This figure was adapted from Onwuegbuzie (2003). Reprinted with kind permission of the Mid-South Educational Research Association and the Editors of *Research in the Schools*.

Stage of the Research Process	Type of threat	Specific Validity Threat	Description
Research Design/Data Collection	Internal Validity	History	Occurrence of events or conditions that are not related to the intervention or independent variable but that occur at some point during the study to yield changes in the dependent variable (i.e., outcome) such that the longer a study lasts, the more likely that history will pose a threat to validity
		Maturation	Processes that reside within a study participant due, at least in part, to the passage of time, which lead to physical, mental, emotional, and/or intellectual changes (e.g., aging, boredom, fatigue, motivation, learning) that can be incorrectly attributed to the independent variable
		Testing	Changes that might occur in participants' scores obtained on the second administration or post-intervention measure arising, at least in part, from having taken the pre-intervention instrument
		Instrumentation	Occurs when scores yielded from a quantitative measure lack the appropriate level of consistency (i.e., low reliability) and/or validity (i.e., inadequate content-, criterion-, and/or construct-related validity)
		Statistical regression	Occurs when participants are selected because of their extremely low or extremely high scores on some pre-intervention/pre-study measure, wherein there is a tendency for extreme scores to regress, or move toward, the mean on subsequent measures
		Differential selection of participants	Substantive differences between two or more of the (typically intact) comparison groups prior to the implementation of the intervention or the study
		Mortality /attrition	Wherein participants who have been selected to participate in a research study either fail to take part at all or do not participate in every phase of the study (i.e., drop out of the study)—resulting in the findings being biased

Stage of the Research Process	Type of threat	Specific Validity Threat	Description
		Selection interaction effects	One or more threats to internal validity interacting with the differential selection of participants to produce an effect that resembles the intervention effect. For example, a <i>selection by mortality</i> threat can occur if one group has a higher rate of mortality than do the other groups, such that any differences between the groups create factors unrelated to the intervention that are greater as a result of differential mortality than was the case prior to the start of the study. Other threats include selection by history and selection by maturation
		Implementation Bias	Stems from differences in the application of the treatment to the intervention groups as a result of differential motivation, time, training, or resources; inadequate knowledge or ability; poor self-efficacy; implementation anxiety; stubbornness; or poor attitudes among those administering the treatment
		Sample augmentation bias	Being the opposite of mortality, it prevails when one or more individuals that were not selected by the researcher join the study
		Behavior bias	Occurs when a participant has a strong personal bias in favor of or against the intervention prior to the beginning of the study. It is most often a threat when participants are exposed to all levels of a treatment
		Order bias	When multiple interventions are being compared in a research study, such that all participants are exposed to and measured under each and every intervention condition, and the effect of the order of the intervention conditions cannot be distinguished from the effect of the intervention conditions
		Observational bias	Occurs when the data collectors have obtained an insufficient sampling of the behavior(s) of interest
		Researcher bias	Occurs when the researcher has a personal bias in favor of one intervention or technique over another, which might be subconsciously transferred to the participants in such a way that their behavior is affected. In addition to affecting the behavior of participants, the researcher's bias could affect study procedures or even contaminate data collection techniques. It could be active or passive

Stage of the Research Process	Type of threat	Specific Validity Threat	Description
		Matching bias	Occurs when the matching is non-optimal after (a) the researcher uses matching techniques to select a series of groups of individuals (e.g., pairs) who are similar with respect to one or more characteristics, and then assigns each individual within each group to one of the intervention conditions; or (b) once participants have been selected for one of the intervention conditions, the researcher finds matches for each member of this condition and assigns these matched individuals to the other intervention group(s)
		Treatment replication error	Occurs when researchers collect data that do not reflect the correct unit of analysis, with the most common form of treatment replication error being when an intervention is administered once to each group of participants or to two or more classes or other existing groups, yet only individual outcome data are collected—which seriously violates the assumption that each replication of the intervention for each and every participant is independent of the replications of the intervention for all other participants in the study
		Evaluation anxiety	Occurs when the performance of one or more participants is affected unduly by debilitating levels of anxiety such that systematic error is introduced into the measurement
		Multiple-treatment interference	Occurs when the same research participants are exposed to more than one intervention, leading to carryover effects from an earlier intervention, making it difficult to assess the effectiveness of a later treatment

Stage of the Research Process	Type of threat	Specific Validity Threat	Description
		Reactive arrangements	Changes in a participant's response(s) that can occur as a direct result of being cognizant that he/she is participating in a research investigation, comprising the following five major components: (a) the Hawthorne effect (i.e., when participants interpret their receiving an intervention as being given special attention, confounding the effects of the intervention); (b) John Henry effect (i.e., when on being informed that they will be in the control group, participants selected for this condition decide to compete with the treatment or intervention by exerting extra effort during the study period); (c) resentful demoralization (i.e., participants in the control group becoming resentful about not receiving the intervention, interpreting this omission as a sign of being ignored or disregarded, and becoming demoralized, thereby leading to a reduction in effort expended and ensuing decrements in performance or other outcomes); (d) the novelty effect (i.e., increased motivation, interest, or participation on the part of study participants merely because they are undertaking a different or novel task, thereby affecting their responses in a way that is not related to the independent variable); and (e) the placebo effect (i.e., participants in the control group attaining more favorable outcomes [e.g., more positive attitudes, higher performance levels] merely because they believed that they were in the intervention group)
		Treatment diffusion	Occurs when members in different intervention groups communicate with each other, such that some of the treatment seeps out or diffuses into the control group—resulting in the study no longer having two or more distinctly different interventions, but overlapping interventions—thereby violating the assumption of independence
		Time x treatment interaction	Occurs when (a) individuals in one group are exposed to an intervention for a longer period of time than are individuals receiving another intervention in such a way that this differentially affects group members' responses to the intervention, (b) participants in different groups receive their respective interventions for the same period of time, but one of these interventions needs a longer period of time for any positive effects to be realized; or (c) the post-measure for one of the intervention groups is delayed long enough for the effect of the intervention to have changed

Stage of the Research Process	Type of threat	Specific Validity Threat	Description
		History x treatment interaction	Occurs when the interventions being compared experience different history events that differentially affect group members' responses to the intervention
	External Validity	Population validity	Extent to which findings are generalizable from the sample of individuals on which a study was conducted to the population from which the sample was drawn
		Ecological validity	Extent to which findings from a study can be generalized across settings, conditions, variables, and contexts—thereby representing the extent to which findings from a study are independent of the setting or location in which the investigation took place
		Temporal validity	Extent to which research findings can be generalized across time—or the extent that results are invariant across time
		Multiple-Treatment Interference	See above. It is a threat to external validity inasmuch as the order that the treatments or interventions are administered (i.e., sequencing effect) reduces a researcher's ability to generalize findings to the population because generalization typically is limited to the particular sequence of interventions that was administered
		Researcher bias	See above. The more unique the researcher's characteristics and values that influence the data collected, the less generalizable the findings
		Reactive arrangements	See above. The five components of reactive arrangements adversely affect external validity because, in their presence, findings pertaining to the intervention are determined by which of these components prevail
		Order bias	See above. The extent to which findings resulting from a particular order of administration of treatments or interventions cannot be generalized to situations in which the sequence of interventions is different
		Matching bias	See above. Extent to which findings from the matched participants cannot be generalized to the results that would have occurred among individuals in the accessible population for whom a match could not be found (i.e., individuals in the sampling frame who were not selected for the study)

Stage of the Research Process	Type of threat	Specific Validity Threat	Description
		Specificity of variables	Occurs when one of the seven variables are so unique to the study that the findings are not generalizable: type of participants, time, location, circumstance, operational definition of the independent variables, operational definition of the dependent variables, and types of instruments used
		Treatment diffusion	See above. Extent to which the intervention is diffused to other treatment conditions in a unique (i.e., unreplicable) way that threatens the researcher's ability to generalize the findings
		Pretest x treatment interaction	Situations in which the administration of a pretest increases or decreases the participants' responsiveness or sensitivity to the intervention or treatment, thereby making the observed findings pertaining to the pretested group unrepresentative of the effects of the independent variable for the unpretested population from which the study participants were selected—allowing the researcher to generalize the findings to pretested groups but not to unpretested groups
		Selection x treatment interaction	Stems from important pre-intervention differences between intervention groups that emerge because the intervention groups are not representative of the same underlying population—making it unjustifiable for the researcher to generalize the results from one group to another group



Stage of the Research Process	Type of threat	Specific Validity Threat	Description
<b>Data Analysis</b>			
	Internal Validity	Statistical regression	See above. Use of techniques that attempt to control statistically for pre-existing differences among the groups being studied, such as analysis of covariance (ANCOVA), that (a) are unlikely to produce unbiased estimates of the intervention effect and (b) might render the residual scores as uninterpretable
		Restricted range	Inappropriately categorizing continuous variables in non-experimental designs, then using analysis of variance (ANOVA) in an attempt to justify making causal inferences, which, instead, leads to a discarding of relevant variance, in turn, leading to a loss of statistical power and reduced effect size
		Mortality/ attrition	See above. The extent to which sub-sampling from a dataset (e.g., casewise deletion or listwise deletion strategies in the presence of missing data; reducing the size of the largest group(s) to resemble more closely the size of the smaller group(s) to undertake a balanced analysis) introduces or adds bias into the analysis
		Non-interaction seeking bias	Neglecting to assess the presence of interactions when testing hypotheses that leads to a statistical model that does not honor, in the optimal sense, the nature of reality that they want to study

Stage of the Research Process	Type of threat	Specific Validity Threat	Description
		Type I to Type X Error	<ul style="list-style-type: none"> <li>• Type I—falsely rejecting the null hypothesis</li> <li>• Type II—incorrectly failing to reject the null hypothesis</li> <li>• Type III—incorrect inferences about result directionality</li> <li>• Type IV—incorrectly following-up an interaction effect with a simple effects analysis</li> <li>• Type V error—internal replication error measured via incidence of Type I or Type II errors detected during internal replication cycles when using methodologies such as the jackknife procedure</li> <li>• Type VI error—reliability generalization error—measured via linkages of statistical results to characteristics of scores on the measures used to generate results (e.g., when researchers fail to consider differential reliability estimates for subsamples within a data set)</li> <li>• Type VII error—heterogeneity of variance/regression—measured via the extent to which data treated via ANOVA/ANCOVA are not appropriately screened to determine whether they meet homogeneity assumptions prior to analysis of group comparison statistics</li> <li>• Type VIII error—test directionality error—referring to the extent to which researchers express alternative hypotheses as directional yet assess results with two-tailed tests</li> <li>• Type IX error—sampling bias error—assessed via disparities in results generated from numerous convenience samples across a multiplicity of similar studies</li> <li>• Type X error—degrees of freedom error—representing the tendency of researchers using certain statistical procedures (e.g., stepwise procedures) inaccurately to compute the degrees of freedom utilized in these procedures (cf. Daniel &amp; Onwuegbuzie, 2000)</li> </ul>
		Observational bias	See above. Occurs whenever inter-rater reliability or intra-rater reliability of the coding scheme is less than 100%
		Researcher bias	See above. Occurs when a researcher is evaluating open-ended responses, or the like, and allows his/her prior knowledge of the participants to influence the scores given (i.e., halo effect), resulting in findings that are biased
		Matching bias	See above. Occurs when the researcher matches groups after the data on the complete sample have been collected that introduces bias as a result of omitting those who were not matched

Stage of the Research Process	Type of threat	Specific Validity Threat	Description
		Treatment replication error	See above. Occurs when the researchers use an inappropriate unit of analysis, even though data are available for them to engage in a more appropriate analysis, such as analyzing individual data to compare cooperative learning groups instead of analyzing group data, the former leading to the independence assumption being violated, resulting in the inflation of Type I error and effect-size estimates
		Violating assumptions	Stems from a failure to check statistical model assumptions
		Multicollinearity	Failure to assess multicollinearity in multiple regression models when multicollinearity is present
		Mis-Specification error	Failure to specify and to test an appropriate statistical model, including non-interaction seeking bias, discussed above
	External validity	Population validity	See above. Occurs when a researcher analyzes a subset of her/his dataset such that there is a discrepancy between those sampled and those not sampled from the full dataset, leading to findings from the subset that are less generalizable than would have been the case if the total sample had been used
		Researcher bias	See above. Occurs when the particular type of bias of the researcher is so unique as that the findings are not generalizable
		Specificity of variables	See above. Depends on the manner in which the independent and dependent variables are operationalized (e.g., use of local norms vs. national/standardized norms)
		Matching bias	See above. Some researchers match individuals in the different intervention groups just prior to analyzing the data. Matching provides a threat to external validity at this stage if those not selected for matching from the dataset are in some important way different than those who are matched, such that the findings from the selected individuals may not be generalizable to the unselected persons
		Mis-specification error	See above. Occurs when the researcher omits one or more important variables (e.g., interaction terms) from the analysis and it is not clear whether the findings would be the same if the omitted variable(s) had been included

Stage of the Research Process	Type of threat	Specific Validity Threat	Description
Data interpretation			
Internal validity			
		Effect size	Occurs when the researcher fails to report and to interpret confidence intervals and effect sizes, leading to under-interpretation of associated $p$ values when sample sizes are small and the corresponding effect sizes are large, and an over-interpretation of $p$ -values when sample sizes are large and effect sizes are small
		Confirmation bias	Occurs when interpretations and conclusions based on new data are overly consistent with preliminary hypotheses
		Statistical regression	See above. Occurs when a study involves extreme group selection, matching, statistical equating, change scores, time-series studies, or longitudinal studies, and findings from this investigation reflect some degree of regression toward the mean
		Distorted graphics	Occurs when the researcher uses only graphical means to inform interpretations instead of triangulating the graphical data with empirical evaluation
		Illusory correlation	Occurs when the researcher overestimates the relationship among variables that are only slightly related or not related at all
		Crud factor	Occurs when the sample size is so large that the researcher identifies and interprets relationships that are not real but represent statistical artifacts
		Positive manifold	Occurs when the researchers misinterpret relationships between variables by failing to recognize that individuals who perform well on one ability or attitudinal measure tend to perform well on other measures in the same domain
		Causal error	Occurs when the researcher infers causality from a correlation between variables
External validity			
		Population validity	See above. Occurs when researchers over-generalize their conclusions across populations
		Ecological validity	See above. Occurs when researchers over-generalize their conclusions across settings or contexts
		Temporal validity	See above. Occurs when researchers over-generalize their conclusions across time

Figure 2: Onwuegbuzie's (2003) Quantitative Legitimation Model: Threats to internal validity and external validity at the research design/data collection, data analysis, and data interpretation phases of quantitative research studies. This figure was adapted from Onwuegbuzie (2003). Reprinted with kind permission of the Mid-South Educational Research Association and the Editors of *Research in the Schools*.

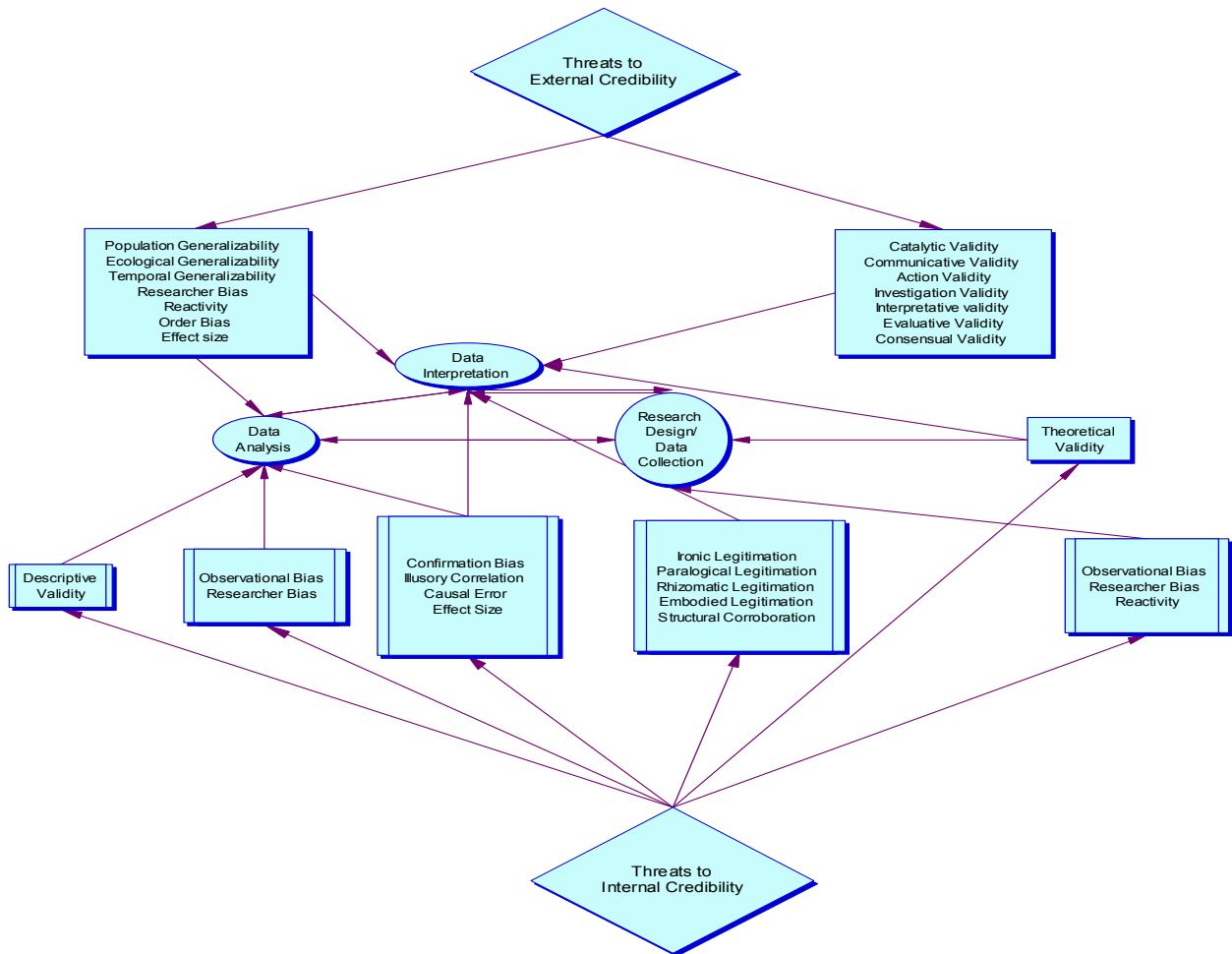


Figure 3: Onwuegbuzie and Leech's (2007) Qualitative Legitimation Model. This figure was adapted from Onwuegbuzie and Leech (2007). Reprinted with kind permission of Springer.

Stage of the Research Process	Type of threat	Specific Validity Threat	Description
Research Design/Data Collection	Internal credibility		
		Descriptive validity	The factual accuracy of the account (e.g., transcripts obtained via an interview, focus group) as documented by the researcher (cf. Maxwell, 1992, 2005)
		Observational bias	Occurs when the researchers have obtained an insufficient sampling of words or behaviors from the study participant(s)—stemming from a lack of persistent observation or prolonged engagement (Lincoln & Guba, 1985)
		Researcher bias	Occurs when the researcher has personal biases or <i>a priori</i> assumptions that he/she cannot bracket (i.e., suspend), which the researcher might subconsciously transfer to the participants in such a manner that their attitudes, behaviors, or experiences are affected; or the researcher could affect study procedures (e.g., ask leading questions in an interview) or even contaminate data collection techniques
		Reactivity	Involves changes in a participant's responses that arise from being aware that he/she is participating in a research investigation
External credibility			
	Observational bias	Occurs when the researcher uses an observational protocol that is unique	
	Order bias	Occurs when the order of the questions that are posed in an interview or focus group or the order in which observations are made unduly affects the dependability and confirmability of the data	

Stage of the Research Process	Type of threat	Specific Validity Threat	Description
<b>Data Analysis</b>			
	Internal credibility		
	Observational bias		Occurs if an insufficient sample of words or behaviors is analyzed from the underlying data
	Researcher bias		Occurs when the researcher has personal biases or <i>a priori</i> assumptions that he/she cannot bracket (i.e., suspend), which unduly affects his/her analysis of the data
	External credibility		
	Catalytic validity		Degree to which a given research study empowers and liberates a research community (Lather, 1986)
	Communicative validity		Involves assessing the legitimation of knowledge claims in a discourse such that legitimation is agreed upon by the collection of researchers (Kvale, 1995)
	Action validity		Justification of the legitimation of the research findings is based on whether or not it works—that is, whether or not the research findings are used by decision makers and other stakeholders (Kvale, 1995)
	Investigation validity		Based on the quality of the researcher's skills, such that legitimation represents the researcher's quality control (e.g., ethicalness)
	Interpretive validity		Extent to which a researcher's interpretation of an account represents an understanding of the perspective of the individuals or group(s) under study and the meanings attached to their words and actions (Maxwell, 1992, 2005)
	Evaluative validity		Extent to which an evaluation framework can be applied to the objects of study, rather than a descriptive, interpretive, or explanatory one (Maxwell, 1992)
	Consensual validity		Based on the opinion of others, with "an agreement among competent others that the description, interpretation, and evaluation and thematics of an educational situation are right" (Eisner, 1991, p. 112)

Stage of the Research Process	Type of threat	Specific Validity Threat	Description
Data interpretation	Internal credibility	Researcher bias	Occurs when the researcher has personal biases or <i>a priori</i> assumptions that he/she cannot bracket (i.e., suspend), which unduly affects his/her interpretations of the findings
		Confirmation bias	Occurs when interpretations and conclusions based on new data are overly congruent with <i>a priori</i> hypotheses, and when there is at least one plausible rival explanation to the underlying findings that might be demonstrated to be superior if the researcher maintained an open mind when interpreting data
		Illusory correlation	Occurs when the researcher identifies a relationship among events, people, and the like, when no such relationship actually exists
		Causal error	Occurs when the researcher provides causal explanations for phenomena without attempting to verify such interpretations
		Effect size	Occurs when the researcher uses quantitative-based terms such as <i>many, most, frequently, several</i> , but does not justify these terms by using some form of quantitative analysis (i.e., effect size) such as counting
		Ironic legitimation	Based on the assumption that there are multiple realities of the same phenomenon such that the truth value of the research depends on its capacity to reveal co-existing opposites (Lather, 1993)
		Paralogical legitimation	Represents that aspect of legitimation that reveals paradoxes (Lather, 1993)
		Rhizomatic legitimation	Arises from mapping data and not only from describing data (Lather, 1993)
		Voluptuous/ Embodied legitimation	Represents the extent to which the researcher's level of interpretation exceeds her/his knowledge base stemming from the data (Lather, 1993)
		Structural Corroboration	The extent to which the researcher utilizes multiple types of data to support or to contradict the interpretation



Stage of the Research Process	Type of threat	Specific Validity Threat	Description
	External credibility		
		Population generalizability	Occurs when researchers over-generalize their findings across populations
		Ecological generalizability	Occurs when researchers over-generalize their findings across settings or contexts
		Temporal generalizability	Occurs when researchers over-generalize their findings across time
		Reactivity	Involves changes in a participant's responses that arise from being aware that he/she is participating in a research investigation that are so unique that it affects the transferability of the findings
		Order bias	Occurs when the order of the questions that are posed in an interview or focus group schedule or the order in which observations are made unduly affects the transferability of the findings
		Effect size	Occurs when the researcher bases interpretations on quantitative-based terms such as <i>many</i> , <i>most</i> , <i>frequently</i> , <i>several</i> , but does not justify these terms by using some form of quantitative analysis (i.e., effect size) such as counting

Figure 4: Onwuegbuzie and Leech's (2007) Qualitative Legitimation Model: Threats to internal credibility and external credibility at the research design/data collection, data analysis, and data interpretation phases of qualitative research studies. This figure was adapted from Onwuegbuzie and Leech (2007). Reprinted with kind permission of Springer.

## Appendix A

Quantitative Dissertations or Quantitative Phase of Mixed Research Dissertations: Threats to Internal Validity and External Validity at the Planning Stage of the Research Process

### Limitations

Several possible threats to internal and external validity of the findings were identified and will be presented as possible limitations to this study in keeping with my pragmatist stance. Threats to the internal and external validity of the findings stemming from the quantitative components will be discussed first. This will be followed by the threats to internal and external credibility and legitimation of the findings stemming from the qualitative and mixed methods components, respectively.

**Quantitative phase.** Onwuegbuzie (2003c) identified 22 threats to internal validity and 12 threats to external validity in quantitative inquiry that occur at the research design/data collection stage and 21 threats to internal validity and five threats to external validity that occur at the data analysis stage; also, seven possible threats to internal validity and three threats to external validity were identified at the data interpretation stage.

***Threats to internal validity at the research design/data collection stage.*** Springer (2010) defines validity as “the extent to which interpretation of the text scores is appropriate, in light of existing evidence and theory” (p. 152). Internal validity refers to the researcher’s ability to make an inference that a causal relationship exists between variables (Cook & Campbell, 1979). For the current study, eight possible threats to internal validity were identified at the research design/data collection stage of the quantitative phase: (a) differential selection of participants, (b) history, (c) mortality, (d) behavior bias, (e) order bias, (f) implementation bias, (g) instrumentation, and (h) multiple-treatment interference (Onwuegbuzie, 2003c).

Differential selection of participants (i.e., selection bias) occurs when there are pre-existing differences among the comparison groups that usually exist when pre-formed groups are utilized as the basis of comparison (Onwuegbuzie, 2003c). Onwuegbuzie (2003c) identified selection bias as a threat in most educational studies because relatively few researchers who utilize whole class groups are able to randomly select the students for the sake of the study. Selection bias was a threat to this study because previously formed, intact classrooms were utilized, and all students within the group who return permission forms were included. However, this study utilized a counter-balanced design that provides for all participants to be exposed to all experimental conditions (Gay, Mills, & Airasian, 2008; Springer, 2010). Because all participants were exposed to all four experimental conditions, they served as their own control group, which should have addressed the possible threat due to selection bias (Gay et al., 2008; Johnson & Christensen, 2010).

A threat due to history occurs when any unplanned event occurs over the course of the study that might or might not have an unintended impact on the outcome measures being utilized (Onwuegbuzie, 2003c). This threat posed a potential threat in the present study because of the time lapse between when the pre-experiment measure was administered in

order to determine the vocabulary words to be taught and when each of the four experimental measures were conducted. Additionally, because students participated in all conditions at varying points over the course of a 4-week period, it was possible that, in the interim, some or all students might have been exposed to one or more of the items to be taught, thereby having an impact on one of the measures that were being utilized. This could have had an impact on the delayed recall measure as well. Although no control can be placed on what occurs outside of the school day, because fifth-grade students were taught by teams of teachers, teachers within the team were notified of the words to be taught within the context of the study so that they refrained from including them in any classroom study.

Mortality is a potential threat to a study when participants in the study either drop out of the study or fail to participate in one or more experimental conditions or measures being administered and, as a result, the differences among the groups cannot accurately be attributed to the experimental treatment (Onwuegbuzie, 2003c). The high transience rate of many of the students in the district where the study was conducted made mortality a threat to this study. In addition, because children are unpredictable, there was a possibility that they might have refused to participate in the assessment measures at any point during the study. However, assuming that any attrition occurred randomly, the target sample was large enough to withstand the loss of a number of students in each group without yielding a significant loss of statistical power. In addition, before the study began, pre-existing group differences were determined and the class groups were matched based on scores from the Texas Assessment of Knowledge and Skills (TAKS; Texas Education Agency, 2008-2009). Participant scores were removed from consideration to keep the groups evenly matched when attrition made it necessary.

All participants were exposed to all levels of the treatment; thus, the study was subject to behavior bias. Behavior bias refers to the pre-existing personal biases that a participant might have toward one or more of the interventions (Onwuegbuzie, 2003c). Although few students were exposed to the use of cartoon mnemonics to learn vocabulary, all of them had used definitions as a vocabulary learning activity prior to the study. However, to determine pre-existing biases prior to the study, students completed a survey, the Vocabulary Learning Strategies/Attitudes Scale (VLSAS), which helped me determine what vocabulary learning strategies participants had utilized previously and their attitudes toward their use. Data gleaned from the VLSAS were utilized to generate questions geared toward identifying any possible selection bias.

Because all participants in the study were exposed to all four experimental conditions, order bias also posed a threat to internal validity. Order of intervention poses a threat when “the effect of the order of the intervention conditions cannot be separated from the effect of the intervention conditions” (Onwuegbuzie, 2003c, p. 77). To control for order bias, a counter-balanced design was utilized whereby the order of the interventions was randomly varied (Gay & Airasian, 2003; Johnson & Christensen, 2010; Onwuegbuzie, 2003c). Indeed, a Latin square was used to counter-balance the treatments—specifically, one of the 576 possible Latin squares that are available when four treatments are involved (Maxwell & Delaney, 1990). Moreover, a *standard square* was used that eliminates systematic patterns

that arise when one treatment always follows another treatment in the Latin square. The order effect was tested by including this variable as a covariate/blocking factor in any statistical comparison of the interventions (Maxwell & Delaney, 1990), such that the data analysis matched the design, as recommended by Fisher (1935/1971) 75 years ago.

Onwuegbuzie (2003c) identified implementation bias as possibly one of the “most frequent and pervasive threats to internal validity at the data collection stage in intervention studies” (p. 77). Implementation bias occurs when teachers are utilized to implement the intervention. Onwuegbuzie (2003c) suggested that teachers’ implementation of the intervention protocol might vary as a function of variables such as lack of motivation, time, and training. As a result of this implementation differential, fidelity to the intervention protocol might have been minimal, resulting in outcomes that were not trustworthy. To minimize implementation bias, I observed one class period for each participating teacher during the 4-week intervention period. A copy of the intervention protocol sheet was used by the teacher to document deviations from the protocol so that I was able to calculate a percentage to determine the fidelity to the protocol. In addition, the teachers’ behaviors were observed to determine the extent to which their dispositions in implementing the interventions potentially affected fidelity.

Six measures were utilized in this study, all of which might have led possibly to instrumentation bias. Scores that “lack the appropriate level of consistency and/or validity” (Onwuegbuzie, 2003c, p. 76) result in instrumentation bias. The pre-experiment measure that was used to determine the words that were included in the treatments is a checklist based on Dale’s (1965) four stages of vocabulary knowledge. This researcher-created instrument was not administered to any other group prior to the study; therefore, there were no previous reliability estimates to use as a basis for comparison. One of the measures utilized to determine differences in groups (i.e., TAKS). The VLSAS was previously piloted and a high degree of internal consistency was reported (Benge, 2009). However, the students involved in the piloting of the VLSAS were ninth- and 10th-grade students, and the wording of the questions was altered for this study so that it was a better fit for the fifth-grade students. The scoring of the sentence and questioning tasks, which were used to determine the level of students’ word understanding, and the scoring of the definition recall items, which was used to assess students’ ability to recall word definition long term, posed the biggest threat to internal consistency. These two measures, either a sentence completion task or a questioning task, depending on the condition, were scored by me. To control for instrumentation bias, a rubric was generated a priori, and 25% of the question task items were randomly selected and scored by a second scorer (McKeown, 1993) and 100% of the sentence task items and definition recall items were scored by a second scorer to determine inter-scorer reliability.

Multiple treatment interference, which occurs when participants in a study are exposed to multiple treatments (Onwuegbuzie, 2003c), was a potential threat to internal validity. To address this threat, as suggested by Onwuegbuzie (2003c), treatments were randomly counterbalanced among the groups. In addition, the treatments were implemented at intervals of 1 week to minimize the threat to internal validity due to multiple treatments (Onwuegbuzie, 2003c).

*Threats to internal validity at the data analysis stage.* One possible threat to internal validity was identified at the data analysis stage. *Observational bias* occurs when data are rated or coded by more than one researcher and less than 100% agreement is reached in the rating of the data (Onwuegbuzie, 2003c). Also, it can pose a potential threat to internal validity when only one coder is involved in coding the data, and there is no basis for comparison to determine inter-rater reliability (Onwuegbuzie, 2003c). As previously discussed, procedures were in place to determine inter-rater reliability using either 100% of student responses or a random selection of student task responses and a second scorer. Table 3 presents the threats to internal validity and their descriptions at the quantitative phase of the study.

Table 3: Threats to Internal Validity at the Quantitative Phase

Stage of Design/ Limitation	Description	Manifestations in Current Study
Research design/ data collection:		
Differential selection of participants	Bias relating to the use of pre-existing groups; selection bias	<ul style="list-style-type: none"> <li>• Pre-formed (i.e., intact) classes were used to form groups.</li> </ul>
History	Relates to an unplanned event that has an impact on the study	<ul style="list-style-type: none"> <li>• A time lapse of 9 weeks occurred between the administration of the Vocabulary Knowledge Check Sheet and the administration of the delayed recall test, allowing for students to be unduly exposed to the words being taught in the intervention.</li> </ul>
Mortality	Occurs when participants' dropping out or failing to participate in the study has an unintended impact on the study	<ul style="list-style-type: none"> <li>• Attrition might have ensued from the high transient rate of the population of students.</li> <li>• Some children might have refused to participate after initially agreeing or might have begun to take the exercise less seriously as the study progressed.</li> </ul>
Behavior bias	Pre-existing personal biases of the participants that have an unintended impact on the results	<ul style="list-style-type: none"> <li>• Participants might have had a preconceived preference to one of the interventions.</li> </ul>

Stage of Design/ Limitation	Description	Manifestations in Current Study
Research design/ data collection:		
Order bias	Occurs when the effect of the order of the interventions have an unintended impact on the results	<ul style="list-style-type: none"> <li>All participants took part in all conditions, making it possible for the order of intervention to become a threat.</li> </ul>
Implementation bias	Occurs when someone other than the researcher implements the intervention and deviates from the protocol	<ul style="list-style-type: none"> <li>All participants took part in all conditions making it possible for the order of intervention to become a threat.</li> </ul>
Instrumentation	Occurs when scores lack consistency or validity	<ul style="list-style-type: none"> <li>Six measures were used during the study, with five measures being researcher-created; therefore, the appropriateness of these five instruments prior to administration was unknown.</li> </ul>
Multiple treatment interference	Occurs when participants in a study are included in multiple treatments	<ul style="list-style-type: none"> <li>All students were involved in all four vocabulary conditions and the effect of one treatment might have carried over to the next treatment (i.e., <i>carryover effect</i>).</li> </ul>

Stage of Design/ Limitation	Description	Manifestations in Current Study
Research design/ data collection:		
Observational bias	Occurs when data are rated or coded by more than one researcher and less than 100% agreement is attained	<ul style="list-style-type: none"> <li>• Only one rater scored 100% of student responses of the questioning task activity; 25% were randomly selected for a second scorer; however, for 75% of the questioning response data, there was no inter-rater reliability information.</li> <li>• Only one coder analyzed all student open-ended response data to be quantitized; a second coder coded 25% of data selected randomly; however, for 75% of the data, there was no inter-rater reliability information.</li> </ul>

This table was adapted from Onwuegbuzie (2003). Reprinted with kind permission of the Mid-South Educational Research Association and the Editors of *Research in the Schools*.

*Threats to external validity.* Springer (2010) defines *external validity* as “the extent to which experimental findings can be generalized beyond the original study” (p. 189). *Population validity* and *ecological validity* were identified by Onwuegbuzie (2003c) as two threats to external validity that can never be excluded as possible threats because “all samples, whether random or non-random, are subject to sampling error” (p. 72). Population validity determines the generalizability between the population of participants and the target population (Onwuegbuzie, 2003c). Ecological validity determines the generalizability “across settings, conditions, variables, and contexts” (Onwuegbuzie, 2003c, p. 80). Random selection of school, random assignment of measure (i.e., sentence task or questioning task), and random selection of treatment order were utilized along with a sample size ( $n > 280$ ) for each study that was large enough to achieve a statistical power of .80 with a 95% confidence interval (Onwuegbuzie & Daniel, 2004). However, population and ecological validity must be considered as potential threats to the study because the participating district reported a disproportionate number of minority and English Language Learners (ELLs), as well as students classified as lower socioeconomic, when compared with the general population.

Three other possible threats to external validity were identified at the research design/data collection stage: (a) multiple-treatment interference; (b) order bias, which were discussed previously; and (c) *specificity of variables* (Onwuegbuzie, 2003c). Onwuegbuzie (2003c) determined specificity of variables to be one of the most prevalent threats to external validity at the research design/data collection stage (p. 84). Because the study utilized a sample of fifth-grade students whose demographic data were relatively unique and, therefore, did not represent the general population, specificity of variables was a threat to external validity. In addition, the measures being utilized also were specific to the study. Therefore, as suggested by Onwuegbuzie (2003c), I “operationally define(ed) variables in a way that has meaning outside of the study setting” (p. 81). However, because five of the six measures that were utilized were researcher created and did not have associated standardized norms, “extreme caution” (Onwuegbuzie, 2003c, p. 81) was utilized in generalizing the findings beyond the sample. Possible threats to external validity at the data analysis stage were population validity and specificity of variables (Onwuegbuzie, 2003c), both of which were discussed previously. Table 4 presents the threats to external validity and their descriptions at the quantitative phase of the study.



Table 4: Threats to External Validity at the Quantitative Phase

Stage of Design/ Limitation	Description	Manifestations in Current Study
Research design/ data collection:		
Ecological validity	Determines the generalizability across settings, conditions, variables, and contexts	The district where the study took place had a large number of minority and ELL students as well as those considered representing lower socioeconomic status as compared with the general population.
Multiple-treatment interference	Occurs when participants in a study are included in multiple treatments	All students were involved in all four vocabulary conditions and the effect of one treatment might have carried over to the next treatment.
Order bias	Occurs when the effect of the order of the interventions have an unintended impact on the results	All participants took part in all conditions making it possible that the order that some students receive the condition might have had an impact on the outcomes.
Research design/ data collection / data analysis		
Population validity	Determines the generalizability between the population of participants and the target population	The district where the study was conducted had a disproportionate number of minority and ELL students; additionally, it had a disproportionate number of students who were classified as lower-socioeconomic.
Specificity of variables	Occurs when a population is unique and findings cannot be generalized beyond the sample	The demographic data of the sample were unique and likely did not represent the general population. The measures being utilized were specific to the study.

This table was adapted from Onwuegbuzie (2003). Reprinted with kind permission of the Mid-South Educational Research Association and the Editors of *Research in the Schools*.

## Appendix B

### Qualitative Dissertations or Qualitative Phase of Mixed Research Dissertations: Threats to Internal Credibility and External Credibility at the Planning Stage of the Research Process

**Qualitative phase.** Onwuegbuzie and Leech (2007) documented 14 possible threats to *internal credibility* and 12 possible threats to *external credibility* in qualitative research. Onwuegbuzie and Leech (2007) defined internal credibility as “the truth value, applicability, consistency, neutrality, dependability, and/or credibility of interpretations and conclusions within the underlying setting or group” (p. 234). Consistent with external validity in quantitative research, external credibility “refers to the degree that the findings of a study can be generalized across different populations of persons, settings, contexts, and times” (Onwuegbuzie & Leech, 2007, p. 235). Four threats to the internal and external credibility of the findings stemming from the study were identified: (a) *researcher bias*, (b) *observational bias*, (c) *reactivity*, and (d) *confirmation bias* (Onwuegbuzie & Leech, 2007). The following procedures were followed in an effort to increase credibility: (a) triangulation, (b) leaving an audit trail, (c) checking for researcher effects/clarifying researcher bias, (d) peer debriefing, and (e) determining effect sizes (Onwuegbuzie & Leech, 2007). Specifics are discussed below.

*Researcher bias* can occur when the researcher’s preconceived ideas or biases threaten to interfere with either the researcher’s or participants’ actions, thereby influencing the outcome of the study (Onwuegbuzie, 2003c). Because I, as the researcher, was considered the “primary instrument” (p. 2) for data gathering, analysis, and interpretation in qualitative research, researcher bias could not be avoided (Onwuegbuzie, Leech, & Collins, 2008). Due to my keen interest in and prior experiences with both methods of direct instruction under study, researcher bias was identified as a possible threat to internal credibility and external credibility in the current study (Onwuegbuzie & Leech, 2007). To minimize researcher bias, prior to collecting interview data, I produced a written reflection by engaging in Moustakas’s (1994) four-part self-report of the lived experiences as they pertained to the interventions under review. This allowed me to document, to acknowledge, and to suspend any “prejudgments, biases, and preconceived ideas about things” (Moustakas, 1994, p. 85) as interview data were collected and analyzed and as interpretations of the data were made. Additionally, researcher bias was minimized through the use of *debriefing the interpretive researcher*, which is intended to encourage reflection on the part of researchers (Onwuegbuzie et al., 2008). The debriefing involved a trusted and knowledgeable second party whose focused questions were aimed at uncovering my biases before, during, and after the implementation of the study (Onwuegbuzie et al., 2008).

Observational bias was identified as a potential threat to internal credibility of the findings at the research design/data collection stage. This threat manifests itself when the researcher fails to gather enough data pertaining to either participant’s behaviors or participant’s words (Onwuegbuzie et al., 2008). Two types of qualitative data were gathered:

participants' written self-reports and interview data. In addition, I conducted three extensive interviews of the teacher participants in the tradition of Moustakas's (1994) transcendental phenomenology so that substantial, quality data were gathered.

Reactivity also was determined to be a threat to internal validity and external validity in the study. Reactivity becomes a threat to findings when the participants become aware that they are involved in a research study (Onwuegbuzie, 2003c). This awareness might possibly lead to (a) the *Hawthorne effect* (i.e., a possible effect due to participants' beliefs that they are receiving special attention) and (b) the novelty effect (i.e., the impact on participants as a result of an event or an item being introduced into their environment that is out of the norm) (Onwuegbuzie & Leech, 2007). Because only a small number of students participated in the interview portion of the study, there was a real possibility that these students might have felt singled out for special attention; moreover, students' interviews were audio-taped, and the presence of the tape recorder might have had an impact on these students' responses. However, participants were aware from the beginning that interviews would be conducted, but they did not know that only a select number would be conducted, thereby limiting the impact of any possible Hawthorne effect. Furthermore, in an effort to mitigate a novelty effect, the tape recorder that was utilized was extremely small and inconspicuous.

Confirmation bias, which occurs at the data interpretation stage, refers to researchers' predispositions toward producing interpretations and conclusions when analyzing the new data that are compatible with preconceived beliefs about the topic under review (Greenwald, Pratkanis, Leippe, & Baumgardner, 1986; Onwuegbuzie & Leech, 2007). Procedures that minimized confirmation bias included a researcher-produced written self-reflection (Moustakas, 1994) and my being interviewed as the interpretative researcher (Onwuegbuzie et al., 2008). See the discussion on researcher bias for more discussion about these procedures. Table 5 presents the threats to internal and external credibility and their descriptions.

Table 5: Threats to Internal and External Credibility at the Qualitative Phase

Limitation	Description	Manifestations in Current Study
Researcher bias	Occurs when the researcher has preconceived ideas or biases that threaten the outcomes of the study	<ul style="list-style-type: none"> <li>I have a keen interest in prior history with the use of cartoon mnemonics and revised definitions that might have been a threat to credibility as I engaged in the qualitative portion of the study.</li> </ul>
Observational bias	Occurs when there is a potential for the researcher to fail to collect enough observational data pertaining to a participant's words or behaviors	<ul style="list-style-type: none"> <li>One data analysis procedure relied on the open-ended responses of fifth-grade students who do not always articulate their thoughts well on paper.</li> <li>Another instrument required that I engage in extended interviews with participants. The thickness and richness of the data depended upon the participation level of the students and teachers.</li> </ul>
Reactivity	Occurs when the participants become aware that they are involved in a research study; might lead to the Hawthorne effect or the novelty effect	<ul style="list-style-type: none"> <li>All students were aware that they were participating in a research study.</li> <li>Only select students were involved in the interviews, and this knowledge might have manifested itself into a Hawthorne effect.</li> </ul>
Confirmation bias	The tendency for a researcher to interpret data in a manner that is favorable to his or her preconceived notions of a phenomena	<ul style="list-style-type: none"> <li>I have a keen interest in prior research conducted with the use of cartoon mnemonics and revised definitions and confirmation bias might have become an issue as I analyzed the data.</li> </ul>

This table was adapted from Onwuegbuzie and Leech (2007). Reprinted with kind permission of Springer.

## Appendix C

## Mixed Research Dissertations: Threats to Legitimation at the Planning Stage of the Research Process

**Mixed research phase.** Onwuegbuzie and Johnson (2006) identified nine legitimation types in their typology of legitimation in mixed research. *Multiple validities*, *sequential*, and *conversion* legitimation types were all determined to be potential threats to the study (Onwuegbuzie & Johnson, 2006). Onwuegbuzie and Johnson (2006), proposing the use of a “bilingual nomenclature” (Teddlie & Tashakkori, 2003, p. 12) when differentiating among quantitative, qualitative, and mixed methods, advocate the use of the word *legitimation* when referring to validity of findings in mixed methods research studies. Legitimation in mixed methods research is of critical concern due to the potentially additive effect of the validity threats associated with both quantitative and qualitative research (Onwuegbuzie & Johnson, 2006). Multiple validities legitimation references the need to address all validities surrounding all methods involved in a mixed methods research study, both quantitative and qualitative (Onwuegbuzie & Johnson, 2006). All threats to validity of the findings stemming from all components of the study were addressed. See the discussions earlier for the procedures utilized to control for possible threats to quantitative validity and qualitative credibility.

Sequential legitimation references the need to reduce any impact that the problems associated with the order of the quantitative and qualitative phases might have had on making quality meta-inferences (Onwuegbuzie & Johnson, 2006). In this study, data were collected sequentially. The nature of the design of this study required the sequential collection of the data (e.g., the instructional interventions had to take place before students and teachers could provide insight into how they perceived the effect of the instructional interventions). Therefore, the ability to make meta-inferences depended on the collection of data in a specific order, thereby attenuating the threat to sequential legitimation.

Finally, Onwuegbuzie and Johnson (2006) define conversion legitimation as “the extent to which the quantizing or qualitzing yields quality meta-inferences” (p. 57). This legitimation type posed a threat to the findings of this study due to the quantizing of open-ended response data. I coded all qualitative data; therefore, in order to minimize the possibility of “over-weighting or underweighting emergent themes” (Onwuegbuzie & Johnson, 2006, p. 58) or falsely scoring a student response as correct when it was not, at least 25% of responses were randomly selected for re-coding by a second coder. Table 6 presents the threats to mixed methods legitimation and their descriptions.

Table 6: Threats to Mixed Methods Legitimation

Limitation	Description	Manifestations in the Current Study
Multiple validities legitimation	References the need to address all validities surrounding all methods in a study	<ul style="list-style-type: none"> <li>Multiple threats to validity were evident within the current study</li> </ul>
Sequential legitimation	References the need to reduce the impact that the order of the quantitative and qualitative phases might have on the ability to make meta-inferences	<ul style="list-style-type: none"> <li>All data were gathered sequentially; thus, the findings might have been an artifact of the sequence of phases (i.e., quantitative phase before the qualitative phase)</li> </ul>
Conversion legitimation	The ability to make quality meta-inferences from both quantitative and qualitative data in a study	<ul style="list-style-type: none"> <li>The quantizing of open-ended response data that were generated in the mixed methods analysis posed a threat; for example, counting qualitative data might not have been appropriate</li> </ul>

This table was adapted from Onwuegbuzie and Johnson (2006). Reprinted with kind permission of the Mid-South Educational Research Association and the Editors of *Research in the Schools*.

## Appendix D

## Quantitative Dissertations or Quantitative Phase of Mixed Research Dissertations: Threats to Internal Validity and External Validity at the Interpretation Stage of the Research Process

**Validation/Legitimation of the Research Findings**

As discussed in Chapter I, several possible threats to internal and external validity, internal and external credibility, and mixed methods legitimation in this study were identified. Because they were discussed at length in Chapter I, I only list each threat that was addressed during this study; however, threats that could not be sufficiently addressed over the course of the study are offered as limitations to the study and specifically discussed in relationship to the findings of the study. Tables 1-4, which were included in Chapter I, have been expanded to include not only the manifestation of the possible threats within the study, but also the manner in which I addressed each threat. Threats to the internal and external validity of the quantitative phases will be discussed, followed by threats to internal and external credibility and legitimation of the qualitative and mixed portions, respectively.

**Quantitative phase.** Threats to internal validity and external validity at the research design/data collection stage will be discussed briefly, followed by threats to internal validity and external validity at the data analysis stage. Next, I will discuss threats to internal validity and external validity at the data interpretation stage.

***Threats to internal validity research design/data collection stage.*** Internal validity refers to the researcher's ability to make inferences that a causal relationship exists between variables (Cook & Campbell, 1979). Eight threats to internal validity were identified at the research design stage of the quantitative phase: (a) differential selection of participants, (b) history, (c) mortality, (d) behavior bias, (e) order bias, (f) implementation bias, (g) instrumentation, and (h) multiple-treatment interference (Onwuegbuzie, 2003c). Table 7 presents a brief description of each of the limitations at the quantitative phase, a description of how each manifested itself in this study, and the manner in which I addressed each threat. Although every effort was made to address each of the identified threats over the course of the study, two threats, mortality and implementation bias, warrant further discussion.

Mortality occurs when participants in the study either drop out or fail to participate in one or more experimental conditions or measures (Onwuegbuzie, 2003c). Due to the high transience rate of the district where the study was conducted and the large number of absences, a high mortality rate did occur. However, after all data were collected and students with missing data were eliminated, assuming that the attrition occurred randomly, a sufficient number of students remained to maintain high statistical power. Further, although no students who had agreed to participate within the study refused to participate, teachers reported via interviews that students exhibited boredom and frustration during some of the instructional conditions. Although I randomly counterbalanced the vocabulary instructional conditions, it is possible that students' boredom and/or frustration with one or more of the instructional conditions might have resulted in students not answering questions or doing their best on an

activity; therefore, mortality should be considered as a possible limitation, and the reader should take this into consideration when interpreting the findings of the study.

Implementation bias results when teachers are utilized to implement the intervention. In order to minimize implementation bias, I observed one class period for each participating teacher during the 4-week intervention period and documented any deviations from the protocol. No deviation was noted that would have resulted in implementation bias; however, disruptions within the classroom caused by students were noted in two schools, and a disruption by another teacher was noted in another school during the implementation of the protocol. Although these interruptions did not result in a deviation from the protocol, the distractions to both the students and the participating teachers might have resulted in students not performing their best on the activity. Therefore, I cannot rule out the possibility that implementation bias occurred during this study, and the reader should take this into consideration when interpreting the findings.

***Threats to internal validity at the data analysis stage.*** One possible threat to internal validity was identified at the data analysis stage. Observational bias was identified and measures enacted to neutralize the threat.



Table 7: Threats to Internal Validity at the Quantitative Phase, Manifestations in Current Study, and Method of Increasing Validity

Stage of Design/ Limitation	Description	Manifestations in Current Study	Method of Increasing Internal Validity
Research design/ data collection:			
Differential selection of participants	Bias relating to the use of pre-existing groups; selection bias	<ul style="list-style-type: none"> <li>• Pre-formed (i.e., intact) classes were used to form groups.</li> </ul>	<ul style="list-style-type: none"> <li>• I utilized a counter-balanced design so that all students were exposed to each instructional condition.</li> <li>• I checked for pre-existing differences prior to data analysis and found none.</li> </ul>
History	Relates to an unplanned event that has an impact on the study	<ul style="list-style-type: none"> <li>• A time lapse of 9 weeks occurred between the administration of the Vocabulary Knowledge Check Sheet and the administration of the delayed recall test, allowing for students to be unduly exposed to the words being taught in the intervention.</li> </ul>	<ul style="list-style-type: none"> <li>• I was unable to control for the possibility that the students would be exposed to the words; however, I specifically selected words that would be unlikely to be known or encountered by fifth-grade students.</li> <li>• GT students who might have been more likely to encounter these words were eliminated from data analysis.</li> </ul>
Mortality	Occurs when participants' dropping out or failing to participate in the study has an unintended impact on the study	<ul style="list-style-type: none"> <li>• Attrition might have ensued from the high transient rate of the population of students.</li> <li>• Some children might have refused to participate after initially agreeing or might have begun to take the exercise less seriously as the study progressed.</li> </ul>	<ul style="list-style-type: none"> <li>• A high attrition rate did occur as a result of absences and withdrawals from school; however, I conducted an a priori power analysis to ensure sufficient statistical power.</li> <li>• No child refused to participate; however, given that boredom and frustration were reported by the teachers, I discussed mortality as a possible limitation in the discussion section.</li> </ul>

Stage of Design/ Limitation	Description	Manifestations in Current Study	Method of Increasing Internal Validity
Behavior bias	Pre-existing personal biases of the participants that have an unintended impact on the results	<ul style="list-style-type: none"> <li>Participants might have had a preconceived preference to one of the interventions.</li> </ul>	<ul style="list-style-type: none"> <li>Several of the questions on the VLSAS pertained to particular attributes of the study; no preference for any of the attributes was documented.</li> <li>Teachers indicated during interviews that they were unfamiliar with and had not used any method other than dictionary only.</li> </ul>
Order bias	Occurs when the effect of the order of the interventions have an unintended impact on the results	<ul style="list-style-type: none"> <li>All participants took part in all conditions, making it possible for the order of intervention to become a threat.</li> </ul>	<ul style="list-style-type: none"> <li>An interaction effect was noted in several of the statistical procedures; as a result, I was conservative in my interpretation and did not interpret pairs that were part of any interaction.</li> </ul>
Implementation bias	Occurs when someone other than the researcher implements the intervention and deviates from the protocol	<ul style="list-style-type: none"> <li>All participants took part in all conditions making it possible for the order of intervention to become a threat.</li> </ul>	<ul style="list-style-type: none"> <li>Although I could not totally control for this, I conducted one observation of each teacher during the administration of the protocol; no deviations were noted; however, I did note that, in several classrooms, disruptions occurred. I discussed this as a limitation of my findings.</li> </ul>

Stage of Design/ Limitation	Description	Manifestations in Current Study	Method of Increasing Internal Validity
Instrumentation	Occurs when scores lack consistency or validity	<ul style="list-style-type: none"> <li>Six measures were utilized during the study, with five measures being researcher-created; therefore, the appropriateness of these five instruments prior to administration was unknown.</li> </ul>	<ul style="list-style-type: none"> <li>VLSAS was piloted prior to use.</li> <li>Internal consistencies were reported for all quantitative instrument data/scores.</li> <li>For the research-scored instruments, a second scorer was utilized and the inter-scorer reliability calculated and reported with a straight percentage as well as Krippendorff's alpha</li> </ul>
Multiple treatment interference	Occurs when participants in a study are included in multiple treatments	<ul style="list-style-type: none"> <li>All students were involved in all four vocabulary conditions and the effect of one treatment might have carried over to the next treatment (i.e., <i>carryover effect</i>).</li> </ul>	<ul style="list-style-type: none"> <li>The four instructional interventions were randomly counterbalanced.</li> <li>Interventions were spaced in intervals of 1 week.</li> </ul>
<b>Data analysis:</b>			
Observational bias	Occurs when data are rated or coded by more than one researcher and less than 100% agreement is attained	<ul style="list-style-type: none"> <li>Only one rater scored 100% of student questioning response of the questioning task activity; 25% was randomly selected for a second scorer; however, for 75% of the questioning response data, there will be no inter-rater reliability information.</li> </ul>	<ul style="list-style-type: none"> <li>100% of sentence generation task data were scored by two scorers.</li> <li>25% of the questioning task scores were randomly selected and scored by a second coder.</li> <li>Inter-scorer reliability was reported via a straight percentage in addition to Krippendorff's alpha. A high degree of reliability was noted and reported in Chapter IV.</li> </ul>

This table was adapted from Onwuegbuzie (2003). Reprinted with kind permission of the Mid-South Educational Research Association and the Editors of *Research in the Schools*.

*Threats to external validity.* Three threats to external validity were identified at the design/data collection stage: (a) ecological validity; (b) multiple-treatment interference; and (c) order bias. Every effort was made to address all three of these threats; however, ecological validity, which can occur when the participants are not representative of the general population, remained a threat to the generalizability of the findings. Because two of the schools in this study had a population that was disproportionately Hispanic and the other two schools had a population that was disproportionately African American, ecological validity must be considered a threat. Therefore, readers are cautioned to take this limitation into consideration when generalizing the findings of this study. Multiple-treatment interference and order bias were thoroughly discussed in Chapter I.

Two threats were identified at the research design/data collection and the data analysis stages: population validity and specificity of variables. Population validity is threatened when there are differences between the target population and the population of participants. Specificity of variables occurs when the demographic data of the sample do not represent the general population and the measures being used are unique to the study. Although every effort was made to control for any limitations to this study due to either of these threats, the reader should take these two threats into consideration when generalizing the results of this study. Table 8 presents the threats to external validity at the quantitative phase, a description of how each manifested itself in this study, and the manner in which I addressed each threat.

Table 8: Threats to External Validity at the Quantitative Phase, Manifestations in Current Study, and Method of Increasing Validity

Stage of Design/ Limitation	Description	Manifestations in Current Study	Method of Increasing External Validity
Research design/ data collection:			
Ecological validity	Determines the generalizability across settings, conditions, variables, and contexts	<ul style="list-style-type: none"> <li>The district where the study took place had a large number of minority and ELL students as well as those considered representing lower socioeconomic status as compared with the general population.</li> </ul>	<ul style="list-style-type: none"> <li>Utilized random selection of school, random assignment of measure to teacher, random selection of treatment order, and large sample size.</li> <li>This was addressed in the narrative as a limitation to the generalizability of the findings of this study.</li> </ul>
Multiple-treatment interference	Occurs when participants in a study are included in multiple treatments	<ul style="list-style-type: none"> <li>All students were involved in all four vocabulary conditions and the effect of one treatment might carry over to the next treatment.</li> </ul>	<ul style="list-style-type: none"> <li>The order of the treatments were randomly counterbalanced</li> <li>Treatments were spaced in 1-week intervals.</li> </ul>
Research design/ data collection:			
Order bias	Occurs when the effect of the order of the interventions has an unintended impact on the results	<ul style="list-style-type: none"> <li>All participants took part in all conditions making it possible that the order that some students receive the condition might have had an impact on the outcomes.</li> <li>An interaction effect was noted in several of the statistical procedures</li> </ul>	<ul style="list-style-type: none"> <li>I was conservative in my interpretation and did not interpret pairs that were a part of any interaction.</li> </ul>

Stage of Design/ Limitation	Description	Manifestations in Current Study	Method of Increasing External Validity
Research design/ data collection / data analysis:			
Population validity	Determines the generalizability of the findings from the participants to the target population	<ul style="list-style-type: none"> <li>The district where the study was conducted has a disproportionate number of minority and ELL students; additionally, there are a disproportionate number of students who are classified as lower-socioeconomic.</li> </ul>	<ul style="list-style-type: none"> <li>Utilized random selection of school, random assignment of measure to teacher, random selection of treatment order, and large sample size.</li> <li>This was addressed within the narrative as a limitation to the generalizability of the findings.</li> </ul>
Research design/ data collection / data analysis:			
Specificity of variables	Occurs when the study is unique with respect to variables such as the participants, time, conditions, context, and variables, thereby making the findings less generalizable	<ul style="list-style-type: none"> <li>The demographic data of the sample are unique and likely do not represent the general population.</li> <li>The measures being utilized are specific to the study.</li> </ul>	<ul style="list-style-type: none"> <li>Operationally defined the variables in a way that had meaning outside of the study setting.</li> <li>This was addressed in the narrative as a limitation to the generalizability of the findings of this study.</li> </ul>

This table was adapted from Onwuegbuzie (2003). Reprinted with kind permission of the Mid-South Educational Research Association and the Editors of *Research in the Schools*.

## Appendix E

### Qualitative Dissertations or Qualitative Phase of Mixed Research Dissertations: Threats to Internal Credibility and External Credibility at the Interpretation Stage of the Research Process

**Qualitative phase.** Four threats to the internal and external credibility of this study were identified: (a) researcher bias, (b) observational bias, (c) reactivity, and (d) confirmation bias (Onwuegbuzie & Leech, 2007). The following procedures were followed in an effort to increase credibility: (a) triangulation, (b) leaving an audit trail, (c) checking for researcher effects/clarifying researcher bias, (d) debriefing of the interpretive researcher, and (e) determining effect sizes (Onwuegbuzie & Leech, 2007).

Although every effort was made to increase credibility for this study, observational bias, which can occur when the researcher fails to collect enough qualitative data, is offered as a limitation to this study. In utilizing the open-ended response questionnaire, I relied on fifth-grade students to be able to write about their perceptions about the four instructional conditions. Even though data from the questionnaire were informative, the questionnaire did not produce the rich data for which I had hoped. Therefore, when interpreting any results pertaining to the open-ended data, the reader should take this limitation into consideration. Table 9 presents the threats to internal and external credibility, a description of how each manifested itself in this study, and the manner in which I addressed each threat.

Table 9: Threats to Internal and External Credibility, Manifestations in Current Study, and Method of Increasing Credibility

Limitation	Description	Manifestations in Current Study	Method of Increasing Credibility
Researcher bias	Occurs when the researcher has preconceived ideas or biases that threaten the outcomes of the study	<ul style="list-style-type: none"> <li>I have a keen interest in and prior experience with the use of cartoon mnemonics and revised definitions that might have been a threat to credibility as I engaged in the qualitative portion of the study.</li> </ul>	<ul style="list-style-type: none"> <li>I engaged in written self-reflection prior to and throughout the interviewing process.</li> <li>I engaged in debriefing by one of my committee co-chairpersons.</li> <li>My self-reflection was included as part of my Chapter IV findings.</li> <li>I utilized member checking with the teachers in order to confirm my impressions.</li> </ul>
Observational bias	Occurs when there is a potential for the researcher to fail to collect enough observational data pertaining to a participant's words or behaviors	<ul style="list-style-type: none"> <li>One data analysis procedure relied on the open-ended responses of fifth-grade students who do not always articulate their thoughts well on paper.</li> <li>Another instrument required that I engage in extended interviews with participants. The thickness and richness of the data depended upon the participation level of the students and teachers.</li> </ul>	<ul style="list-style-type: none"> <li>I conducted three extensive interviews of each of the teacher participants over the course of the data collection period.</li> <li>I conducted one interview with each of 12 student participants</li> <li>The results of the students' open-ended responses are offered as a limitation to this study.</li> </ul>



Limitation	Description	Manifestations in Current Study	Method of Increasing Credibility
Reactivity	Occurs when the participants become aware that they are involved in a research study; might lead to the Hawthorne effect or the novelty effect	<ul style="list-style-type: none"> <li>• All students were aware that they were participating in a research study.</li> <li>• Only select students were involved in the interviews and this knowledge might have manifested itself into a Hawthorne effect.</li> </ul>	<ul style="list-style-type: none"> <li>• Although students were aware that there was a possibility that they could be interviewed, interviews did not take place until after all quantitative data collection had been completed; therefore, enough time had passed so that reactivity would have been at a minimum.</li> <li>• I engaged in triangulation by clarifying all responses with teachers and confirming my conclusions with all teacher participants. Additionally, I utilized quantitative data to triangulate any teacher or student responses.</li> </ul>

Limitation	Description	Manifestations in Current Study	Method of Increasing Credibility
Confirmation bias	The tendency for a researcher to interpret data in a manner that is favorable to his or her preconceived notions of a phenomena	<ul style="list-style-type: none"> <li>I have a keen interest in prior research conducted with the use of cartoon mnemonics and revised definitions and confirmation bias might have become an issue as I analyzed the data.</li> </ul>	<ul style="list-style-type: none"> <li>I engaged in written self-reflection (epoché) prior to and throughout the interviewing process.</li> <li>I engaged in debriefing with one of my committee co-chairpersons.</li> <li>My self-reflection was included as part of my Chapter IV findings.</li> <li>I utilized extreme cases for student interviews.</li> <li>I created an audit trail that included my audio and video recordings of all teacher, student and interpretive interviewer debriefs, my transcriptions, and my research journal.</li> </ul>

This table was adapted from Onwuegbuzie and Leech (2007). Reprinted with kind permission of Springer.

## Appendix F

## Mixed Research Dissertations: Threats to Mixed Research Legitimation at the Interpretation Stage of the Research Process

**Mixed research phase.** Three legitimation types were all determined to be potential threats to the proposed study: (a) multiple validities, (b) sequential, and (c) conversion (Onwuegbuzie & Johnson, 2006). Although all threats to this study that could be addressed were addressed, mortality, implementation bias, ecological validity, multiple-treatment inference, order bias, population validity, specificity of variables, and observational bias were further discussed earlier as limitations to the quantitative phase of the study; and researcher bias, observational bias, reactivity, and confirmation bias were further discussed earlier as limitations to the qualitative phase of the study; therefore, multiple validities legitimation remained a threat. Table 10 presents the threats to mixed methods legitimation, a description of how each manifested itself in this study, and the manner in which I addressed each threat.

Table 10: Threats to Mixed Methods Legitimation, Manifestations in Current Study, and Method of Increasing Legitimation

Limitation	Description	Manifestations in the Current Study	Method of Increasing Legitimation
Multiple validities legitimation	References the need to address all validities surrounding all methods in a proposed study	<ul style="list-style-type: none"> <li>Multiple threats to validity were evident within the current study.</li> </ul>	<ul style="list-style-type: none"> <li>All threats to validity that could be addressed were addressed.</li> <li>Population validity, specificity of variables, and ecological validity were discussed as limitations to this study.</li> </ul>
Sequential legitimation	References the need to reduce the impact that the order of the quantitative and qualitative phases might have on the ability to make meta-inferences	<ul style="list-style-type: none"> <li>All data were gathered sequentially; thus, the findings might have been an artifact of the sequence of phases (i.e., quantitative phase before the qualitative phase).</li> </ul>	<ul style="list-style-type: none"> <li>Teachers' interviews took place sequentially over the course of the data-gathering period.</li> <li>The interview data that I gathered from the students relied on it taking place after all quantitative data were gathered.</li> <li>The nature of the open-ended data also mandated that it take place after quantitative data were gathered.</li> </ul>

Limitation	Description	Manifestations in the Current Study	Method of Increasing Legitimation
Conversion legitimation	The ability to make quality meta-inferences from both quantitative and qualitative data in a study	<ul style="list-style-type: none"> <li>The quantizing of open-ended response data were utilized in the mixed methods analysis posed a threat; for example, counting the qualitative open-ended responses was not appropriate.</li> <li>The quantizing of the questioning task and sentence generation task also posed a threat</li> </ul>	<ul style="list-style-type: none"> <li>The relatively weak open-ended question responses of the students rendered the quantizing of the themes inappropriate, so I did not quantify students' responses.</li> <li>Both the questioning task and sentence generation task results were rigorously subjected to the following:               <ul style="list-style-type: none"> <li>100% of sentence generation task data were scored by two scorers.</li> <li>25% of the questioning task scores were randomly selected and scored by a second coder.</li> <li>Inter-scorer reliability was reported via a straight percentage in addition to Krippendorff's alpha.</li> </ul> </li> </ul>

This table was adapted from Onwuegbuzie and Johnson (2006). Reprinted with kind permission of the Mid-South Educational Research Association and the Editors of *Research in the Schools*.

### Copyright Disclaimer

Copyright reserved by the author(s).

This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).