ASSESSMENT OF THE CONCLUSION VALIDITY FOR EMPIRICAL RESEARCH STUDIES PUBLISHED IN THE *JOURNAL OF SPEECH, LANGUAGE, AND HEARING RESEARCH*

A Dissertation

by

GLENDA ELKINS BYRNS

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2007

Major Subject: Educational Psychology

ASSESSMENT OF THE CONCLUSION VALIDITY FOR EMPIRICAL RESEARCH STUDIES PUBLISHED IN THE *JOURNAL OF SPEECH, LANGUAGE, AND HEARING RESEARCH*

A Dissertation

by

GLENDA ELKINS BYRNS

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Co-Chairs of Committee,	James F. McNamara
	Salvador Hector Ochoa
Committee Members,	Linda H. Parrish
	Linda Skrla
Head of Department,	Michael R. Benz

May 2007

Major Subject: Educational Psychology

ABSTRACT

Assessment of the Conclusion Validity for Empirical Research Studies Published in the *Journal of Speech, Language, and Hearing Research*. (May 2007)

Glenda Elkins Byrns, B.S., Southwest Texas State University;

M.Ed., Southwest Texas State University

Co-Chairs of Advisory Committee: Dr. James F. McNamara
Dr. Salvador Hector Ochoa

Research-based decision making has been advanced as a way for professionals to make a determination about the effectiveness of a potential treatment. However, informed consumers of research need to be able to determine what constitutes evidence-based practices and what criteria can be used to determine if evidence-based practices have been met.

This study was a synthesis of research that involved a critical review of the empirical research studies reported in Volume 47 of the *Journal of Speech, Language, and Hearing Research* (*JSLHR*) published in 2004. This methodological research synthesis evaluated (a) the research designs used in the *JSLHR* studies, (b) information and rationale used to inform population validity assessment decisions, and (c) the extent to which the sampling designs, population validity rating, data analysis procedures, and the specification of generalizations and conclusions provide sufficient evidence to determine an overall rating of conclusion validity.

Results indicated that less than one-fifth of the 105 research synthesis population of studies used experimental research designs. Additionally, the vast majority of the research synthesis population of studies (83.8%) were observational research designs.

Only five studies out of the research synthesis population of studies (4.8%) were determined to have high population validity. In contrast, 84.8 percent of the research synthesis population of studies were found to have low population validity. That is, the studies did not contain adequate information or description of the essential sampling concerns.

The vast majority or 75.3 percent of the research synthesis population of studies were rated as having low conclusion validity. Approximately one-fifth of the 105 research synthesis study population (22 studies or 20.9%) were found to have moderate conclusion validity while less than five percent of the total studies (4 of 105 studies or 3.8%) were found to have high conclusion validity.

A meaningful relationship between population validity ratings and conclusion validity ratings was established. Since 81 of 105 studies have identical ratings for both population and conclusion validity, the accuracy of the prediction model developed for this study is 77.1 percent.

To my best friend, my husband

Bob Byrns

To the spirits of my being, my daughters

Megan Byrns Davis

Erin Byrns Cutts

ACKNOWLEDGMENTS

The acknowledgments below will, at best, serve only as a small token of my gratitude to my committee members and friends who have made this journey with me.

My greatest appreciation goes to the members of my doctoral committee. They are:

Dr. Jim McNamara, committee co-chair. Thank you for being a visionary, for taking the first vague idea and transforming that idea into a finished product. I appreciate your guidance and support. You define what being a teacher means.

Dr. Hector Ochoa, committee co-chair. You are a role model to others as you live your commitment. Your ability to see the best in people is truly appreciated.

Dr. Linda Parrish, committee member. The passion you have for children is an inspiration. Your zest for life is contagious and you have passed that feeling to many undergraduates in your career.

Dr. Linda Skrla, committee member. You have the ability to challenge beliefs while at the same time allowing others to see beyond. I strive to achieve your level of professionalism.

My heartfelt thanks goes to the doctoral students who have spent the last year reading articles and coding information. The completion of this dissertation would not have become reality without your help and support. Susan Skidmore, Toni Ivey, Dane Bozeman, and Judy Cryer—thank you.

While completing this process called a Ph.D., my family has increased by two. Adam Cutts and Andrew Davis—I am glad you have joined the family. You have enriched my life and added dimensions that I did not realize were lacking.

It has been said that completing a dissertation requires a friend who is taking the journey with you. Kendra Williams-Diehm, thank you for making the journey.

There are many others who were always there with a kind word of encouragement, I thank you. Specifically,

My parents, Glyn and Ellen Elkins. Thank you for being who you are.

Celia Goode-Haddock and Bill Haddock, for the Sunday brunches which provided a link to sanity and normalcy.

Kelly Schmid who always provided a touch of reality with a twist of humor.

All the members of the Red Letter Reflection Sunday School class.

Connie Fournier, Deborah Simmons, and Jorge Gonzales. Thank you for always offering support and words of encouragement.

In loving memory of Robert E. Byrns and Lily Byrns.

TABLE OF CONTENTS

CHAPTER Page

CHAPTER Page

CHAPTER Page

Page

LIST OF TABLES

TABLE Page

TABLE Page

TABLE Page

CHAPTER I

INTRODUCTION

Educators are being challenged to provide sound data on which to base decisions and recommendations for programming. The No Child Left Behind Act of 2001 references scientifically-based research as a way to improve policy. To achieve scientifically-based research, Feuer, Towne, and Shavelson (2002) argue "that the primary emphasis should be on nurturing and reinforcing a scientific culture of educational research" (p. 4). Scientific culture, they claim, provides "a set of norms and practices and an ethos of honesty, openness, and continuous reflection, including how research quality is judged" (p. 4).

When the frame of reference shifts from producers of research to consumers of research, the literature often reflects this transition by moving from the construct "scientifically-based research" to the construct "evidence-based practice".

Historically, evidence-based practice (EBP) is a term that originated in clinical medicine. EBP has recently been adopted in several other professions, including education. By definition EBP is "…the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of

---

This dissertation follows the style of the *American Educational Research Journal*.

individual patient … [by] integrating individual clinical expertise with the best available external clinical evidence from systematic research" (Sackett, Rosenberg, Gray, Haynes, & Richardson, 1996, p. 71). Evidence-based practice is, then, the integration of the science of research and the practical field application of the research.

In medicine and related professions, evidence-based practices integrate an individual's professional clinical expertise with the best available clinical evidence produced by systematic research. History provides numerous examples where treatments based solely on the recommendations of professionals rather than evidence-based practice or scientifically-based research have resulted in procedures, which were later determined to be detrimental to patient well-being. For example, in the 19th century, William Osler claimed that opium could limit the progress of diabetes (Sackett, Haynes, Guyatt, & Tugwell, 1991). In the 1940s the practice of oxygenating premature infants to prevent retrolental fibroplasia was later found to cause the condition (Meehl, 1997). The tongues of stutterers were surgically sliced in the early 1900s to treat stuttering (Van Riper, 1965). In the 1950s doctors performed frontal lobotomies to treat schizophrenia and obsessive-compulsive disorders (Pressman, 1998). In all these examples, treatments were based on clinical opinions of the time and lacked rigorous scientific testing.

Professional Domain of Interest

To prevent these sometimes less than beneficial and, at times, potentially harmful treatments, professional organizations are mirroring the national mandate and are urging their members to engage in evidence-based practice. The American Speech-Language-Hearing Association (ASHA) is an example of an organization that is advocating for members to engage in research-based decision making.

The American Speech-Language-Hearing Association (ASHA) is the professional, scientific, and credentialing association for more than 102,724 certified speech-language pathologists and 12,798 certified audiologists (ASHA, 2006a). This organization has issued both a Technical Report (2004) and a Position Statement (2005) on evidence-based practice in communication disorders. In their Technical Report, ASHA (2004) proposes that an evidence-based framework with explicit criteria be "used to evaluate the quality of evidence available to support clinical decisions" (p. 2). Additionally, the ASHA Position Statement (2005) calls for speech-language pathologists to "evaluate the quality of evidence appearing in any source or format, including journal articles ..." (p. 1). Justice and Fey (2004) advocate for heightened standards in the quantity and quality of research in the field of speech-language pathology to ensure that the research findings are "accessible to and assimilable by" (p.5) practitioners.

At the heart of this issue of evidence-based practice, professionals are called upon to make the determination whether the research findings described in professional journals yield valid findings. This is not an easy task as professional journals often contain articles that cover a variety of topics with an array of sampling procedures, research designs, and interpretations of findings. With so much discussion about and the focus on evidence-based practices, one might ask, “What constitutes evidence-based practices and how does one determine if the research contained in professional journals meets that criteria?” A synthesis of research can make a contribution answering this question.

Research Synthesis

Synthesis of prior research plays an important role identifying trends in research published in professional journals and by adding to the accumulation of knowledge within a profession. Light and Pillemer (1984) argue that a synthesis of research is actually an effort to “discover what is known” (p. ix) by learning from existing findings. An important facet of a synthesis of research involves organizing existing evidence in a systematic way. McMillan and Schumacher (2001) provide clarification that “the researcher constructs a picture that takes shape as he or she collects data and examines the parts” (p. 94). In research synthesis, there is an emergent research design that reformulates questions as the data collection evolves. The synthesis, or information-seeking with a purpose, can be one of two types: substantive or methodological.

A substantive review is a synthesis of findings and often takes the form of a meta-analysis. “Meta-analysis” was coined by Glass (1976) as the “analysis of analyses.” “In a meta-analysis the investigator gathers together all the studies relevant to an issue and then constructs at least one indicator of the relationship under investigation from each of the studies” (Cook, Cooper, Cordray, Hartmann, Hedges, Light, Louis, & Mosteller, 1992, pp. 4-5). McNamara (1997) explains that a meta-analysis can be used to identify trends in research. There are two basic types of meta-analysis: (a) comparison or mean difference studies which determine which method or group have higher scores on a criterion variable of interest, and (b) correlation studies that attempt to determine if a meaningful relationship exists between two variables. Substantive reviews typically ask questions that focus on identifying trends. Examples of questions from such an inquiry might be: “Is there a trend that emerges across studies?” or “Do these trends differ between different target populations?” Answers to substantive questions result in a better understanding of the nature or trends of the phenomenon under study.

The characteristics of a study that are not related to the substantive aspects involve methodological or procedural aspects. As such, a synthesis of research methods yields a synthesis of the procedures and methods used in studies. Feldman (1971) claimed that “systematically reviewing and integrating…the literature of a field may be considered a type of research in its own right--one using a characteristic set of research techniques and methods”

(p. 86). He further argued that the work of others should be viewed as the researcher's raw data. Synthesis of research methods can include variations in designs, research procedures, quality of measures, and forms of data analysis. Lipsey (1994) explains that the "…study results provides useful information about which aspects of research procedures make the most difference and, hence, should be most carefully selected by researchers" (p. 115). A synthesis of research methods would answer questions such as, "What research questions or problems are more likely to use single-case studies?" or "What data analysis procedures are more likely to be used with categorical data?" To take this concept a step further, a synthesis of research methods of articles contained within a specific journal for a specified period of time provides consumers a snapshot of what research methods and strategies are being used within the profession.

Research syntheses are beginning to emerge as dissertation topics in graduate schools in education. Specifically at Texas A&M University, Thompson (1993) did a substantive review on job satisfaction research published in the *Educational Administration Quarterly*. Wang (1996) undertook a synthesis of research methods of the survey sample designs published in the *Educational Administration Quarterly.* Kier (1999) conducted a synthesis of correlational research methods published in the *Journal of Consulting and Clinical Psychology.* Koehler (2000) completed a research methods synthesis that compared the 1997 narrative (vote counting method) literature review findings

on gender differences published in *Youth and Society* to her 1999 meta-analysis findings generated from the same population of studies used in the narrative review. Moore (2001) completed a synthesis of correlational research methods published in the *Journal of Counseling Psychology.*

Three characteristics of the set of the five research synthesis dissertations referenced above deserve mention here. First, empirical evidence used in each of these research syntheses came from a single research journal. Second, four of the five dissertations involved conducting a synthesis of research methods rather than a synthesis of substantive findings. Third, the written record in each of these five research synthesis dissertations is organized into a set of chapters that reflect the actual sequential phases put forth in the design of the inquiry.

Research Domain of Interest

This study is a synthesis of research methods that involves a critical review of research procedures and their impact on conclusion validity within the *Journal of Speech, Language, and Hearing Research* (*JSLHR).* The *JSLHR*, a professional journal in the fields of speech/language pathology and audiology, is published by the American Speech-Language-Hearing Association. Articles included in this journal focus on studies of the processes, diagnosis, and treatment of speech, language, and hearing disorders. The articles include experimental reports; theoretical, tutorial, or review papers; research notes; and

letters to the editor. A summary classification of the research topics addressed in all the articles included in any given volume is published two years later. For example, the classification of all articles published in 2004 is provided in a 2006 issue.

Although the profession of speech-language pathology has declared an interest in utilizing evidence-based practices, a simple preliminary review of the last four years of the *JSLHR* indicates that a brief summary classification that shows the rates of acceptance and rejection for manuscripts and submission-to-decision interval provided two years after each published volume, is the only synthesis that has been undertaken in the *JSLHR*. Neither a substantive review (to uncover meaningful trends) nor a synthesis of research methods have been reported in the *JSLHR*.

Intent of the Inquiry

In light of the absence of a research synthesis in the *JSLHR*, the purpose of this inquiry is to conduct a methodological research synthesis that determines both the population validity and conclusion validity of the empirical research studies reported in Volume 47 of the *Journal of Speech, Language, and Hearing Research* (*JSLHR*) published in 2004.

Thinking in terms of the two validity concerns referenced above, the specific intent of this methodological research synthesis is to conduct a critical review of (a) the accuracy and completeness of the sampling designs reported in

these *JSLHR* articles, and (b) the extent to which the actual characteristics of these sampling designs are reflected in the **discussions** of the rationale for selecting the data analysis procedures, the **application** of these procedures, and the **narrative report** of the information dealing with the findings, conclusions, implications for practice, and recommendations for future research.

Design of the Inquiry

The intent of this methodological research synthesis is accomplished by implementing a research design involving seven sequential phases. These seven phases **begin** with the creation of a theoretical framework to guide the synthesis (Phase One), **move** toward linking population and conclusion validity (Phases Two through Six), and **end** with the specification of a propositional inventory of recommendations for improving both future studies undertaken by individual researchers and the reports of their findings in the *JSLHR* (Phase Seven). Each of these seven sequential phases is identified in Table 1.1 and briefly described in the narrative that follows.

Phase One. Creating the Theoretical Framework

The purpose of phase one is to develop a theoretical framework to guide this inquiry. This theoretical framework has three parts that include (a) developing a category system with which to classify empirical studies according to their purpose and research design, (b) creating a set of questions that

Table 1.1

Design of the Inquiry

Phase 1:	Creating a theoretical framework
Phase 2:	Identifying the research methodological synthesis population
Phase 3:	Conducting an independent methodological research synthesis for studies that focus on hearing
Phase 4:	Conducting an independent methodological research synthesis for studies that focus on speech
Phase 5:	Conducting an independent methodological research synthesis for studies that focus on language
Phase 6:	Constructing an inventory of trend across the three research domains of hearing, speech, and language
Phase 7:	Specifying the recommendations for conducting future research studies and for presenting these findings in the *JSLHR*

provides an accurate description of sampling designs used in the empirical studies, and (c) framing a set of question that assesses the impact of sampling design on conclusion validity.

The first part of this theoretical framework involves building a classification system for types of research reported in the *JSLHR*. The classification system of articles is based on the reporting standards and guidelines recommended by the American Speech, Language, and Hearing Association (ASHA). This system is comprised of six mutually exclusive categories: (1) true experiments; (2) randomized research designs to evaluate interventions; (3) nonrandomized research designs to evaluate interventions; (4) instrument development with concurrent validity; (5) instrument development without concurrent validity; (6) single case studies; and (7) other nonexperimental quantitative research studies. The latter category is further classified to evaluate both the research objectives and time dimensions involved in the studies.

The second part of this theoretical framework is creating a set of questions to guide a methodological review of sampling designs used in the *JSLHR* articles. Central to the development of this sampling design classification is the need to separate the samples in individual studies into three broad categories. These are probability sampling designs and nonprobability sampling designs known as purposive sampling designs that involve either purposive sampling or quota sampling (Babbie, 1990).

The third part of the theoretical framework involves framing a set of questions that assesses the impact of sampling design on conclusion validity. The use of inferential statistics is related to the sampling design used in that study. Therefore the inferences drawn from the statistical analysis is dependent upon the sample design. The questions guiding this methodological research synthesis focuses on the impact of those sampling designs on the conclusion validity in each of the studies.

Phase Two. Identifying the Methodological Research Synthesis Population

The purpose of phase two is to describe the research synthesis population of both articles and studies in Volume 47 of the *JSLHR* along three dimensions: (a) substantive concerns addressed in each study; (b) research design used; and (c) country of origin where each study was conducted.

Phase Three. Conducting an Independent Methodological Research Synthesis for Studies that Focus on Hearing

Phase three is the first of three research syntheses. It looks at the 20 articles dealing with hearing research. More detailed substantive concerns in hearing research, sampling designs and validity assessments, are addressed. Findings are summarized as responses to a set of population and conclusion validity questions specified in the theoretical framework developed in phase one.

Phase Four. Conducting an Independent Methodological Research Synthesis for Studies that Focus on Speech

Phase four is the second of three research syntheses. It focuses on the 19 articles dealing with speech research. The response procedures for this phase parallel the ones used in phase three. More detailed substantive concerns in speech research, sampling designs and validity assessments, are addressed. Findings are summarized using the response procedures elaborated in phase three.

Phase Five. Conducting an Independent Methodological Research Synthesis for Studies that Focus on Language

Phase five is the third of the three research syntheses in this inquiry. It focuses on the 55 articles dealing with language research. The response procedures for this phase parallel the ones used in phase three. More detailed substantive concerns in language research, sampling designs and validity assessments, are addressed. Findings are summarized using the response procedures elaborated in phase three.

Phase Six. Constructing an Inventory of Trends Across the Three Research Domains of Hearing, Speech, and Language

In phase six, the purpose is to summarize the similarities and differences in the three previous independent research syntheses conducted in phases

three, four, and five. Analysis in this phase is well-defined as it uses the common set of questions developed in phase one and used to conduct the three sequential independent research syntheses.

Phase Seven. Specifying the Recommendations for Conducting Future Research Studies and for Presenting These Findings in the *Journal of Speech, Language, and Hearing Research*

Phase seven provides recommendations regarding editorial policies of the *JSLHR*, and for conducting future individual studies. The recommendations to editors target procedures for presenting findings that are in line with the policies of the *Journal of Speech, Language, and Hearing Research* and best practices in educational research. Recommendations for future research suggest ways of constructing research and then reporting findings in professional journals.

Definition of Terms

Two constructs are essential for understanding the intent of this inquiry. Both constructs are defined below.

Population validity is the degree to which the sample of participants included in the study is representative of the population from which it was selected (Gall, Gall & Borg, 1999). In order to establish population validity the selected sample must be shown to be similar to both the target population

(theoretical) and the accessible population (actual list). The more evidence the researcher provides in support of the relationship between the sample, the accessible population, and the target population the more confidence there is in the ability to generalize the findings of the study.

Conclusion validity relates to the results that are drawn in the study. McNamara (2003) explains that accurate results yield high conclusion validity. Reaching high conclusion validity requires that the researcher achieve compliance by (a) selecting the correct procedure for data analysis, (b) applying correctly all steps set forth by the procedure, and (c) offering only the conclusions that can be reached from the patterns and trends revealed in the data analysis.

Significance of the Inquiry

This study is an academic inquiry intended to contribute to the current knowledge base in the field of speech-language pathology. At least four major benefits can be expected of this inquiry.

First, this study is the initial inquiry to thoroughly and systematically examine the accuracy and completeness of the sampling designs reported in the *JSLHR* articles. Using both the phase one theoretical framework and the phase two sampling design typology findings from this study (a) reveals what are the most frequent types of sampling designs used in *JSLHR* articles, and (b)

identifies which types of sampling designs are most likely to have inaccurate and incomplete research design information.

Second, this study is also the initial inquiry to thoroughly and systematically examine the conclusion validity of articles published in the *JSLHR*. Accordingly, in the third through sixth phases, this study reports findings on (a) the extent to which the actual characteristics of these sampling designs are reflected in the discussions of the rationale for selecting the data analysis procedures, (b) the extent to which these data analysis procedures are correctly applied in individual studies, and (c) the extent to which the narrative report includes only the findings and conclusions that can be linked directly to the data analysis results.

Third, the final phase of this methodological research synthesis is devoted to sharing recommendations for two specific groups. These are individual researchers who plan to conduct future speech, language, and hearing research and for editors who are responsible for reviewing, selecting, and publishing prospective research articles.

Fourth, this inquiry presents an exploratory avenue that provides professionals with a new and innovative method for more systematic and rigorous evaluation of quantitative research. In addition, this evaluation strategy can easily be adapted for use in a wide range of behavioral science journals.

Organization of the Dissertation

The organization of this research synthesis follows the reporting strategy used in previous dissertations that have undertaken methodological research syntheses (Kier, 1999; Koehler, 2000; Thompson, 1993; Moore, 2001; Wang, 1996). Accordingly, this study is organized into nine chapters beginning with the purpose and research design elaboration in Chapter I. Chapter II provides the theoretical framework that guides the empirical aspects of the inquiry.

The next five chapters are designed to share the empirical results undertaken in this study. Chapter III describes the population included in the methodological research synthesis for this study. Chapters IV, V, and VI present the independent research synthesis findings for the three research domains of hearing, speech, and language. Chapter VII presents an inventory of trends across the research syntheses provided in the three previous chapters.

The last two chapters are used to explore implications and conclusions. Chapter VIII provides the implications as a set of recommendations for both researchers and editors. The final chapter, Chapter IX, summarizes the purpose, design, conclusions, and recommendations that emerge from this study.

References and appendices follow the chapters. The appendices provide information of interest and documentation of procedures and outcomes used in this methodological research synthesis. Appendix A and Appendix B provide the citations for all the articles and studies in Volume 47 of the *JSLHR*.

CHAPTER II

PHASE ONE

This chapter presents the findings for phase one which is dedicated to developing the theoretical framework that guides this inquiry. This theoretical framework has five components. These components are:

An elaboration of basic research concepts as they appear in research methods texts frequently used in the behavioral sciences,

An elaboration of basic sampling strategies as they are described in behavioral science reference sources dealing with research design, statistical methods, and data analysis,

A set of classification systems that can be used to categorize essential elements of empirical studies reported in Volume 47 of the *Journal of Speech, Language, and Hearing Research (JSLHR)*,

An inventory of questions that provides a basis to determine the population validity in these *JSLHR* studies, and

An inventory of questions that provides a basis to determine the conclusion validity in these *JSLHR* studies.

With this information at hand, the chapter is organized into six parts. Each of the first five parts shares information for one of the five components developed in the theoretical framework. The final part is used to summarize this information and to provide a few comments about how specific elements in the

theoretical framework influence tasks to be completed in the second through sixth phase of this research synthesis.

Basic Research Concepts

The initial component of the theoretical framework describes nine essential research concepts. The first two research concepts provide additional insights for population validity and conclusion validity, two terms that were briefly defined in the initial chapter dealing with the intent and design of this inquiry. The remaining seven research concepts are discussed in a single section dealing with basic sampling concepts used in inferential statistics.

Population Validity

Inferential statistics enables researchers to make inferences about a population based on descriptive statistics that are calculated on data from a sample. Population validity refers to the extent to which the sample of individuals in a study is representative of the population from which it was selected (Gall, Gall, & Borg, 2005).

In survey research, high population validity is achieved by satisfying two conditions. First, it must be demonstrated that the selected sample is similar to the accessible population (the actual list from which the researchers drew their sample). Second, and equally important, researchers must also demonstrate that the accessible population is similar to the target population which is defined

as the population to which the researchers intend to apply their research findings (Rosier, 1988).

Sampling theorists such as Schaeffer, Mendenhall, and Ott (1996) and Lohr (1999) are more likely to use the term “selection bias” as a means to establish population validity. They claim selection bias occurs when some part of the target population is not in the sampling frame (list of sampling units). Accordingly, high population validity for a sampling theorist implies freedom from selection bias. Moreover, the greater the selection bias, the lower the population validity.

In experimental studies, population validity is the extent to which the result of an experiment (treatment effect) can be generalized **from** the sample that participated in the experiment **to** a particular population of interest. Low population validity results when representativeness has not been ensured either by using an appropriate sampling procedure, such as random sampling, or by demonstrating equivalence through empirical comparisons (Tate, 1988).

For the purpose of this study it is essential to keep in mind that accurate and meaningful generalizations to a population of interest can only be made when either an experiment or a survey yields high population validity (Vogt, 2007).

Conclusion Validity

In general terms, conclusion validity should ensure that results reported in a research article are both accurate and meaningful. Tate (1988) provides a similar perspective. Thinking in terms of experimental studies, Tate suggests that the design of an experiment should ensure "adequate validity or truthfulness of conclusions" (p. 94).

Although experiments and observational studies share the common need for conclusion validity to reflect accurate and meaningful results, the approach used for determining conclusion validity in these two general types of studies is different. Both of these approaches are elaborated below.

Experiments. In experimental research texts, the term conclusion validity is seldom used. In these texts, conclusion validity is more likely to be discussed using the caption "threats to validity" (Mertens, 2005; Cooper & Hedges, 1994). Four specific validities that experimental researchers use to determine the overall conclusion validity of their study are statistical conclusion validity, internal validity, construct validity, and external validity (Tate, 1988). Each of these four validities and the corresponding question used to explore their threat to accurate and meaningful conclusions are elaborated in Table 2.1.

Observational Studies. The shift from experimental to nonexperimental studies involves a shift **from** using a cause-effect model **to** using one of three observational study models. The actual observational study choices are **(a)** a model that focuses on parameter estimation (survey research), **(b)** a model that

Table 2.1

Threats to Conclusion Validity in Experimental Research

Cook and Campbell (1979) identified four different validities that should be assessed to ensure accurate and meaningful conclusions of experimental research studies. Since that time almost all educational and behavioral science research handbooks have presented this perspective. Elaborated below is an example of how the original Cook and Campbell (1979) perspective is likely to be expressed in contemporary research handbooks using key research design questions.

Statistical Conclusion Validity	Does the empirical relationship between the operationalized variables exist in the population?
Internal Validity	If a relationship does exist, is it a causal relationship?
Construct Validity of Cause and Effect	Can the uncovered relationship between the operationalizations for treatment and response variables be generalized to the treatment and outcome constructs of interest?
External Validity	Can the sample relationship be generalized to individuals and situations beyond those involved in the study?

The elaborations shared above are based primarily on information provided in Tate (1988). This source also contains clear and detailed descriptions for assessing the threats to each of the four validities. Excellent validity assessment information is also available in Cooper & Hedges (1994) and more recently in Mertens (2005).

captures covariation (correlation research), or **(c)** a model that predicts a dependent variable of interest using one or more independent variables (prediction research).

When one conducts an observational study, only three of the four threats to validity elaborated in Table 2.1 must be assessed to determine the overall conclusion validity. The three validities to be assessed are statistical conclusion validity, internal validity, and external validity. An assessment of construct validity focusing on cause and effect (causal inferences) is not required (Cohen, Cohen, West, & Acken, 2003).

In general terms, these three validities address the following concerns in observational studies: **(a)** statistical conclusion validity in observational studies requires linking the actual research design to the correct statistical method for data analysis, **(b) i**nternal validity centers on producing valid and reliable measures for all variables of interest, and **(c)** external validity reflects the extent to which the study results can be generalized to individuals and situations beyond those included in the study (Gall, Gall, & Borg, 2005).

A Common Perspective. Given the conclusion validity information shared above, it is essential to recognize two important considerations. First, to verify that study results are accurate and meaningful, an assessment of conclusion validity must be undertaken for both experimental and observational studies. Second, it is also important to recognize that established procedures

have been advanced for assessing the conclusion validity in either type of study (i.e. experiments or observational studies).

With these two considerations in mind, three general assessment questions are used to organize and conduct individual conclusion validity assessments (McNamara, 2003). Referenced in the definition section of the first chapter, these three general assessment questions deal with (a) selecting the correct procedure for data analysis, (b) applying correctly all steps in this data analysis procedure, and (c) reporting only conclusions that can be justified from patterns and trends uncovered in the data analysis.

Conclusion Validity Assessments. A few insights into what is involved in constructing responses to these three general validity assessment questions are offered below.

For selecting the correct statistical data analysis procedures, the following guidelines are essential considerations. First, inferential statistical methods are the correct data analysis methods only when the research design involves the use of probability sampling (Sheskin, 2004).

Second, nonparametric statistical methods are the correct data analysis methods when probability sampling data do not meet parametric statistical model assumption or when probability sampling data are ordinal or nominal in nature (Mertens, 2005).

Finally, for nonprobability samples, descriptive statistical methods are the correct procedures for data analysis. This is the case because nonprobability

samples do not provide the information required to estimate variance due to sampling error (McNamara, 1994).

Once the synthesis research experts conducting the conclusion validity assessment are satisfied that the correct data analysis procedure was selected, they are then expected to review all statistical conclusion validity requirements associated with the actual procedure selected for data analysis. Two common cases illustrating how threats to statistical conclusion validity can emerge are given below.

When researchers in individual studies use test statistics to evaluate theoretical hypotheses of interest, they are expected to share information regarding (a) verification of the assumptions required for using the test statistic, (b) specification of the statistical power associated with the actual sampling design, and (c) specification of the effect size associated with the actual test result (Ottenbacher, 1989). Failure to share this required information clearly poses a threat to statistical conclusion validity.

When researchers in individual studies use statistical estimation procedures, they are expected to report (a) both the point and interval estimates for each parameter to be estimated, and (b) the corresponding confidence interval for each pair of parameter estimates. Failure to share this information clearly compromises statistical conclusion validity, and ultimately also compromises the overall conclusion validity for the published study results.

The research synthesis experts who will offer a response to the third and final general question posed for the conclusion validity assessment are expected to verify the extent to which the conclusions reported in a journal article (i.e., inferences, generalizations, emerging patterns, trends, etc.) are qualified or constrained to reflect the actual threats to internal and external validity (Rosier, 1988; Tate, 1988).

Two sources for these threats to validity deserve mention here. First, there are validity threats reported by the study authors. These threats are usually published in either the research methods section as study design limitations or in the discussion section as considerations likely to qualify or constrain the conclusions presented in the study. The second source of validity threats is the inventory of additional threats identified by the research synthesis experts.

More detailed information on the population validity and conclusion validity assessments to be conducted in this research synthesis are presented in the fourth and fifth components of this theoretical framework.

Sampling Concepts

In research methods texts frequently used in the behavioral sciences, basic sampling concepts are most likely to be described in two sequential steps (Mertens, 2005; Vogt, 2007). The first step is used to define basic sampling concepts as they relate directly to inferential statistics and probability sampling

designs. The second step is to relate this same set of concepts to their restricted use in descriptive statistics and nonprobability sampling designs. The typical transition from step one to step two in these descriptions includes a reference to the accepted superiority of probability sampling methods, often followed by some positive reference to the information value of the nonprobability sampling alternative in cases where probability sampling is not feasible or prohibitively expensive.

A similar two part elaboration of sampling concepts is also given in basic sampling theory texts (Lohr, 1999; Scheaffer, Mendenhall, & Ott, 1996) and in basic survey research methods texts (Alreck & Settle, 1995; Babbie, 1990).

Using information provided in the research methods, sampling theory, and survey research texts mentioned above, an overview of seven basic probability sampling concepts used in inferential statistics is provided in Table 2.2. Three additional comments regarding the use of these seven sampling concepts are offered below.

Nonprobability Sampling. How the basic sampling concepts in Table 2.2 are modified when research text authors shift from probability sampling (step one descriptions) to nonprobability sampling (step two descriptions) can be seen in the following elaboration.

When a nonprobability sampling alternative is selected for an observational study, the **sample population** and the **sample** (as these terms

Table 2.2

Sampling Concepts

Element	A unit on which measurements are taken (Scheaffer, Mendenhall & Ott, 1986). The information collected on the elements is the basis of analysis (Babbie, 1990).
Universe	A hypothetical term that conceptualizes a population at the broadest level. A universe is the aggregation of all the elements for a study (Babbie, 1990).
Population	A theoretical concept that refers to the aggregation of elements of a study (Rubin & Babbie, 2001). In a study, the population is the group that one attempts to describe and to which inferences are made (McMillan & Schumacher, 2001). The population must be defined prior to collecting a sample (Scheaffer, Mendenhall & Ott, 1996).
Sample population	The sample population is the actual group studied. This is the group from which the sample is selected (Rubin & Babbie, 2001).
Sample	A sample is a small proportion of a population that has been selected for observation and analysis (Vogt, 1999). If the sample is selected using probability sampling designs then inferences can be made to describe the population of interest (Scheaffer, Mendenhall & Ott, 1986).
Sampling error	The difference between the statistic for the sample and the population parameter is sampling error (Gall, Gall, & Borg, 1999). With probability sampling the sampling error can be estimated (Babbie, 1990).
External validity	The extent to which the findings of a study extend to participants and settings beyond those included in the study (Vogt, 1999, 2007).

are defined in Table 2.2) are **identical**. This is the case because the sample proportion is 1.0 (implying 100 percent of the population is observed).

Also noteworthy is the following consequence. When a nonprobability sampling alternative is used in an observational study, descriptive rather than inferential statistical methods are the correct procedures for data analysis. Moreover, the only **population** for which the descriptive statistics hold is for the group of **elements** in the **sample**. No inferential statistics can be derived from the descriptive statistics generated for the sample because, without probability sampling, no **sampling error** (also a concept entry in Table 2.2) can be estimated.

A nontechnical example should help to clarify the consequence put forth in the above elaboration. Assume a survey uses only volunteer participants. This sample of volunteers is also the only population under study. Statistics calculated for this sample of volunteers (also the population) cannot be used to create an inference (generalization) for any other group. In a word, the study ends with conclusions that apply only to the actual set of volunteers who completed the survey.

It is this generalizability restriction (i.e., the inability to apply the survey findings to anyone other than the volunteers who participated in the study) that prompts the authors of research methods texts to reference the accepted superiority of probability sampling.

Sampling Concepts in Experimental Studies. To better understand experimental methodology, Christensen (1988) suggests that it is helpful to think about two different types of experimental studies. **Idealistic experimental studies** are those that use random selection of participants and random assignment of participants to treatments. **Realistic experimental studies** are those that use nonrandom selection of participants but maintain random assignment of these participants to treatments. Christensen (1988) and Vogt (2007) both note that ideal experimental studies are seldom encountered in behavioral science research.

When idealistic experimental studies are implemented, the sampling concepts in Table 2.2 have direct application. Specifically, the distinction between populations and samples is maintained. Moreover, the random selection improves the external validity in that the experimental researcher can easily and confidently generalize to the population from which the samples of participants was randomly selected (Vogt, 2007).

When realistic experimental studies are implemented, the basic sampling concepts in Table 2.2 are less informative. This less informative situation is the case because in realistic experiments the attention shifts **from** the Table 2.2 emphasis on populations and corresponding samples as essential concepts **to** a new sampling concept called the **experimentally accessible population** (Tate, 1988). In realistic experiments, it is the experimentally accessible population that is randomly assigned to treatments.

Vogt (2007) provides an interesting way to think about sampling in realistic experiments. Specifically, Vogt suggests that experimenters using random assignment of nonrandom samples to treatments essentially **create** populations (the populations of treated and nontreated subjects) rather than **sample from** an actual population.

Other behavioral science researchers such as Tate (1988) like to think of experimentally accessible populations as **subpopulations**. This perspective allows researchers to easily and confidently generalize to this subpopulation, usually in terms of confidence intervals reflecting differences in treatment effects.

External Validity. The third and final comment relating to the use of sampling concepts provided in Table 2.2 addresses external validity (the final concept in this table). Our concern in framing this final comment is to reflect the difference in the literature between **external validity** and **transferability**. Mertens (2005) references two distinct paradigms as a means to understand the difference in these two generalizability concepts.

The Mertens (2005) perspective begins by assuming that both external validity and transferability deal with the ability of researchers and consumers of research to extend or generalize the findings of a particular study beyond the specific individuals and settings in which that study occurred.

With this point of clarification in hand, Mertens (2005) suggests that the postpositivist paradigm uses the term **external validity**. Within the postpositivist

paradigm, the external validity or generalizability of a study depends on the design and implementation of the sampling strategy. Specifically, the postpositivist researcher typically follows the probability sampling strategy where a **target population** is defined as the sample population (see Table 2.2) and also as the population to which the researcher wishes to generalize.

Drawing directly from the treatment of qualitative research methods given in Denzin and Lincoln (2000) and the specific definition of “thick description” offered in Lincoln and Guba (2000), Mertens (2005) notes that the constructivist paradigm uses the term **transferability** rather than **external validity** to talk about the ability to generalize the findings of a particular case or case study beyond the specific individuals and settings in which that study occurred.

In more specific terms, Mertens (2005) explains that in the constructivist paradigm the **researcher’s task** is to provide sufficient thick description about the case or case study under investigation so that **consumers of research** (readers of the research report) can understand the contextual variables operating in that setting.

Given both the thick description provided by the **researcher** and the corresponding understanding that can be acquired by **readers**, Mertens (2005) suggests that “the burden of generalizability then lies with the readers, who are assumed to be able to generalize subjectively form the case in question to their own personal experiences” (p. 309).

Sampling Strategies

The second component of the theoretical framework summarizes basic sampling strategies as they are described in behavioral science research methods handbooks. This summary is organized into three major sections, with information provided in each section having direct and immediate value for constructing the research design classification system to be presented as the third component of the theoretical framework.

Types of Quantitative Research Studies

Behavioral science research methods handbooks are organized to present two different types of quantitative research studies. These two types of studies are summarized in Table 2.3.

The first panel in this table acknowledges the idealistic versus realistic distinction in experimental studies put forth in Christensen (1988). The second panel in this table is used to elaborate three of the most common types of quantitative observational studies encountered in behavioral science research. For the record, it is of interest to note that a prestudy exploratory review of recent volumes of the *JSLHR* indicated that both experimental and observational studies are frequently encountered in individual issues of this journal.

Table 2.3

Types of Studies Used in Quantitative Research

Experimental studies make comparisons between groups. The random assignment of subjects creates statistical equivalence between the groups and allows for the use of inferential statistics (McMillan & Schumacher, 2001). Experimental studies attempt to predict and establish cause-and-effect relationships (Moore, 1983). Two types of experimental studies, A and B, are discussed below.

Experimental Study A	Idealistic Experiment (Christensen, 1998) Step 1: Subjects are randomly selected from the population. Step 2: Those randomly selected are randomly assigned to treatment and control groups (Vogt, 2007).
Experimental Study B	Realistic Experiment (Christensen, 1998) Step 1: Subjects from a well-defined population are identified through purposive rather than random sampling. Step 2: The subjects are randomly assigned to treatment and control groups (McMillan & Schumacher, 2001).

Observational studies incorporate descriptive designs. These studies are nonexperimental because they evaluate relationships between non-manipulated variables (Best & Kahn, 2003). Three basic types of observational studies are discussed here: survey studies, correlational studies, and causal-comparative studies.

Survey	Surveys attempt to gather and present information on a topic of interest collected from a sample. This information is reflective of a population (Babbie, 1990). Data are collected through questionnaires or interviews about attitudes, beliefs, and behaviors (McMillan & Schumacher, 2001). Care must be taken to ensure that the sampling frame is complete and that response rate is high (Gall, Gall, & Borg, 2005).
Correlational	Correlational studies evaluate relationships between variables (Gall, Gall, & Borg, 2005). A statistical correlation is reported that provides information about the degree of the relationship or correlation (McMillan & Schumacher, 2001).
Causal-comparative	This is also known as *comparative studies*. These studies attempt to establish relationships between past events or conditions already existing (Best & Kahn, 2003).

Sampling Designs for Observational Studies

When quantitative observational studies are undertaken, several options exist for designing sampling plans. A summary of these sampling design options is provided in Table 2.4. This table is divided into three panels. The first panel summarizes basic probability sampling designs. The second and third panels elaborate the most common types of nonprobability sampling designs used in behavioral science research. Additional comments and points of clarification for the sampling designs identified in these three panels are offered below.

Probability Sampling Designs. Strictly speaking, only probability sampling designs allow researchers to use inferential statistics (Krathwol, 1993). Although all probability sampling designs involve random sampling at some stage in the sampling process, the correct definition of probability sampling implies that every element in the population has a known nonzero probability of being selected (Kish, 1965; Rosier, 1988). Accordingly, different strata in a single stratified sampling design can have different probabilities of selection. In general terms, larger selection probabilities are used when strata have larger variability and small selection probabilities are used when strata are less variable or more homogeneous (Agresti & Finlay, 1999).

For this synthesis of research methods, two probability sampling features are important to recognize and address in the assessment of sampling designs. First, as sample sizes increase, sampling error is reduced and statistical power is increased (Rosier, 1988; Rubin & Babbie, 2001). Second, nonresponse error

Table 2.4

Sampling Designs Used in Observational Studies

Three basic types of sampling designs used in observational studies are summarized below.

I. Probability sampling designs allow researchers to make inferences about a population (Krathwohl, 1993). Probability samples involve random sampling from the population and are, therefore, measurable (Kish, 1965). In random selection, which is key to probability sampling, each unit has an equal chance of being selected independent of any other event (Rubin & Babbie, 2001). This allows for the use of inferential statistics which enables generalization from a sample to the population (Hinkle, Wiersma & Jurs, 2003). Four types of probability sampling designs highlighted are: simple random, systematic random, stratified random, and cluster random.

Simple random	Each unit in the population of interest has an equal and independent chance of being selected for the sample (McMillan & Schumacher, 2001). Inferential statistics can be used when analyzing the data and allows for generalizations beyond the sample.
Systematic random	Every k^{th} element in the population list is systematically chosen for inclusion in the sample. The first k^{th} element is randomly selected (Moore, 1983). Inferential statistics can be used and generalizations beyond the sample can be made (Agresti & Finlay, 1999).
Stratified random	The population of interest is divided into subgroups, or strata, on the variable of interest. This stratification provides homogenous subsets from which samples are randomly selected (Rubin & Babbie, 2001). Inferential statistics can be used and generalizations beyond the sample can be made (Agresti & Finlay, 1999).
Cluster random	The population of interest is divided into a large number of groups, or clusters (Agresti & Finlay, 1999) that are heterogeneously balanced, or have the same mixture of characteristics as any other group (Moore, 1983). The clusters are then randomly selected as the sample. Inferential statistics can be used when analyzing the data and allows for generalizations beyond the sample.

Table 2.4 (Continued)

II. Basic nonprobability sampling designs do not include random sampling which questions the representativeness of the sample (Krathwohl, 1993). If the topic of interest occurs in a small percentage of the population, the sample is intended to approximate the population (Hultsch, MacDonald, Hunter, Maitland, & Dixon (2002). However, because the sample is not representative of the population, the use of inferential statistics should be questioned (Agresti & Finlay, 1999). Two types of nonprobability sampling designs discussed are: convenience and purposive.	
Convenience	This is also known as *reliance on available subjects* and a *grab sample*. The use of this sample may be warranted if a researcher is attempting to study the characteristics of a specific group at a specific time under a specific condition (Rubin & Babbie, 2001). Inferential statistics should not be used as there is no way to generalize findings beyond the participants in the study (McMillan & Schumacher, 2001).
Purposive	This is also known as *judgmental* sampling. It requires the researchers to use their experience and knowledge of the topic of interest to select a sample that is representative of the population (Krathwohl, 1993). Conclusions reached from data of purposive or judgmental samples can only be used for the sample from which it came.
III. Other nonprobability sampling designs use other methods with which to gain a sample. These include quota and snowball.	
Quota	Quotas are established for characteristics of interest are distributed in the sample as they are in the population (Krathwohl, 1993). A quota frame, or the proportions that the different cells represent, is developed and data are collected from participants who had all the characteristics of a given cell (Rubin & Babbie, 2001). Generalizations hold only for the sample used.
Snowball	When members of a population are difficult to locate, snowball sampling may be used. Researchers identify a few members that have the desired quality and they are asked who they know that also possesses this quality. This process is continued until the names are repeated (Krathwohl, 1993). Data hold only for the sample used.

reduces the confidence placed on generalizations provided by inferential statistics (Babbie, 1990). Moreover, threats to the confidence placed on these generalizations can be reduced (but never completely eliminated) by using established procedures for investigating the influence of nonresponses (Holt, 1988).

Nonprobability Sampling Designs. While randomized probability samples are also considered the ideal for observational studies, Mertens (2005) suggests that probability samples are not commonly used in educational and psychological research. Similarly, Vogt (2007) notes that “much sampling in educational and medical research is not random or probability sampling” (p. 80).

For observational studies conducted in educational, psychological and medical research, both Mertens (2005) and Vogt (2007) acknowledge that the two most common nonprobability sampling strategies are convenience sampling and purposive sampling which is also called judgment sampling in many research methods handbooks. These two sampling strategies are described in the second panel of Table 2.4. Three general comments are offered here to extend the information provided in this panel.

First, Ross (1988) provides the following general contrasts as a means to clarify the actual limitations of convenience sampling. Sampling designs in experimental studies are strong in terms of internal validity. Sampling designs in survey research and in other forms of observational studies are strong with

respect to external validity. However, convenience sampling designs are weak on both internal and external validity.

Vogt (2007) provides these two additional insights for the limitation of convenience sampling. He initially notes that “inferential statistics do not make a lot of sense with convenience samples, although this almost never prevents people from computing these statistics” (p. 81). Vogt then goes on the claim that researchers who use convenience sampling hardly ever wants to say what they should say about convenience sampling. Specifically what Vogt suggests they should say is as follows: “I studied these folks for no very good reason except that it was easy; and there is no very good reason for you to be interested in my conclusions because the sample is wholly unrepresentative” (p. 81).

Second, Ross (1988) indicates that nonprobability sampling is educational and social science research “has mostly taken the form of judgment sampling in which expert choice is used to guide the selection of typical or representative samples” (p. 528). Vogt (2007) confirms that this trend of having judgment sampling being the most common form of nonprobability sampling is still true in contemporary social science research.

The third comment provides the rationale researchers frequently use to justify why they believe that inferential statistics can be used for nonprobability sampling data collected in purposive and judgmental sampling designs.

This rationale begins with a position advanced in both Walker and Burnhill (1988) and Ross (1988), and recently reaffirmed by Vogt (2007). This position

takes the form of an “If I then” statement. Specifically, if a researcher has extensive knowledge about a population, it is possible that a judgment or purposive sample may yield more precise information about the population than would a sample of the same size drawn at random.

The second and final part of this rationale for using inferential statistics for judgment or purposive sample data is called the default option. The Walker and Burnhill (1988) description of the default option implies that researchers **assume implicitly** that the judgment or purposive sample data are equivalent to a sample of the same size drawn from an infinite population by simple random sampling.

Put briefly, adopting this **implicit assumption** declares that the possibility that a judgment sample **may yield** more accurate information about a population **is replaced** by a deterministic belief statement that the judgment sample actually **does yield** more accurate information about the population of interest. While Walker and Burnhill (1988) suggest that this declaration could result in misleading conclusions, they also note that this approach is “adopted in most standard statistical packages and hence in much published work” (p. 105).

Other Nonprobability Sampling Designs. Most research methods handbooks also acknowledge less frequently used nonprobability sampling designs. Two of these infrequently used nonprobability sampling designs are elaborated in the third panel of Table 2.4.

Moving beyond basic information provided in this panel, it is important to recognize that snowball sampling is used in medical research and in qualitative social science research.

Snowball sampling (also called chain sampling) is often used in medical research to create a study population of those who have a rare medical condition. Snowball sampling is a very valuable sampling strategy to use when no population list is available or in cases where it is prohibitively expensive or too time consuming to assemble such a list.

Snowball sampling is also a nonprobability sampling strategy used in qualitative research to assemble a set of key informants who are considered to be knowledgeable about a program or community under study (Mertens, 2005).

Sampling Designs in Qualitative Research

Reflecting on the classification of sampling strategies, Mertens (2005) shares the following distinction. Researchers working in the quantitative research tradition (postpositivist approach) divides sampling strategies into **probabilistic** and **nonprobabilistic**. On the other hand, researchers working in the qualitative research tradition (constructivist approach) rarely use the term **nonprobabilistic**. They prefer to classify their sampling strategy options using the terms **theoretical** or **purposive**.

Selection Strategies. For qualitative inquiries, Gall, Gall, and Borg (2005) claim that the goal of purposive sampling is “to select individuals for case

study who are likely to be information-rich with respect to the researchers' purpose" (p. 310). Those actually selected in these purposive sampling plans are expected to have special knowledge or perspective that makes them especially important for obtaining emic perspectives.

In qualitative inquiries, the term **emic perspective** is seen to be the research participants' perceptions and understanding of their social reality. This term is contrasted with the term **epic perspective** which reflects the researchers' conceptual and theoretical understanding of the research participants' social reality (Gall, Gall, & Borg, 2005).

Table 2.5 describes fifteen purpose sampling strategies frequently used in qualitative inquiries. Following the descriptions given in Patton (2001), this table has four panels which groups purposive sampling options in terms of their underlying rationale. Two comments are offered below to provide additional insights regarding qualitative research sampling methods.

Probability Sampling. From Table 2.5 one can observe that a wide range of sampling methods can be used to collect data from individual research participants. Closer inspection of these methods indicates that not all sampling options in qualitative research are nonprobability sampling designs. Specifically, the last option in the first panel of Table 2.5 is a probability sampling design. Actual use of probability sampling in qualitative research is further discussed in Mertens (2005, Chapter 11).

Table 2.5

Sampling Designs Used in Qualitative Research

Researchers frequently use case studies to describe, explain, or evaluate specific social phenomena of interest. Fifteen sampling strategies that can be used in qualitative case study research are elaborated below. These strategies are organized into four groups, based on their underlying rationale.

I. To Capture an Essential Feature of Interest seven different strategies are explained.

Extreme or *deviant case*	Sampling unit that reflects a feature of interest at extreme opposites of the spectrum of interest
Intensity	Sampling unit that reflects a feature of interest at a high or low extent, but not to the extreme
Typical case	Sampling unit that reflects a feature of interest to an average
Maximum variation	Sampling units that reflects the range of the feature of interest
Stratified	Sampling units where the feature of interest occurs at an even distribution that are predefined
Homogeneous	Sampling units where the feature of interest are similar
Purposeful random	Sampling units that possess the feature of interest are selected at random from the population

Table 2.5 (Continued)

II. To Capture Conceptual Theory through research can be done with five strategies. These include critical case, theory-based, confirming/disconfirming, criterion, and cases that are politically important.	
Critical case	Sampling units where the feature of interest is tested
Theory-based or *operational construct*	Sampling units where an unproven topic of interest is involved
Confirming or *disconfirming case*	Sampling units where the feature of interest is supported or not supported
Criterion	Sampling units meet a specific standard
Politically important case	Sampling units that have political or social standing
III. Strategies that Evolve as the Research Develops allows the researcher the opportunity to take advantage of topics of interest as they emerge. These include opportunistic and snowball.	
Opportunistic	Sampling units where the feature of interest is highlighted during data collection. The researcher pursues this avenue of inquiry.
Snowball or *chain*	Sampling units that refer others who posses the feature of interest
IV. Strategy that does not have a Rationale consists of convenience samples. This strategy lacks a rationale that prevents it from being placed in one of the three previous categories.	
Convenience	Sampling units that may or may not possess the feature of interest and are used because they are available

Note. *This table is adapted from information provided in Chapter 10 in Gall, Gall, & Borg (2005).*

Research Accountability. Sowden and Keeves (1988) argue that in reporting qualitative research “it is essential for the researchers to accept the responsibility of being accountable, and to present clearly a statement on the analytical procedures employed” (p. 514). Included in this reporting effort is the need to elaborate the actual sampling strategy so that future readers (consumers of research) can make informed judgments about transferability.

On these written report requirements for sampling in qualitative inquiries, Mertens (2005) provides these two normative expectations. First, in the treatment of sampling considerations she notes “although the goal is not generalizability from a sample to a population, it is important that the researcher make clear the sampling strategy and its associated logic to the reader” (p. 317).

Later on in a focused discussion of research methodology Mertens (2005) again stresses the need for qualitative researchers to discuss the criteria for the selection of the participants and the setting of the study. Here she also notes that qualitative studies typically occur in natural settings, and equally important, all individuals in that setting are considered to be study participants. With this perspective in mind, Mertens declares that researchers should specifically “describe the method that will be used to identify those participants who will serve as a subsample to provide in-depth information” (p. 446).

Synthesis

In this second component of the theoretical framework, care has been taken to explain in detail that all researchers—those who operate in either the qualitative or quantitative tradition—are expected to accurately describe the sample of all participants who are directly involved in their studies.

Classification Systems

This part of Chapter II provides the findings that emerged from developing the third component of the theoretical framework. The results for this developmental effort yield two specific outcomes. The primary outcome is a research design classification system that can be used to accurately categorize the research designs encountered in this inquiry. The secondary outcome is the identification of the set of additional classification systems that are used in the research synthesis.

This part of Chapter II is organized into eight sections. Both the methods and results used to develop an accurate and meaningful research design classification system are presented in the first seven sections. The final section describes other classification systems to be used in this research synthesis.

Rationale

Initial reflections on the first two components of the theoretical framework (research concepts and sampling strategies), while helpful, are insufficient to

generate an accurate and meaningful research design classification system for the empirical research articles in Volume 47 of the *JSLHR.* The solution to this problem is the major reason for creating the third component of this theoretical framework. Work on this component capitalizes on additional research design recommendations provided by the editors and publisher of the *JSLHR* in the journal itself and on the American Speech-Language-Hearing Association (ASHA) website (ASHA, 2006b).

Editors and Publisher Recommendations

ASHA provides generic recommendations for publication in each of their five journals, which includes the *JSLHR.* In addition to the generic recommendations, ASHA provides more specific research design recommendations on their website (ASHA, 2006b). Each of these research design recommendations is linked to a type of study. These study types are experimental studies, intervention evaluation research studies, and diagnostic instrument development studies. These research design recommendations and their sources are elaborated below.

ASHA's research design recommendations for experimental studies are described in the Consolidated Standards of Reporting Trials (CONSORT) (Moher, Schulz, & Altman, 2001) located in Appendices C and D.

ASHA's research design recommendations for intervention evaluation research studies are discussed in the Transparent Reporting of Evaluations with

Nonrandomized Designs (TREND) (Des Jarlais, Lyles, Crepaz & TREND Group, 2004) found in Appendix E.

ASHA's research design recommendations for diagnostic instrument development studies are given in the Standards for Reporting of Diagnostic Accuracy (STARD) (Meyer, 2003) in Appendices F and G.

Stage One Classification System

The three sets of recommendations provided by the ASHA editors and publisher were used to create the stage one classification system summarized in Table 2.6.

This initial research design classification system consists of seven elements organized under three study types (experimental studies, intervention evaluation research studies, and diagnostic instrument development studies). Notice that the first five designs that are included in the second and third panels of Table 2.6 all involve the use of experimental methods, but for two different purposes.

Specifically, experimental studies (see panel two) are designed to extend the theoretical knowledge-base in a given field. On the other hand, intervention evaluation research studies (see panel three) uses experimental methods to evaluate one or more procedures for use in professional practices.

The two final designs (see panel four) are both used for research focusing on developing a new diagnostic instrument. In the first case (Diagnostic

Table 2.6

Stage One Research Design Classification System

The stage one design classification system consists of seven elements organized under the three broad concepts of experimental studies, intervention evaluation research studies, and diagnostic instrument development studies.

Experimental Studies. In this classification system experimental studies reflect the authors' intent to extend theoretical knowledge-base in a field of study. The difference between the two types of experimental studies listed below relates to how the sample for the experiment is selected.

Experimental Study I	Random selection of sample with random allocation to treatments
Experimental Study II	Nonrandom selection of sample with random allocation to treatments

Intervention Evaluation Research Studies. The studies that focus on intervention evaluation research are experiments that reflect the authors' intent to evaluate procedures used in professional practice. The first two types of intervention evaluation research studies differ according to how the sample for the experiment is selected.

Intervention Evaluation Research I	Random selection of sample with random allocation to treatments
Intervention Evaluation Research II	Nonrandom selection of sample with random allocation to treatments
Intervention Evaluation Research III	Single-case studies

Diagnostic Instrument Development Studies. The primary intent of studies concerned with diagnostic instrument development is the development of a new testing instrument to identify a condition or outcome of interest. The first type has the psychometric advantage of being able to use concurrent validity. In the second type of study, the concurrent validity advantage is not available.

Diagnostic Instrument Development I	With concurrent validity
Diagnostic Instrument Development II	Without concurrent validity

Instrument Development I) the instrument development can be influenced by concurrent validity. Concurrent validity is most likely achieved by using a traditional diagnostic instrument that shares the same purpose as the newly proposed instrument. The second case (Diagnostic Instrument Development II) also provides for developing new instruments; however, psychometric instrument comparisons cannot be made. This second case is most likely to occur when instrument development addresses a new diagnostic assessment concern.

Rationale Revisited

The seven categories provided in the stage one classification system (Table 2.6), while helpful, do not provide an adequate system to accurately and meaningfully classify the entire set of journal articles in the research synthesis population. In more specific terms, those articles not classified are nonexperimental quantitative studies. Thus, there still remains a need to extend the classification efforts again to clearly represent these nonexperimental quantitative studies.

To help in the development of a more adequate research design classification system, an additional literature search that focused on classifying nonexperimental quantitative studies was conducted. This search of the literature uncovered a useful classification system developed by Johnson (2001).

Nonexperimental Research Design

Johnson (2001) elaborates a theoretical classification system that classifies nonexperimental research studies across two dimensions. The first dimension is the **horizon of time** used in individual studies. The second dimension reflects the author's declaration of **research intent** or **research objective**. Combining these two dimensions produces a method for categorizing nonexperimental research designs. An overview of this research design classification system is given in Table 2.7.

The second panel in Table 2.7 highlights the horizon of time when data were collected. These time horizons, in the order elaborated in the table, represent cross-sectional research studies (present), longitudinal research studies (forecast future), and retrospective research studies (historical).

The third panel in Table 2.7 discusses the second dimension that provides information about the research objective using the three categories of descriptive research, predictive research, and explanatory research.

The Johnson (2001) framework for classifying nonexperimental research allows studies to be combined under the three descriptive horizons of time and the three descriptive dimensions of research objective. This framework yields nine descriptive categories that provide an accurate and meaningful classification system.

Table 2.7

Nonexperimental Research Design

Johnson (2001) provides a classification system for nonexperimental research designs that describes the intent of the study across two dimensions. The two dimensions describe the time horizon captured in the study and the authors' declared research objective. Both of these dimensions are elaborated below.

The **horizon of time** defines the period when the data are collected. Explained below are the three time horizons of cross-sectional research, longitudinal research, and retrospective research.

Cross-sectional research studies	are used to collect data at a single point in time or during a brief period of time (contemporaneous collection)
Longitudinal research studies	are used to collect data during more than one data collection period as a means to document changes in variable of interest.
Retrospective research studies	are used to assemble available archival data and to, at times, collect current data for establishing comparisons of interest.

The **primary research objective** explains what the researcher is trying to do in a study. The three types of research objectives are linked to a specific type of nonexperimental research. These three types are descriptive research, predictive research, and explanatory research.

Descriptive nonexperimental research studies	are primarily concerned with describing the population of interest or documenting characteristics of the population of interest.
Predictive nonexperimental research studies	are primarily concerned with predicting an event in a population of interest by using constructs in the population.
Explanatory nonexperimental research studies	are primarily concerned with explaining how or why a population of interest operates.

Stage Two Classification System

The information provided by Johnson (2001) was used to construct the nine part classification system elaborated in Table 2.8. This elaboration preserves all nine of Johnson's classification categories and, at the same time, fulfills the need to specify a classification system to categorize all of the quantitative nonexperimental research studies residing in the research synthesis population.

Final Research Design Classification System

The quest to develop an accurate and meaningful research design classification system for all studies published in Volume 47 of the *JSLHR* has adopted information found in several different sources. Using these sources produced the final classification system. This final research design classification system (Table 2.9) was achieved by combining the stage one classification system (Table 2.6) with the stage two classification system (Table 2.8). This final classification system has sixteen categories.

A complete history of all procedures and evidence needed to develop and validate the research design classification system is documented in Appendix H.

Table 2.8

Nonexperimental Quantitative Research Designs Obtained by

Crossing Horizons of Time and Research Objective

The nine descriptive categories of nonexperimental quantitative research designs (NQRD) are created by crossing the time horizon of data collection and the research objective. The nine categories are briefly described below. Johnson (2001) provides additional information on this research design.

NQRD I	Cross-sectional Descriptive	Data are collected at a single point in time or over a brief period of time and are primarily concerned with describing the population of interest or documenting characteristics of the population of interest.
NQRD II	Cross-sectional Predictive	Data are collected at a single point in time or over a brief period of time and are primarily concerned with predicting an event in a population of interest by using constructs in the population.
NQRD III	Cross-sectional Explanatory	Data are collected at a single point in time or over a brief period of time and are primarily concerned with explaining how or why a population of interest operates.
NQRD IV	Longitudinal Descriptive	Data are collected during more than one data collection period and are primarily concerned with describing the population of interest or documenting characteristics of the population of interest.
NQRD V	Longitudinal Predictive	Data are collected during more than one data collection period and are primarily concerned with predicting an event in a population of interest by using constructs in the population.
NQRD VI	Longitudinal Explanatory	Data are collected during more than one data collection period and are primarily concerned with explaining how or why a population of interest operates.
NQRD VII	Retrospective Descriptive	Available archival data and, at times, current data are used and are primarily concerned with describing the population of interest or documenting characteristics of the population of interest.
NQRD VIII	Retrospective Predictive	Available archival data and, at times, current data are used and are primarily concerned with predicting an event in a population of interest by using constructs in the population.
NQRD IX	Retrospective Explanatory	Available archival data and, at times, current data are used and are primarily concerned with explaining how or why a population of interest operates.

Table 2.9

Classification System Used in This Inquiry

Original Code	Description	Final Code
Experimental Study I	Extends theoretical knowledge-base in a field of study with random selection of sample with random allocation	1
Experimental Study II	Extends theoretical knowledge-base in a field of study with nonrandom selection of sample with random allocation	2
Intervention Evaluation Research I	Evaluates procedures used in professional practice with random selection of sample with random allocation	3
Intervention Evaluation Research II	Evaluates procedures used in professional practice with nonrandom selection of sample with random allocation	4
Intervention Evaluation Research III	Evaluates procedures used in professional practice using single-case studies	5
Diagnostic Instrument Development I	Diagnostic instrument development with concurrent validity	6
Diagnostic Instrument Development II	Diagnostic instrument development without concurrent validity	7
NQRD I: Cross-sectional Descriptive	Data are collected at a single point in time or over a brief period of time and are primarily concerned with describing the population of interest or documenting characteristics of the population of interest.	8
NQRD II: Cross-sectional Predictive	Data are collected at a single point in time or over a brief period of time and are primarily concerned with predicting an event in a population of interest by using constructs in the population.	9
NQRD III: Cross-sectional Explanatory	Data are collected at a single point in time or over a brief period of time and are primarily concerned with explaining how or why a population of interest operates.	10

Table 2.9 (Continued)

Original Code	Description	Final Code
NQRD IV: Longitudinal Descriptive	Data are collected during more than one data collection period and is primarily concerned with describing the population of interest or documenting characteristics of the population of interest.	11
NQRD V: Longitudinal Predictive	Data are collected during more than one data collection period and is primarily concerned with predicting an event in a population of interest by using constructs in the population.	12
NQRD VI: Longitudinal Explanatory	Data are collected during more than one data collection period and is primarily concerned with explaining how or why a population of interest operates.	13
NQRD VII: Retrospective Descriptive	Available archival data and, at times, current data are used and are primarily concerned with describing the population of interest or documenting characteristics of the population of interest.	14
NQRD VIII: Retrospective Predictive	Available archival data and, at times, current data are used and are primarily concerned with predicting an event in a population of interest by using constructs in the population.	15
NQRD IX: Retrospective Explanatory	Available archival data and, at times, current data are used and are primarily concerned with explaining how or why a population of interest operates.	16

Other Classification Systems

Five additional classification systems are used to organize and present the empirical evidence gathered in this research synthesis. The first two classification systems deal with population validity and conclusion validity.

The population validity classification system having three distinct categories is introduced below in the fourth component of the theoretical framework and also discussed in more detail in Appendix I.

The conclusion validity classification system having three distinct categories is introduced below in the fifth component of the theoretical framework and also discussed in more detail in Appendix J.

The other three classification system concerns and their corresponding appendix letters are substantive topics addressed in individual studies (Appendix K), country of origin for studies (Appendix L), and the format of individual articles (Appendix M). Each of these three appendices provides the information necessary to understand the developmental rationale, the construction of the classification system, and its actual use in the research synthesis.

Population Validity Assessment

The fourth component of the theoretical framework summarizes the findings that emerged from **developing** and **validating** the procedures used to conduct the population validity assessment.

This summary is organized into two sections. The first section centers on describing the steps taken to develop and implement the assessment strategy. The second section focuses on describing the steps taken to validate this proposed assessment strategy.

Assessment Strategy

The population validity assessment strategy is designed to yield one of three assessment outcomes for each of the 105 studies in the research synthesis population. These outcomes are **(a)** high population validity, **(b)** moderate population validity, and **(c)** low population validity.

Four specific steps are used to determine which of the three assessment outcomes best reflects the population validity of an individual study. A detailed description of this four step process is given in Appendix I. Given this documentation, just a brief overview of these four steps is presented here.

Step One. Since two or more researchers are needed to asses each study, the initial step in this process required the identification of several researchers to complete the population validity assessment. Those identified to assist in this effort were the same five researchers who participated in conducting the research design classification task.

Step Two. The second step in this process involved these five researchers directly in constructing an inventory of questions whose answers

taken collectively would provide a firm basis for making informed judgments regarding the allocation of these three population validity assessment outcomes.

Eight essential questions emerged from this effort. These questions are elaborated in Table 2.10 and revisited in Appendix I.

Step Three. The third step in this process involved extensive training sessions that prepared these researchers to construct accurate written responses for each question in the inventory.

Step Four. The final step in designing the assessment strategy involved constructing a series of focus group sessions. Each focus group session was designed to complete three specific tasks for a set of approximately ten individual studies.

The first task provided an opportunity for all researchers who rated the same study to share their responses to the eight essential population validity questions.

Once all researchers shared their individual written responses for a single study, the second task in the focus groups session was for researchers to indicate which of the three assessment outcomes they believed best fit the population validity of the study.

The final task of the focus group session was to reconcile (if needed) any differences in their individual assessment outcomes. Thus, following this process provided a consensus assessment outcome for each of the set of studies addressed in a focus group session.

Table 2.10

Essential Questions Used to Inform

Population Validity Assessment Decisions

What information and rationale are provided to:

- **Link** the sample to the purpose of the study?
- **Define** the target population?
- **Describe** the extent to which the sample represents the target population?
- **Define** the decision rules for sample size(s)?
- **Link** the sampling selection decisions to the proposed data analysis strategy?
- **Link** the sampling selection decisions to the potential generalizability of study findings?
- **Describe** the extent to which survey return rates or experimental attrition rates compromise study findings?
- **Ensure** that sampling procedures can be replicated in future studies?

Validity Concerns

Two tasks were required to establish the validity of the proposed assessment procedures elaborated above. First, there was a need to ensure that the inventory of essential questions was sufficient for researchers to yield an informed judgment regarding the three population validity assessment outcomes. Second, there was a need to ensure interrater reliability. Activities undertaken to meet these two validity requirements are detailed in Appendix I.

Conclusion Validity Assessment

The fifth and final component of the theoretical framework summarizes the findings that emerged from **developing** and **validating** the procedures used to conduct the conclusion validity assessment.

This summary is organized into four sections. The first section summarizes the intent of the assessment. The second section specifies the assessment outcomes. The third section elaborates the assessment procedures. The final section introduces the steps taken to establish the validity of this assessment effort.

Intent of the Assessment

Recognizing that three possible population validity assessment outcomes have already been specified to represent the accuracy and completeness of the sampling designs reported in each of the 105 studies in the synthesis

population, the conclusion validity assessment is used **to evaluate** the extent to which the actual characteristics of these individual sampling designs are reflected in the corresponding study information presented in the methods and results sections.

The evaluation effort used to provide meaningful information for making an overall conclusion validity assessment has four major evaluation components.

First Evaluation Component. In the methods section of each study, the initial component of the evaluation effort focuses on determining the extent to which the actual characteristics of the sampling design (both its positive features and its limitations) are reflected in the rationale given for selecting the data analysis procedures.

Central to this part of the overall evaluation effort is the need to ensure that inferential statistical methods are selected in the data analysis plan only when studies employ true probability sampling strategies or when purposive sampling strategies are correctly implemented to approximate a true probability sample.

Second Evaluation Component. Using information provided in both the methods and results sections, the second component of the evaluation effort is designed to assess the extent to which the study researchers follow the established procedures designated for each of their proposed data analysis procedures.

When these established procedures are followed, one can infer that the data analysis efforts have been correctly implemented.

Thinking in terms of statistical conclusion validity, four major concerns are central to implementing the second evaluation component.

First, study researchers are expected to provide information regarding their understanding and verification of the data assumptions required for a valid application of each proposed statistics procedure. This information is almost always given in the methods section.

Second, study researchers are expected to share information regarding statistical power considerations when statistical tests are proposed either to test hypothesis or to test statistical model assumptions. They are also expected to share information about proposed margins of error when estimation and prediction studies are undertaken. This information (if presented) is also almost always given in the methods section.

Third, study researchers are expected to identify the reference sources or the computer programs they used to generate their statistical results. This information is usually also given in the methods section. Equally important here is the fact that this information is an essential ingredient needed to replicate study results.

Fourth, turning toward the results section, the established procedures for statistical conclusion validity require researchers to share their actual effect

sizes when inferential test statistics are used and to share their actual margins of error when probability estimation models are used.

Third Evaluation Component. This component of the overall evaluation effort is devoted to determining the extent to which the actual characteristics of the sampling design (both its positive features and its limitations) are reflected in the study narrative dealing with **(a)** presentation of statistical findings, **(b)** specification of generalizations and conclusions, and **(c)** elaboration of implications and recommendations.

Clearly, these three evaluation considerations almost always span the entire narrative given in the results section. This claim can be easily verified by examining Appendix M which shares how individual *JSLHR* authors have chosen to organize their own study articles.

Two specific insights deserve mention when research synthesis evaluators concentrate on the third evaluation component.

First, this evaluation component reflects positions taken by the study researchers themselves rather than positions consumers of this research derive once they have read the published study narrative.

Second, the research synthesis evaluators are expected to evaluate the extent to which the results section narrative that study researchers put forth in their article is both accurate and justified in light of the actual research design and data analysis efforts. This is especially important when examining the

narrative detailing study researchers' position on generalizations and conclusions.

The Fourth Evaluation Component. The final component of the evaluation effort is dedicated to evaluating information provided for consumers of research. In more specific terms, this component is designed to evaluate the extent to which **(a)** the sampling design information presented in the methods section and **(b)** information on qualifications imposed on study findings which is shared in the results section *are sufficient for a research consumer to think in terms of generalizing to individuals and situations beyond those involved in the study.*

Three specific insights deserve mention when research synthesis evaluators concentrate on the fourth evaluation component.

First, drawing on the formal treatment of **external validity** and **transferability** elaborated earlier in this chapter (see the first component narrative on basic research concepts), transferability is the concept that can best guide the fourth component research effort.

Second, it is helpful to note that Bracht and Glass (1968) extended the concept of external validity to include population validity and ecological validity. In their theoretical framework population validity refers to generalizations made from the accessible population to the target population. Ecological validity addresses the issue of generalization in terms of the possibility of replicating empirical results using other participants residing in other settings. Accordingly,

ecological validity information (to be found in the study narrative) describing the sample population, the study setting, and the procedures implemented in the study is essential for research consumers when exploring the possibility of transferability.

Third, the first step in implementing a fourth component evaluation for a single study is to put forward the study position on the description of the target population. One of three possibilities will result from implementing this step.

If the target population is adequately described in the methods section of the study, this information can directly inform a research consumer's investigation of transferability.

If the target population is not described, but the information on study participants provided in the methods section allows readers to adequately imply a realistic target population, this derived information can help a research consumer's investigation of transferability.

If the target population is not described and the information on study participants provided in the methods section does not allow readers to adequately imply a realistic target population, the study yields little if any meaningful help to a research consumer's investigation of transferability.

Clearly, the third possibility elaborated above will almost always ensure that the fourth evaluation component will not contribute to a high conclusion validity rating.

A summary of the specific intents for each of the four evaluation components is given in Table 2.11.

Assessment Outcomes

Taken collectively, the judgments made for each of the four evaluation components should provide a research evaluator with sufficient evidence to declare an overall conclusion validity assessment outcome.

This exploratory research synthesis effort is designed to yield one of three assessment outcomes for each of the 105 studies in the research synthesis population. These outcomes are **(a)** high conclusion validity, **(b)** moderate conclusion validity, and **(c)** low conclusion validity.

Assessment Strategy

Following the procedures used to determine the population validity assessment outcomes, four specific steps are used to specify which one of the three predetermined assessment outcomes best reflects the conclusion validity of an individual study. A detailed description of this four step process is given in Appendix J. Given this documentation, just a brief overview of these four steps is presented here.

Table 2.11

Four Evaluation Components Used to Assess Conclusion Validity

Taken collectively, the judgments made for each of the following four evaluation components should provide sufficient evidence to determine an overall rating of conclusion validity within a study.

Determine the extent to which the actual characteristics of the sampling design (both its positive features and its limitations) are reflected in the rationale given for selecting the data analysis procedures.

Assess the extent to which the established procedures designated for each of the proposed data analysis procedures are followed.

Determine the extent to which the actual characteristics of the sampling design (both positive features and limitations) are reflected in the study narrative dealing with **(a)** presentation of statistical findings, **(b)** specification of generalizations and conclusions, and **(c)** elaboration of implications and recommendations.

Evaluate the extent to which the sampling design and information of qualifications imposed on study findings are shared.

Step One. Since two or more researchers are needed to assess each study, the initial step in this process required the identification of several researchers to complete the conclusion validity assessments. Those identified to assist in this assessment effort were the same five researchers who conducted the population validity assessment task.

Step Two. The second step in this process involved these five researchers directly in constructing the four-part evaluation that provides a firm basis for making informed judgments regarding the allocation of these three conclusion validity assessment outcomes.

Step Three. The third step in this process involved extensive training sessions that prepared these researchers to construct accurate written records for the evaluation of each study.

Step Four. The final step in designing the conclusion validity assessment strategy involved constructing a series of focus group sessions. Each focus group session was designed to complete three specific tasks for a set of approximately 20 individual studies.

The first task provided an opportunity for all researchers who rated the same study to share their responses to the four evaluation components. Once this was accomplished, the second task asked each researcher to share their conclusion validity assessment outcomes. The final task in a focus group session was to reconcile (if needed) any differences in their individual assessment outcomes. Following this procedure provided a consensus

assessment outcome for each of the set of studies addressed in a focus group session.

Validity Concerns

Two tasks were required to establish the validity of the proposed assessment procedures specified above. First, there was a need to ensure that the inventory of four evaluation components was sufficient for researchers to yield an informed judgment regarding the three conclusion validity assessment outcomes. Second, there was a need to ensure interrater reliability. Activities undertaken to meet these two validity requirements are detailed in Appendix J.

Summary

This chapter details the theoretical framework that guides the findings reported in the subsequent phases of this inquiry. Specifically, five components were developed. These components provided:

An elaboration of basic research concepts,

An elaboration of basic sampling strategies,

A set of classification systems that capture essential elements of the 105 empirical studies reported in Volume 47 of the *JSLHR*,

An inventory of questions that guides the assessment of population validity, and

An inventory of evaluation concerns that guides the assessment of conclusion validity.

CHAPTER III

PHASE TWO

This chapter presents the findings for phase two which is dedicated to identifying two research synthesis populations. Specifically, two essential research synthesis populations that emerge from Volume 47 of the *Journal of Speech, Language, and Hearing Research* (*JSLHR*) are described. The first population is the research synthesis population of empirical articles. The second population is the research synthesis population of studies.

To respect the two distinct populations this chapter is arranged in two parts. Part one uses four sections to share basic information about the research synthesis population of articles. Part two uses two sections to share basic information about the research synthesis population of studies. More in-depth information about the studies will be addressed in subsequent chapters that are devoted to elaborating the findings for the next three phases. Moreover, Appendix B provides further information about the transformation recode between the article synthesis population and study synthesis population.

Research Synthesis Population of Empirical Articles

Part one is organized into four sections, with each section providing information about the empirical articles included in this research synthesis. Section one offers a preliminary analysis of all articles. Section two provides an

evaluation of the substantive concerns addressed in each of the empirical articles. Section three shares an assessment of the research designs used in each of the empirical articles. Section four identifies the country of authorship for each article.

Preliminary Analysis

The first section of part one provides an overview of the findings of the preliminary analysis of the individual articles published in the *JSLHR* in 2004. Table 3.1 highlights these findings in three panels.

The **first panel** in Table 3.1 reveals that the majority of the articles of the 109 total article population are empirical studies. In contrast, the remaining 17 articles do not contain empirical studies.

The content of these 17 articles, which are not retained in the synthesis population of articles, is described in more depth in Table 3.2. Strictly speaking, the one meta-analysis in this table could be classified as an empirical article. This option was not used. Inspection of this table reveals that the most common article not retained in the research synthesis article population offers a research note, which commonly contains tutorials about current methods.

The **second panel** in Table 3.1 classifies the articles into articles with a single study or articles containing multiple studies. Interestingly, most articles (89.1%) report the findings of single studies. This constitutes a one-to-one

Table 3.1

Article Population and Study Population

	Frequency	Percent
Total Article Population = 109		
Empirical Study Articles	**92**	**84.4**
Nonempirical Articles	16	14.7
Meta-analysis Article	1	0.9
Empirical Article Population = 92		
Articles with single study	**82**	**89.1**
Articles with two studies	7	7.6
Articles with three studies	3	3.3
Study Population = 105		
82 articles with single study	**82**	**78.1**
7 articles with two studies	14	13.3
3 articles with three studies	9	8.6

Note. *Modal frequency in each panel is in bold print.*

Table 3.2

Articles Not Retained in the Research Synthesis Population of Articles by Type

N=17

Article Type	Frequency	Percent	Cumulative Percent
Research Note	**8**	**47.1**	47.1
Letters to Editor	3	17.6	64.7
Responses to Letters to Editor	3	17.6	82.3
Summary	1	5.9	88.2
Literature Review	1	5.9	94.1
Meta-analysis	1	5.9	100.0
Total	**17**	**100.0**	**--**

Note. *Modal frequency is in bold print.*

correspondence between study population and article population. The remaining ten articles share the results of two or three studies contained in a single article, thus creating the need for a more clearly defined study population which is addressed in the third panel.

The **third panel** in Table 3.1 provides the link between the research synthesis population of empirical articles (N=92) and the research synthesis population of studies (N=105).

Substantive Concerns

The second section of part one provides information about the substantive domain of interest for all of the 92 articles in the synthesis population of empirical articles. Table 3.3 classifies the 92 articles by one of three substantive concerns…hearing, speech, or language.

The information contained in Table 3.3 is based on the classification system assigned by the editor of the *JSLHR* and reported in each individual journal issue. Inspection of the table reveals that over half (57.6%) of the 92 empirical articles address the domain of language. The remaining 39 articles are almost evenly divided between the topics of hearing (21.7%) and speech (20.7%). A more in-depth analysis of the three major substantive domains of interest (hearing, speech, and language) is provided in phases three, four, and five, respectively.

Table 3.3

Substantive Domains in the Empirical Article Population

N=92

Substantive Concern	Frequency	Percent
Hearing	20	21.7
Speech	19	20.7
Language	**53**	**57.6**
Total	**92**	**100.0**

Note. *Modal frequency is in bold print.*

Research Designs

The results of the classification of the research synthesis population of 92 empirical articles are summarized in Table 3.4 and Table 3.5. This classification effort uses the 16 category classification system developed in the theoretical framework and defined in Table 2.9.

The findings for the 82 empirical article population having single studies are given in Table 3.4. Inspection of the findings reveals that only 13.4 percent of the studies fall in the first seven categories covered in this classification system. Almost one-third of articles with single studies (29.3%) use cross-sectional explanatory research design. Furthermore, over one-third of the articles (37.8%) use cross-sectional descriptive research design. Taken collectively, these two categories account for over half of the articles with single studies (67.1%).

Table 3.5 presents the findings for the research design classifications used in articles with multiple studies. Of the ten articles with multiple studies, only three articles use multiple research designs. The majority of the articles use the same research design in all studies included in the specific article.

Table 3.4

Research Design Classifications for Articles Having Single Studies

N=82

Code	Research Design	Frequency	Percent	Cumulative Percent
1	Experimental Study I	--	--	--
2	Experimental Study II	3	3.7	3.7
3	Intervention Evaluation Research I	--	--	3.7
4	Intervention Evaluation Research II	2	2.4	6.1
5	Intervention Evaluation Research III	4	4.9	11.0
6	Diagnostic Instrument Development I	--	--	11.0
7	Diagnostic Instrument Development II	2	2.4	13.4
8	Cross-sectional Descriptive	**31**	**37.8**	51.2
9	Cross-sectional Predictive	2	2.4	53.6
10	Cross-sectional Explanatory	**24**	**29.3**	82.9
11	Longitudinal Descriptive	8	9.9	92.8
12	Longitudinal Predictive	2	2.4	95.2
13	Longitudinal Explanatory	2	2.4	97.6
14	Retrospective Descriptive	1	1.2	98.8
15	Retrospective Predictive	--	--	98.8
16	Retrospective Explanatory	1	1.2	100.0
	Total	**82**	**100.0**	**--**

Note. *The two largest frequencies are in bold print.*

Table 3.5

Research Design Classifications for Articles Having Multiple Studies

N=10

Article	Study	Same Research Design	Different Research Design
5			
	1	Cross-sectional Descriptive	
	2	Cross-sectional Descriptive	
	3	Cross-sectional Descriptive	
77			
	1	Cross-sectional Explanatory	
	2	Cross-sectional Explanatory	
	3	Cross-sectional Explanatory	
32			
	1		Experimental Study II
	2		Experimental Study II
	3		Cross-sectional Explanatory
37			
	1	Diagnostic Instrument Development II	
	2	Diagnostic Instrument Development II	
20			
	1	Cross-sectional Descriptive	
	2	Cross-sectional Descriptive	
62			
	1	Cross-sectional Descriptive	
	2	Cross-sectional Descriptive	
78			
	1	Cross-sectional Explanatory	
	2	Cross-sectional Explanatory	
104			
	1	Cross-sectional Explanatory	
	2	Cross-sectional Explanatory	
69			
	1		Intervention Evaluation Research I
	2		Longitudinal Descriptive
59			
	1		Cross-sectional Descriptive
	2		Intervention Evaluation Research III

Country of Origin

The fourth and final section of part one reports on the country of origin for the 92 empirical articles. This classification system is described in Appendix L. A snapshot of the countries of origin is provided in Table 3.6.

The first column in Table 3.6 shows that articles came from ten different countries. The United States was the largest contributor providing over two-thirds of the articles. Of the 28 articles originating outside the United States, the United Kingdom contributed almost half of those articles (12 of 28 articles). The remaining 16 articles were from eight other countries.

Table 3.7 provides information as to how the country of origin was determined. It is important to note that in over half (36 of 64 articles or 56.3%) of the articles established as being from the United States, the determination was made from a source other than information and description contained in the methods section of the article.

Thinking in terms of this research synthesis, it is of interest to note here that the **country of origin** should be clearly identified in the article, primarily because this identification is an essential element in the accurate definition of the study's target population.

Table 3.6

Country of Origin of Articles

N=92

Country	Frequency	Percent	Cumulative Percent
United States	**64**	**69.6**	69.6
United Kingdom	12	13.0	82.6
Israel	4	4.3	86.9
Australia	3	3.3	90.2
Hong Kong, SAR	3	3.3	93.5
The Netherlands	2	2.1	95.6
Finland	1	1.1	96.7
Belgium	1	1.1	97.8
South Africa	1	1.1	98.9
France	1	1.1	100.0
Total	**92**	**100.0**	--

Note. *Modal frequency is in bold print.*

Table 3.7

Determination of Country of Origin

Country	Number of articles	Determination made by			
		Information in article		Author information	
		(*f*)	(%)	(*f*)	(%)
United States	64	28	43.7	36	56.3
United Kingdom	12	9	75.0	3	25.0
Israel	4	2	50.0	2	50.0
Australia	3	3	100.0	--	--
Hong Kong, SAR	3	1	33.3	2	66.7
The Netherlands	2	2	100.0	--	--
Finland	1	--	--	1	100.0
Belgium	1	1	100.0	--	--
South Africa	1	--	--	1	100.0
France	1	--	--	1	100.0
Total	**92**	**46**	**--**	**46**	**--**

Notes. (*f*) reflects the frequency count. (%) reflects the percent.

Research Synthesis Population of Studies

Part two of this phase is organized into two sections, with each section sharing basic information about the research synthesis population of studies. Section one offers a description of the substantive domains of interest addressed in each study in this population. Section two shares a description of the research designs used in each study.

Preliminary Analysis

Table 3.8 identifies the broad substantive domain of interest addressed in each of the 105 studies.

Inspection of Table 3.8 reveals that over half of the studies (59 of 105 studies or 56.2%) address some concern under the topic of language. The rest of the studies are equally divided under the topics of hearing (23 of 105 studies or 21.9%) and speech (23 of 105 studies or 21.9%).

Research Designs

The results of the classification of the research synthesis population of 105 studies are summarized in Table 3.9. This classification uses the 16 category classification system developed in the theoretical framework and defined in Table 2.9.

Table 3.8

Substantive Domains in the Study Population

N=105

Substantive Concern	Frequency	Percent
Hearing	23	21.9
Speech	23	21.9
Language	**59**	**56.2**
Total	**105**	**100.0**

Note. *Modal frequency is in bold print.*

Table 3.9

Research Design Classifications for Studies

N=105

Code	Research Design	Frequency	Percent	Cumulative Percent
1	Experimental Study I	--	--	--
2	Experimental Study II	5	4.8	4.8
3	Intervention Evaluation Research I	1	0.9	5.7
4	Intervention Evaluation Research II	2	1.9	7.6
5	Intervention Evaluation Research III	5	4.8	12.4
6	Diagnostic Instrument Development I	--	--	12.4
7	Diagnostic Instrument Development II	4	3.8	16.2
8	Cross-sectional Descriptive	**39**	**37.1**	53.3
9	Cross-sectional Predictive	2	1.9	55.2
10	Cross-sectional Explanatory	**32**	**30.5**	85.7
11	Longitudinal Descriptive	9	8.7	94.4
12	Longitudinal Predictive	2	1.9	96.3
13	Longitudinal Explanatory	2	1.9	98.2
14	Retrospective Descriptive	1	0.9	99.1
15	Retrospective Predictive	--	--	99.1
16	Retrospective Explanatory	1	0.9	100.0
	Total	**105**	**100.0**	**--**

Note. *The two largest frequencies are in bold print.*

Inspection of the top panel in Table 3.9 reveals that only 16.2 percent of the studies fall in the first seven categories covered in this classification system.

Also noteworthy is the fact that the majority of the studies in this panel (13 of 17 studies) are experimental type studies. Only four studies are dedicated to diagnostic instrument development.

Inspection of the second panel of Table 3.9 reveals that the large majority of studies published in Volume 47 of the *JSLHR* (88 of 105 studies or 83.8%) are observational (nonexperimental) studies.

It is of interest to note that a large majority of the observational studies (73 of 105 total studies or 69.5%) use cross-sectional research designs. Specifically, the two most common research design types for these 88 observational studies are cross-sectional descriptive (39 of the 105 total studies or 37.1%) and cross-sectional explanatory (32 of the 105 total studies or 30.5%).

Summary

This chapter presented the findings for phase two of this methodological research synthesis. Two distinct research synthesis populations contained in Volume 47 of the *Journal of Speech, Language, and Hearing Research* (*JSLHR*) emerged. Specifically, these populations were the research synthesis population of empirical articles and the research synthesis population of studies.

For the 92 articles identified in the research synthesis population of empirical articles four major trends emerged.

The large majority of the 92 articles in the empirical article population contain single studies (82 of 92 articles or 89.1%). Just ten articles (10.9%) contained multiple studies.

Over half of the article population (53 of 92 articles or 57.6%) addresses the substantive domain of language.

The two most common research designs used in articles having single studies were cross-sectional descriptive (31 of 82 articles with single studies or 37.8%) and cross-sectional explanatory (24 of 82 articles with single studies or 29.3%). Specifically, the time frame is that data are collected at a single point or during a brief period of time with the intent to either describe or explain a phenomenon of interest.

Ten different countries contributed articles to the *JSLHR* in 2004. Unsurprisingly, authors from the United States contribute the majority of articles (69.6%).

For the 105 studies identified in the research synthesis population of studies two major trends emerge, both of which mirror the trends in the empirical article population.

Over half of the studies (59 of the 105 studies or 56.2%) address a concern under the substantive topic of language.

The two most common research designs used in studies are cross-sectional descriptive (39 of the 105 studies or 37.1%) and cross-sectional explanatory (32 of the 105 studies or 30.5%).

This research synthesis population of 105 studies will be explored in more depth in the next three phases devoted to the substantive concerns of hearing, speech, and language.

CHAPTER IV

PHASE THREE

This chapter presents the findings for phase three which is dedicated to a synthesis of the 23 studies that focus on hearing research. The first part elaborates the distribution of specific substantive hearing research concerns encountered in these 23 studies. The second part identifies the research designs used in each study. The third part discusses the 23 population validity ratings. The fourth part discusses the 23 conclusion validity ratings.

Substantive Concerns

An overview of the classification of the 23 hearing studies by specific substantive concern is provided in Table 4.1. Examination of this table reveals that 11 specific concerns were addressed in these 23 studies.

Over one-fourth of these studies (six of 23 or 26.1%) addressed the specific substantive concern of speech perception. Another four of these 23 studies (17.4%) addressed the specific substantive concern dealing with cochlear implants. Thus, just two of these 11 substantive hearing research concerns account for almost half (43.5%) of the specific substantive concerns addressed in the synthesis population of hearing research studies.

Table 4.1

Substantive Concerns Addressed in Hearing Research Studies

N=23

Substantive Hearing Concern	Study Number	Frequency	Percent	Cumulative Percent
Cochlear implants	1, 29, 30, 69	4	17.4	17.4
Masking-level	2	1	4.3	21.7
Speech recognition	86, 87, 31	3	13.1	34.8
Effects of aging	16, 54	2	8.8	43.6
Binaural asymmetry	17	1	4.3	47.9
Spectrographic displays	28	1	4.3	52.2
Speech acquisition	43	1	4.3	56.5
Speech perception	44, 57, 58, 99, 100, 101	**6**	**26.1**	82.6
Phonology	45	1	4.3	86.9
Prelingual deafness	55, 56	2	8.8	95.7
Auditory processing	59	1	4.3	100.0
Total		**23**	**100.0**	**--**

Notes. *Modal frequency is in bold print.*
Study numbers are described in Appendix B.

Research Designs

Information on the research designs used in each of the 23 studies focusing on hearing research concerns is given in Table 4.2 and Table 4.3. These two tables use the 16 category research design classification system developed in the theoretical framework and defined in Table 2.9.

From the information contained in Table 4.2, three key points deserve mention.

First, inspection of Table 4.2 indicates that better than nine out of ten hearing research studies (21 of 23 studies or 91.4%) are observational studies rather than experimental studies.

Second, detailed in the lower panel of Table 4.2 is that 18 of the 21 observational studies (85.7%) used a cross-sectional research design. In contrast, only two of 21 observational studies (9.5%) used a longitudinal research design.

Third, an inspection of these findings reveals that the two most frequently used research designs in the studies concerned with hearing are cross-sectional descriptive (39.1%) and cross-sectional explanatory (34.7%). These two types of research designs are used in almost three-fourths (73.8%) of the studies that target issues related to hearing.

Table 4.3 provides a bivariate distribution that classifies each of the 23 hearing research studies in terms of both research design and substantive concern. Three interesting points are worth highlighting.

Table 4.2

Research Designs Used in Hearing Research Studies

N=23

Research Design	Frequency	Percent	Cumulative Percent
Experimental Studies			
Experimental Study II	2	8.6	8.6
Observational Studies			
Cross-sectional Descriptive	**9**	**39.1**	47.7
Cross-sectional Predictive	1	4.4	52.1
Cross-sectional Explanatory	**8**	**34.7**	86.8
Longitudinal Predictive	1	4.4	91.2
Longitudinal Explanatory	1	4.4	95.6
Retrospective Descriptive	1	4.4	100.0
Total	**23**	**100.0**	**--**

Note. *The two largest frequencies are in bold print.*

Table 4.3

Hearing Research Studies Classified by Research Design and Substantive Concern

Research Design	Cochlear implants	Masking-level	Speech recognition	Effects of aging	Binaural asymmetry	Spectrographic displays	Speech acquisition	Speech perception	Phonology	Prelingual deafness	Auditory processing	**Total**
Experimental Studies												
Experimental Study II	--	--	--	--	--	--	--	--	--	2	--	**2**
Observational Studies												
Cross-sectional Descriptive	2	--	2	--	1	1	1	2	--	--	--	**9**
Cross-sectional Predictive	--	--	--	--	--	--	--	1	--	--	--	**1**
Cross-sectional Explanatory	1	1	--	2	--	--	--	3	--	--	1	**8**
Longitudinal Predictive	--	--	--	--	--	--	--	--	1	--	--	**1**
Longitudinal Explanatory	1	--	--	--	--	--	--	--	--	--	--	**1**
Retrospective Descriptive	--	--	1	--	--	--	--	--	--	--	--	**1**
Total	**4**	**1**	**3**	**2**	**1**	**1**	**1**	**6**	**1**	**2**	**1**	**23**

First, the only one of the eleven specific substantive topics to use an experimental research design dealt with prelingual deafness.

Second, longitudinal research designs are used for only two specific substantive concerns, with a single study addressing cochlear implants and a single study addressing speech perception.

Third, the 18 studies using cross-sectional research designs are distributed over ten of the eleven substantive hearing research concerns.

Population Validity

Information on the population validity ratings assigned to each of the 23 hearing research studies is given in Tables 4.4 to 4.6. The procedures used in making informed decisions about these 23 population validity assessment ratings are explained in Appendix I.

Table 4.4 provides an initial glimpse of the population validity ratings of the 23 hearing research studies. Two noteworthy points emerge.

First, of the 23 hearing research studies, no studies received a high population validity rating.

Second, the vast majority of the hearing research studies (21 of 23 studies or 91.4%) received a low population validity rating.

Table 4.5 provides a bivariate distribution that classifies each of the 23 hearing research studies in terms of both research design and population validity rating. Three interesting points are worth highlighting.

Table 4.4

Univariate Distribution of Population Validity Ratings of the Hearing Research Studies

N=23

Population Validity Rating	Frequency	Percent	Cumulative Percent
High Population Validity	--	--	--
Moderate Population Validity	2	8.6	8.6
Low Population Validity	**21**	**91.4**	100.0
Total	**23**	**100.0**	--

Note. *Modal frequency is in bold print.*

Table 4.5

Hearing Research Studies Classified by Research Design and Population Validity Rating

N=23

Research Design	Population Validity					
	High		Moderate		Low	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
Experimental Studies						
Experimental Study II	NA	--	--	--	2	8.7
Observational Studies						
Cross-sectional Descriptive	NA	--	1	4.3	8	34.8
Cross-sectional Predictive	NA	--	--	--	1	4.3
Cross-sectional Explanatory	NA	--	--	--	8	34.8
Longitudinal Predictive	NA	--	1	4.3	--	--
Longitudinal Explanatory	NA	--	--	--	1	4.3
Retrospective Descriptive	NA	--	--	--	1	4.3
Total	**--**	**--**	**2**	**8.6**	**21**	**91.4**

Note. *NA implies the cell of interest has zero count.*

Table 4.6

Hearing Research Studies Classified by Substantive Concern and Population Validity Rating

N=23

Substantive Hearing Concern	Population Validity					
	High		Moderate		Low	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
Cochlear implants	NA	--	--	--	4	17.5
Masking-level	NA	--	--	--	1	4.3
Speech recognition	NA	--	--	--	3	13.1
Effects of aging	NA	--	--	--	2	8.7
Binaural asymmetry	NA	--	--	--	1	4.3
Spectographic displays	NA	--	1	4.3	--	--
Speech acquisition	NA	--	--	--	1	4.3
Speech perception	NA	--	--	--	6	26.2
Phonology	NA	--	1	4.3	--	--
Prelingual deafness	NA	--	--	--	2	8.7
Auditory processing	NA	--	--	--	1	4.3
Total	**--**	**--**	**2**	**8.6**	**21**	**91.4**

Note. *NA implies the cell of interest has zero count.*

An examination of the crosstab data in Table 4.5 reveals that across the 23 hearing research studies, none were rated as having high population validity. Two studies (8.6%) were judged to have moderate population validity. The rest of the studies (21 of 23 or 91.4%) were considered to have low population validity rating.

Given these data, there is no relationship between the three levels of population validity and the 16 levels of research design used in the study.

Inspection of the crosstab data in Table 4.6 provides insights into a possible bivariate relationship between the specific substantive topics of hearing research and population validity ratings. No studies addressing the eleven substantive concerns of hearing were identified as having high population validity. Only two studies, addressing the substantive topics of spectrographic displays and phonology, were rated as having moderate population validity. Most importantly, the vast majority (21 of 23 studies or 91.4%) of the studies addressing the substantive concerns of hearing were found to have low population validity.

Given these findings, there is no relationship between the three levels of population validity and the eleven specific substantive topics addressed in hearing research.

Conclusion Validity

Information on the conclusion validity ratings assigned to each of the 23 hearing research studies is given in Tables 4.7 to 4.10. The procedures used in making informed decisions about these 23 conclusion validity assessment ratings are explained in Appendix J.

Table 4.7 provides an initial glimpse of the conclusion validity ratings of the hearing research studies. Two noteworthy points emerge.

First, of the 23 hearing research studies, no studies received a high conclusion validity rating. Second, the vast majority of the hearing research studies (21 of 23 studies or 91.4%) received a low conclusion validity rating.

Table 4.8 provides a bivariate distribution that classifies each of the 23 hearing studies in both research design and conclusion validity rating.

An examination of Table 4.8 reveals that across the seven research designs used in the 23 hearing research studies, none were rated as having high conclusion validity. Single studies in two of the research designs (cross-sectional descriptive and longitudinal predictive) were judged to have moderate conclusion validity. The majority of the studies (21 of 23 or 91.4%) were considered to have low conclusion validity rating.

Given these findings, there is no relationship between the three levels of conclusion validity and the 16 levels of research design used in the study.

Table 4.7

Univariate Distribution of Conclusion Validity Ratings of the Hearing Research Studies

N=23

Conclusion Validity Rating	Frequency	Percent	Cumulative Percent
High Conclusion Validity	--	--	--
Moderate Conclusion Validity	2	8.6	8.6
Low Conclusion Validity	**21**	**91.4**	100.0
Total	**23**	**100.0**	--

Note. *Modal frequency is in bold print.*

Table 4.8

Hearing Research Studies Classified by Research Design and Conclusion Validity Rating

N=23

Research Design	Conclusion Validity					
	High		Moderate		Low	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
Experimental Studies						
Experimental Study II	NA	--	--	--	2	8.7
Observational Studies						
Cross-sectional Descriptive	NA	--	1	4.3	8	34.9
Cross-sectional Predictive	NA	--	--	--	1	4.3
Cross-sectional Explanatory	NA	--	--	--	8	34.9
Longitudinal Predictive	NA	--	1	4.3	--	--
Longitudinal Explanatory	NA	--	--	--	1	4.3
Retrospective Descriptive	NA	--	--	--	1	4.3
Total	**--**	**--**	**2**	**8.6**	**21**	**91.4**

Note. *NA implies the cell of interest has zero count.*

Table 4.9

Hearing Research Studies Classified by Substantive Concern and Conclusion Validity Rating

N=23

Substantive Hearing Concern	Conclusion Validity					
	High		Moderate		Low	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
Cochlear implants	NA	--	1	4.3	3	13.1
Masking-level	NA	--	--	--	1	4.3
Speech recognition	NA	--	--	--	3	13.1
Effects of aging	NA	--	--	--	2	8.7
Binaural asymmetry	NA	--	--	--	1	4.3
Spectographic displays	NA	--	--	--	1	4.3
Speech acquisition	NA	--	--	--	1	4.3
Speech perception	NA	--	--	--	6	26.3
Phonology	NA	--	1	4.3	--	--
Prelingual deafness	NA	--	--	--	2	8.7
Auditory processing	NA	--	--	--	1	4.3
Total	**--**	**--**	**2**	**8.6**	**21**	**91.4**

Note. *NA implies the cell of interest has zero count.*

Table 4.10

Hearing Research Studies Classified by Population Validity Rating and Conclusion Validity Rating

N=23

Population Validity Rating	Conclusion Validity Rating					
	High		Moderate		Low	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
High Population Validity	NA	--	NA	--	NA	--
Moderate Population Validity	NA	--	1	4.3	1	4.3
Low Population Validity	NA	--	1	4.3	20	87.1
Total	**--**	**--**	**2**	**8.6**	**21**	**91.4**

Note. *NA implies the cell of interest has zero count.*

Table 4.9 provides a bivariate distribution that classifies each of the 23 hearing research studies in both the substantive concerns addressed in the hearing research study and the corresponding conclusion validity rating.

Inspection of this table reveals that none of the studies addressing the substantive concerns of hearing research were judged to have high conclusion validity. The vast majority (21 of 23 studies or 91.4%) were rated as having low conclusion validity.

Given these findings, there is no relationship between the eleven substantive concerns addressed in the hearing research study and the three levels of conclusion validity.

Table 4.10 provides a bivariate distribution that classifies each of the 23 hearing research studies in terms of both the population validity rating and the corresponding conclusion validity rating.

As could be expected given the preceding findings, no studies were found to have both high population validity and high conclusion validity. The 21 studies (21 of 23 studies or 91.4%) that fall on the principal diagonal of the crosstab were given the same rating for both population validity and conclusion validity.

For 20 of the 21 diagonal entries in Table 4.10, low ratings were given for both population validity and conclusion validity. This means that in these 20 studies **(a)** sampling concerns (i.e., accurate description of the actual study sample and a clear specification of the corresponding target population) were not discussed in the methods section and **(b)** the extent to which actual

sampling design characteristics should qualify or constrain study findings was not discussed in the results section.

For the other diagonal entry in Table 4.10, moderate ratings were given for both population validity and conclusion validity. Accordingly, this study received a moderate population validity rating because some useful information on sampling concerns was provided in the methods section narrative. Additionally, this study received a moderate conclusion validity rating because discussion of findings in the results section narrative was tempered by sharing some limitations of the sampling design.

The one case above the principal diagonal of the Table 4.10 crosstab was rated as having moderate population validity but low conclusion validity. This indicates that there was some description of sampling concerns in the methods section, but findings were not clarified in the discussion of results.

The one case below the principal diagonal of the Table 4.10 crosstab was rated as having low population validity but moderate conclusion validity. This indicates that there was no description of sampling concerns in the methods section, but the results section provided a discussion of the actual sampling characteristics in order to clarify the results.

The bivariate distribution given in Table 4.10 reveals a strong relationship between population validity and conclusion validity ratings. Thinking in terms of a prediction model, population validity ratings provide an accurate prediction for 21 of the 23 studies in the hearing research synthesis population.

In more specific terms, this prediction model declares that population validity ratings are identical to conclusion validity ratings. For example, 20 of the 21 studies having a low population validity rating also have a low conclusion validity rating and one of two studies having a moderate population validity rating also has a moderate conclusion validity rating.

Since 21 of 23 studies have identical ratings for both population and conclusion validity, the accuracy of this prediction model is 91.3 percent.

CHAPTER V

PHASE FOUR

This chapter presents the findings for phase four which is dedicated to a synthesis of the 23 studies that focus on speech research. The first part elaborates the distribution of specific substantive speech research concerns encountered in these 23 studies. The second part identifies the research designs used in each study. The third part discusses the 23 population validity ratings. The fourth part discusses the 23 conclusion validity ratings.

Substantive Concerns

An overview of the classification of the 23 speech studies by specific substantive concern is provided in Table 5.1. Examination of this table reveals that nine specific concerns were addressed in these 23 studies.

Almost one-fourth of these studies (5 of 23 studies or 21.7%) addressed the specific substantive concern of voice disorders. Speech perception and fluency were primary topics in four studies each (17.4% each). The remaining studies (10 of 23 studies or 43.5%) addressed six additional substantive topics.

Table 5.1

Substantive Concerns Addressed in Speech Research Studies

N=23

Substantive Speech Concern	Study Number	Frequency	Percent	Cumulative Percent
Articulation	5, 7, 71	3	13.0	13.0
Dysphagia	4, 34	2	8.8	21.8
Respiration	6	1	4.3	26.1
Speech perception	83, 84, 85, 20	**4**	**17.4**	43.5
Voice disorders	3, 18, 32, 93, 94	**5**	**21.7**	65.2
Fluency	19, 21, 33, 70	**4**	**17.4**	82.6
Dysarthria	46	1	4.3	86.9
Word frequency	102, 103	2	8.8	95.7
Anatomy	60	1	4.3	100.0
Total		**23**	**100.0**	**--**

Notes. *Modal frequencies are in bold print.*
Study numbers are described in Appendix B.

Research Designs

Information on the research designs used in each of the 23 studies focusing on speech research concerns is given in Table 5.2 and Table 5.3. These two tables use the 16 category research design classification system developed in the theoretical framework and defined in Table 2.9.

From the information contained in Table 5.2, three key points deserve mention.

First, inspection of Table 5.2 indicates that better than nine out of ten speech research studies (22 of 23 studies or 95.6%) are observational studies rather than experimental studies.

Second, detailed in the lower panel of Table 5.2 is that 19 of the 22 observational studies (86.3%) used a cross-sectional research design. In contrast, only two of the 22 observational studies (9.1%) used a longitudinal research design.

Third, over half (13 of 23 studies or 56.5%) of the studies concerned with speech research are cross-sectional descriptive.

Table 5.3 provides a bivariate distribution that classifies each of the 23 speech research studies in terms of both research design and substantive concern. Three interesting points are worth noting.

First, the only one of the nine specific substantive topics to use an experimental research design dealt with voice disorders.

Table 5.2

Research Designs Used in Speech Research Studies

N=23

Research Design	Frequency	Percent	Cumulative Percent
Experimental Studies			
Intervention Evaluation Research III	1	4.4	4.4
Observational Studies			
Cross-sectional Descriptive	**13**	**56.4**	60.8
Cross-sectional Explanatory	6	26.0	86.8
Longitudinal Descriptive	1	4.4	91.2
Longitudinal Explanatory	1	4.4	95.6
Retrospective Explanatory	1	4.4	100.0
Total	**23**	**100.0**	**--**

Note. *Modal frequency is in bold print.*

Table 5.3

Speech Research Studies Classified by Research Design and Substantive Concern

N=23

Research Design	Articulation	Dysphagia	Respiration	Speech perception	Voice disorders	Fluency	Dysarthria	Word frequency	Anatomy	**Total**
Experimental Studies										
Intervention Evaluation Research III	--	--	--	--	1	--	--	--	--	**1**
Observational Studies										
Cross-sectional Descriptive	2	1	--	3	3	2	1	--	1	**13**
Cross-sectional Explanatory	1	1	--	--	1	1	--	2	--	**6**
Longitudinal Descriptive	--	--	1	--	--	--	--	--	--	**1**
Longitudinal Explanatory	--	--	--	1	--	--	--	--	--	**1**
Retrospective Explanatory	--	--	--	--	--	1	--	--	--	**1**
Total	**3**	**2**	**1**	**4**	**5**	**4**	**1**	**2**	**1**	**23**

Second, the 19 studies using a cross-sectional research design are distributed over eight of the nine substantive speech research concerns.

Third, respiration was the only substantive topic of speech that did not use a cross-sectional research design.

Population Validity

Information on the population validity ratings assigned to each of the 23 speech research studies is given in Tables 5.4 to 5.6. The procedures used in making informed decisions about these 23 population validity assessment ratings are explained in Appendix I.

Table 5.4 provides an initial glimpse of the population validity ratings of the 23 speech research studies. Two noteworthy points emerge.

First, none of the 23 speech research studies received a high population validity rating.

Second, the overwhelming majority of the speech research studies (21 of 23 studies or 91.4%) received a low population validity rating.

Table 5.5 provides a bivariate distribution that classifies each of the 23 speech research studies in terms of both research design and population validity rating. Five interesting points are worth highlighting.

An examination of the crosstab data contained in Table 5.5 reveals that across the 23 speech research studies, none were rated as having high population validity. Inspection of the top panel in Table 5.5 reveals that the only

Table 5.4

Univariate Distribution of Population Validity Ratings of the Speech Research Studies

N=23

Population Validity Rating	Frequency	Percent	Cumulative Percent
High Population Validity	NA	--	--
Moderate Population Validity	2	8.6	8.6
Low Population Validity	21	91.4	100.0
Total	**23**	**100.0**	**--**

Note. *NA implies the cell of interest has zero count.*

Table 5.5

Speech Research Studies Classified by Research Design and Population Validity Rating

N=23

Research Design	Population Validity					
	High		Moderate		Low	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
Experimental Studies						
Intervention Evaluation Research III	NA	--	--	--	1	4.3
Observational Studies						
Cross-sectional Descriptive	NA	--	2	8.6	11	47.8
Cross-sectional Explanatory	NA	--	--	--	6	26.1
Longitudinal Descriptive	NA	--	--	--	1	4.3
Longitudinal Explanatory	NA	--	--	--	1	4.3
Retrospective Explanatory	NA	--	--	--	1	4.3
Total	**--**	**--**	**2**	**8.6**	**21**	**91.4**

Note. *NA implies the cell of interest has zero count.*

Table 5.6

Speech Research Studies Classified by Substantive Concern and Population Validity Rating

N=23

Substantive Speech Concern	Population Validity					
	High		Moderate		Low	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
Articulation	NA	--	--	--	3	13.0
Dysphagia	NA	--	--	--	2	8.7
Respiration	NA	--	--	--	1	4.3
Speech perception	NA	--	--	--	4	17.4
Voice disorders	NA	--	2	8.6	3	13.0
Fluency	NA	--	--	--	4	17.4
Dysarthria	NA	--	--	--	1	4.3
Word frequency	NA	--	--	--	2	8.7
Anatomy	NA	--	--	--	1	4.3
Total	**--**	**--**	**2**	**8.6**	**21**	**91.4**

Note. *NA implies the cell of interest has zero count.*

study having an experimental research design was determined to be of low population validity. The last two items of interest come from the bottom panel in Table 5.5 which looks at 22 observational studies.

First, the large majority (19 of 22 observational studies or 86.4%) used cross-sectional research designs.

Second, two of these 19 cross-sectional research design studies (10.5%) had moderate population validity ratings, while the remaining 17 studies (89.5%) all had low population validity ratings.

Given these data, there is no relationship between the three levels of population validity and the 16 levels of research design used in the study.

Inspection of the crosstab data in Table 5.6 provides insights into a possible bivariate relationship between the specific substantive topics of speech research and population validity ratings. No studies addressing the nine substantive concerns of speech were identified as having high population validity. Only the two studies that both addressed the substantive topic of voice disorders were considered to have moderate population validity. The remaining 21 studies distributed over the nine substantive concerns of speech were all found to have low population validity.

Given these findings, there is no relationship between the three levels of population validity and the nine specific substantive topics addressed in speech research.

Conclusion Validity

Information on the conclusion validity ratings assigned to each of the 23 speech research studies is given in Table 5.7 to 5.10. The procedures used in making the informed decisions about the conclusion validity assessments are explained in Appendix J.

Table 5.7 provides a distribution of the conclusion validity ratings for the 23 speech research studies. Two points are worth mentioning.

First, none of the 23 studies received a high conclusion validity rating.

Second, over half of the studies (15 of 23 studies or 65.2%) were rated as having low conclusion validity.

Table 5.8 provides a bivariate distribution that classifies each of the 23 speech studies in terms of both research design and conclusion validity rating.

Inspection of the top panel reveals that only one experimental research design was used in all of the 23 research synthesis population of speech studies. This single experimental study was judged to have a moderate conclusion validity rating.

Information contained in the second panel of Table 5.8 pertains to the 22 observational studies in this research synthesis population. Here it is important to note three items of interest.

First, the majority of the observational studies addressing the speech research concerns (15 of 22 studies or 68.2%) received a low conclusion validity rating. These 15 low conclusion validity ratings are distributed across five

Table 5.7

Univariate Distribution of Conclusion Validity Ratings of the Speech Research Studies

N=23

Conclusion Validity Rating	Frequency	Percent	Cumulative Percent
High Conclusion Validity	NA	--	--
Moderate Conclusion Validity	8	34.8	34.8
Low Conclusion Validity	15	65.2	100.0
Total	**23**	**100.0**	**--**

Note. *NA implies the cell of interest has zero count*

Table 5.8

Speech Research Studies Classified by Research Design and Conclusion Validity Rating

N=23

Research Design	Conclusion Validity					
	High		Moderate		Low	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
Experimental Studies						
Intervention Evaluation Research III	NA	--	1	4.3	--	--
Observational Studies						
Cross-sectional Descriptive	NA	--	7	30.5	5	21.8
Cross-sectional Explanatory	NA	--	--	--	7	30.5
Longitudinal Descriptive	NA	--	--	--	1	4.3
Longitudinal Explanatory	NA	--	--	--	1	4.3
Retrospective Explanatory	NA	--	--	--	1	4.3
Total	**--**	**--**	**8**	**34.8**	**15**	**65.2**

Note. *NA implies the cell of interest has zero count.*

Table 5.9

Speech Research Studies Classified by Substantive Concern and Conclusion Validity Rating

N=23

Substantive Speech Concern	Conclusion Validity					
	High		Moderate		Low	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
Articulation	NA	--	--	--	3	13.1
Dysphagia	NA	--	--	--	2	8.7
Respiration	NA	--	--	--	1	4.3
Speech perception	NA	--	3	13.1	1	4.3
Voice disorders	NA	--	4	17.4	1	4.3
Fluency	NA	--	--	--	4	17.4
Dysarthria	NA	--	1	4.3	--	--
Word frequency	NA	--	--	--	2	8.7
Anatomy	NA	--	--	--	1	4.3
Total	**--**	**--**	**8**	**34.8**	**15**	**65.2**

Note. *NA implies the cell of interest has zero count.*

Table 5.10

Speech Research Studies Classified by Population Validity Rating and Conclusion Validity Rating

N=23

Population Validity Rating	Conclusion Validity Rating					
	High		Moderate		Low	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
High Population Validity	NA	--	NA	--	NA	--
Moderate Population Validity	--	--	2	8.7	--	--
Low Population Validity	--	--	6	26.1	15	65.2
Total	**--**	**--**	**8**	**34.8**	**15**	**65.2**

Note. *NA implies the cell of interest has zero count.*

research design categories that span cross-sectional, longitudinal, and retrospective type studies.

Second, the remaining 7 of the 22 observational studies (32.8%) received a moderate conclusion validity rating. All seven of these observational studies used a cross-sectional descriptive research design.

Third, thinking in terms of only the 12 observational studies that used a cross-sectional descriptive research design, one can infer that it is almost equally likely that a cross-sectional descriptive speech research study could receive either a moderate or low conclusion validity rating.

Given both the small research synthesis population of speech studies and the Table 5.8 elaborations given above, it is best to declare that there is no clear overall relationship between the three levels of conclusion validity and the six unique categories of research design used in this synthesis population.

Table 5.9 provides a bivariate distribution that classifies each of the 23 speech research studies in terms of both the substantive concerns addressed in the speech research study and the corresponding conclusion validity rating. Inspection of this table suggests that three essential points deserve elaboration.

First, none of the 23 studies that address one of the nine substantive concerns put forth for speech research were judged to have high conclusion validity.

Second, the majority of speech research studies (15 of 23 studies or 65.2%) have a low conclusion validity rating that is distributed across eight of the nine substantive concerns.

Third, the remaining eight studies (34.8% of this synthesis population) have a moderate conclusion validity rating which is distributed across three of the nine substantive concerns.

For these three substantive concerns, the ratios of moderate to low conclusion validity are four to one for the five voice disorder studies, three to one for the four speech disorder studies, and unity (i.e., one moderate and no low ratings) for the sole dysarthria study.

Given these elaborations for the data presented in Table 5.9, moderate conclusion validity ratings are **more likely** for three specific substantive concerns and **less likely** for the other six substantive concerns. Thus, this likelihood difference suggests that there is a small but meaningful relationship between the nine substantive concerns addressed in speech research and the three levels of conclusion validity. However, suggesting that in general it is automatically easier to get moderate conclusion validity ratings for speech perception, voice disorder, and dysarthria research studies is not an inference this exploratory research synthesis can justify, especially in light of the small research synthesis population encountered in this single volume of *JSLHR*.

Table 5.10 provides a bivariate distribution that classifies each of the 23 speech research studies in terms of both the population validity rating and the

corresponding conclusion validity rating. Examination of this table yields four noteworthy elaborations.

First, none of the 23 studies in the speech research synthesis population were found to have simultaneously both high population validity and high conclusion validity ratings. Thus, the upper left position of the principal diagonal has an “NA” entry.

The logical interpretation for this upper left diagonal entry implies that it should be awarded only when a speech study has manifested two established research characteristics. These are **(a)** a methods section narrative for the study that provides an accurate specification of essential sampling concerns (including an explicit specification of the target population, an accurate description of the actual study sample or study participants, and a clear rationale for the study’s sample size) and **(b)** a results section narrative that clearly documents the extent to which the actual sampling design characteristics should qualify or constrain study findings. Given the “NA” entry, this research synthesis failed to uncover any study having these two research characteristics.

Second, the middle principal diagonal entry in Table 5.10 documents that two studies had both moderate population validity ratings and moderate conclusion validity ratings. Accordingly, the narrative in these two studies provided some useful information on sampling concerns and, equally important, shared some comments on how the study findings should be tempered by the limitations of the sampling design.

Third, the lower right entry of the principal diagonal indicates that 15 speech research studies were given both low population validity ratings and low conclusion validity ratings. This means that in each of these 15 studies, neither basic sampling concerns nor the constraints to be placed on study findings due to the actual sampling design characteristics were adequately discussed in the study narrative.

Fourth, six additional studies fell into the center cell below the principal diagonal. This cell indicates that each of these six studies failed to adequately address basic sampling concerns in the methods section narrative, but the narrative in the results section did offer some information regarding how the study findings should be tempered due to limitations associated with the actual sampling design.

The bivariate distribution given in Table 5.10 reveals a strong relationship between population validity and conclusion validity ratings. Thinking in terms of a prediction model, population validity ratings provide an accurate prediction for 17 of the 23 studies in the speech research synthesis population.

In more specific terms, this prediction model declares that population validity ratings are identical to conclusion validity ratings. For example, 15 of the 21 studies having a low population validity rating also have a low conclusion validity rating and both studies having a moderate population validity rating also have a moderate conclusion validity rating.

Since 17 of the 23 studies in this synthesis population have identical ratings for both population and conclusion validity, the accuracy of this prediction model is 73.9 percent.

CHAPTER VI

PHASE FIVE

This chapter presents the findings for phase five which is dedicated to a synthesis of the 59 studies that focus on language research. The first part elaborates the distribution of specific substantive language research concerns encountered in these 59 studies. The second part identifies the research designs used in each study. The third part discusses the 59 population validity ratings. The fourth part discusses the 59 conclusion validity ratings.

Substantive Concerns

An overview of the classification of the 59 language studies by specific substantive concern is provided in Table 6.1. Examination of this table reveals that twelve specific concerns were addressed in these 59 studies.

Over one-third of the language studies (22 of 59 studies or 37.3%) addressed the single substantive concern of language disorders in children. The remaining studies (37 of 59 studies or 62.7%) address eleven additional substantive concerns.

Research Designs

Information on the research designs used in each of the 59 studies focusing on language research concerns is given in Table 6.2 and Table 6.3.

Table 6.1

Substantive Concerns Addressed in Language Research Studies

N=59

Substantive Language Concern	Article Number	Frequency	Percent	Cumulative Percent
Autism / Pervasive Developmental Disorder	9, 48, 67	3	5.1	5.1
Brain activity	11	1	1.7	6.8
Child language disorders	8, 10, 88, 89, 90, 24, 91, 92, 35, 37, 38, 39, 40, 42, 47, 95, 96, 52, 53, 63, 74, 80	**22**	**37.3**	44.1
Child language normative	15, 61, 62	3	5.1	49.2
Gestures	41	1	1.7	50.9
Multilingual/ multicultural issues	14, 27, 49, 50, 51, 66, 104, 105, 82	9	15.3	66.2
Phonology	12, 13, 23, 25, 68	5	8.4	74.6
Assessment	22	1	1.7	76.3
AAC	26, 64, 72, 81	4	6.8	83.1
Adult language	36, 97, 98, 65, 79	5	8.4	91.5
Memory	73, 76	2	3.4	94.9
Morphology	75, 77, 78	3	5.1	100.0
Total		**59**	**100.0**	**--**

Note. *Modal frequency is in bold print.*

These two tables use the 16 category research design classification system developed in the theoretical framework and defined in Table 2.9.

Several key points emerge from inspection of the information in Table 6.2.

Five different research designs are used in both experimental studies and observational studies. Closer inspection, however, reveals that there are over three times as many observation studies (45 studies) as experimental studies (14 studies).

The data in Table 6.2 also reveal that cross-sectional explanatory (30.5%) and cross-sectional descriptive (28.8%) research designs account for almost two-thirds of the research designs (59.3%) used in the 59 language research studies.

Table 6.3 provides a bivariate distribution that classifies each of the 59 language research studies in terms of both research design and substantive concern. Five points are worth highlighting.

First, an examination of the upper panel of Table 6.3 indicates that the 14 experimental studies are distributed over six of the 12 substantive concerns identified for the language research synthesis population.

Second, inspection of the lower panel of Table 6.3 indicates that the 45 observational studies are distributed over 11 of the 12 substantive concerns specified for the 59 language research studies.

Third, a review of the 22 child language disorder studies (see the third data column in Table 6.3) are distributed over seven of the ten research design

Table 6.2

Research Designs Used in Language Research Studies

N=59

Research Design	Frequency	Percent	Cumulative Percent
Experimental Studies			
Experimental Study II	3	5.1	5.1
Intervention Evaluation Research I	1	1.7	6.8
Intervention Evaluation Research II	2	3.4	10.2
Intervention Evaluation Research III	4	6.8	17.0
Diagnostic Instrument Development II	4	6.8	23.8
Observational Studies			
Cross-sectional Descriptive	**17**	**28.8**	52.6
Cross-sectional Predictive	1	1.7	54.3
Cross-sectional Explanatory	**18**	**30.5**	84.8
Longitudinal Descriptive	8	13.5	98.3
Longitudinal Predictive	1	1.7	100.0
Total	**59**	**100.0**	**--**

Note. *The two largest frequencies are in bold print.*

Table 6.3

Language Research Studies Classified by Research Design and Substantive Concern

N=59

Research Design	Autism / Pervasive Developmental Disorder	Brain activity	Child language disorders	Child language normative	Gestures	Multilingual / multicultural issues	Phonology	Assessment	AAC	Adult language	Memory	Morphology	**Total**
Experimental Studies													
Experimental Study II	--	--	2	--	--	--	--	--	--	1	--	--	**3**
Intervention Evaluation Research I	--	--	1	--	--	--	--	--	--	--	--	--	**1**
Intervention Evaluation Research II	--	--	--	--	--	--	--	--	2	--	--	--	**2**
Intervention Evaluation Research III	2	--	--	--	--	--	--	--	1	--	--	1	**4**
Diagnostic Instrument Development II	--	--	2	1	--	--	--	--	1	--	--	--	**4**
Observational Studies													
Cross-sectional Descriptive	1	1	4	1	--	3	1	1	--	2	2	1	**17**
Cross-sectional Predictive	--	--	1	--	--	--	--	--	--	--	--	--	**1**
Cross-sectional Explanatory	--	--	7	1	--	6	3	--	--	1	--	--	**18**
Longitudinal Descriptive	--	--	5	--	1	--	--	--	--	1	--	1	**8**
Longitudinal Predictive	--	--	--	--	--	--	1	--	--	--	--	--	**1**
Total	**3**	**1**	**22**	**3**	**1**	**9**	**5**	**1**	**4**	**5**	**2**	**3**	**59**

alternatives identified for the 59 language research studies. This pattern is not repeated for any other substantive concern specified in this table.

Fourth, inspection of the first row in the lower panel of Table 6.3 uncovers another trend in this research synthesis population; namely, the cross-sectional descriptive research design is used in 17 individual studies that address ten different substantive research concerns. Language research studies that are linked to any of the other nine research design categories are clearly not distributed over this many substantive research concerns.

Verification of the trend uncovered in the fourth elaboration above can be easily accomplished by observing that the third row in the lower panel of Table 6.3 is dedicated to demonstrating that the cross-sectional explanatory research design is used in 18 individual studies that address only five different substantive research concerns. Moreover, each of the eight other research designs used in language research are distributed over fewer than five different substantive concerns.

Fifth, most of the observational studies in the language research domain (36 of 45 studies or 80.0%) use cross-sectional research designs. Only nine observational studies in this domain (20.0%) use longitudinal research designs, with eight of these nine studies all using a longitudinal descriptive research design.

Trends uncovered in the five elaborations shared above will be revisited in phase six that is dedicated to comparing research trends and relationships uncovered in phases three, four, and five.

Population Validity

Information on the population validity ratings assigned to each of the 59 language research studies is given in Tables 6.4 to 6.6. The procedures used in making informed decisions about the 59 population validity assessment ratings are explained in Appendix I.

Table 6.4 provides an initial glimpse of the population validity ratings of the 59 language research studies. Three noteworthy points emerge.

The first thing worth noting is that the language research studies received ratings across all three levels of population validity —high, moderate, and low. However, only 12 of the 59 language studies were considered to be of either high or moderate population validity. The vast majority of studies in this research synthesis (47 of 59 studies or 79.7%) were of low population validity.

Table 6.5 provides a bivariate distribution that classifies each of the 59 language research studies in terms of both research design and population validity rating. Three trends deserve mention.

First, the five studies in Table 6.5 considered to have high population validity were all observational studies.

Table 6.4

Univariate Distribution of Population Validity Ratings of the Language Research Studies

N=59

Population Validity Rating	Frequency	Percent	Cumulative Percent
High Population Validity	5	8.5	8.5
Moderate Population Validity	7	11.9	20.4
Low Population Validity	47	79.6	100.0
Total	**59**	**100.0**	--

Table 6.5

Language Research Studies Classified by Research Design and Population Validity Rating

N=59

Research Design	Population Validity					
	High		Moderate		Low	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
Experimental Studies						
Experimental Study II	--	--	--	--	3	5.1
Intervention Evaluation Research I	--	--	1	1.7	--	--
Intervention Evaluation Research II	--	--	--	--	2	3.4
Intervention Evaluation Research III	--	--	--	--	4	6.8
Diagnostic Instrument Development II	--	--	--	--	4	6.8
Observational Studies						
Cross-sectional Descriptive	--	--	1	1.7	**16**	**27.1**
Cross-sectional Predictive	--	--	1	1.7	--	--
Cross-sectional Explanatory	3	5.1	3	5.1	**12**	**20.3**
Longitudinal Descriptive	2	3.4	--	--	6	10.1
Longitudinal Predictive	--	--	1	1.7	--	--
Total	**5**	**8.5**	**7**	**11.9**	**47**	**79.6**

Note. *The two largest cell frequencies are in bold print.*

Table 6.6

Language Research Studies Classified by Substantive Concern and Population Validity Rating

N=59

Substantive Language Concern	Population Validity					
	High		Moderate		Low	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
Autism/ Pervasive Developmental Disorder	--	--	--	--	3	5.1
Brain activity	--	--	--	--	1	1.7
Child language disorders	3	5.1	4	6.8	15	25.4
Child language normative	--	--	--	--	3	5.1
Gestures	--	--	--	--	1	1.7
Multilingual / multicultural issues	1	1.7	--	--	8	13.6
Phonology	--	--	3	5.1	2	3.4
Assessment	--	--	--	--	1	1.7
AAC	--	--	--	--	4	6.8
Adult language	--	--	--	--	5	8.5
Memory	--	--	--	--	2	3.4
Morphology	1	1.7	--	--	2	3.4
Total	**5**	**8.5**	**7**	**11.9**	**47**	**79.7**

Second, six of the seven studies in Table 6.5 that were awarded moderate population validity ratings were observational studies. The seventh study given a moderate population validity rating was one of the 14 experimental studies identified in the upper panel of this table.

Third, since the 47 language research studies in Table 6.5 having low population validity are distributed across seven of the ten research design categories, it is not wise to declare that there is a meaningful overall relationship between the three levels of population validity and the ten research design categories used in this research synthesis population. Moreover, the fact that 16 cells in this bivariate distribution having 30 actual cells are empty also contributes to the failure to uncover a meaningful overall relationship for these two classification variables.

Table 6.6 provides a bivariate distribution that classifies each of the 59 language research studies in terms of both substantive language research concerns and population validity ratings. An analysis of crosstab data in this table uncovers four points of interest, with each of the first three points of interest linked to a single population validity level.

First, the five language research studies in Table 6.6 having high population validity are distributed over just three of the 12 substantive concerns identified in this table.

Second, the seven language research studies in Table 6.6 having moderate population validity can be linked to only two of the 12 substantive research concerns identified in this table.

Third, the remaining 47 language research studies in Table 6.6 are distributed over all 12 substantive concerns specified in this table. Accordingly, it is unwise to advance the possibility that there is an overall meaningful relationship between the three levels of population validity and the 12 substantive concern categories.

Four, on the other hand, what can be advanced as a meaningful trend is the fact that the most likely population validity rating for 11 of the 12 substantive language research concerns is a low population validity rating. This consistent likelihood also adds clear evidence to support the lack of a meaningful overall relationship between the two variables of population validity and substantive concern.

Conclusion Validity

Information on the conclusion validity ratings assigned to each of the 59 language research studies is given in Tables 6.7 to 6.10. The procedures used in making informed decisions about these 59 conclusion validity assessment ratings are explained in Appendix J.

Table 6.7 provides a distribution of the conclusion validity ratings for the 59 language research studies. Four key points deserve mention here.

Table 6.7

Univariate Distribution of Conclusion Validity Ratings of the Language Research Studies

N=59

Conclusion Validity Rating	Frequency	Percent	Cumulative Percent
High Conclusion Validity	4	6.8	6.8
Moderate Conclusion Validity	12	20.3	27.1
Low Conclusion Validity	43	72.9	100.0
Total	**59**	**100.0**	--

Note. *Modal frequency is in bold print.*

Table 6.8

Language Research Studies Classified by Research Design and Conclusion Validity Rating

N= 59

Research Design	Conclusion Validity					
	High		Moderate		Low	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
Experimental Studies						
Experimental Study II	--	--	--	--	3	5.1
Intervention Evaluation Research I	--	--	1	1.7	--	--
Intervention Evaluation Research II	--	--	--	--	2	3.4
Intervention Evaluation Research III	1	1.7	2	3.4	1	1.7
Diagnostic Instrument Development II	--	--	1	1.7	3	5.1
Observational Studies						
Cross-sectional Descriptive	--	--	5	8.5	**12**	**20.3**
Cross-sectional Predictive	--	--	--	--	1	1.7
Cross-sectional Explanatory	2	3.4	2	3.4	**14**	**23.7**
Longitudinal Descriptive	1	1.7	1	1.7	6	10.2
Longitudinal Predictive	--	--	--	--	1	1.7
Total	**4**	**6.8**	**12**	**20.3**	**43**	**72.9**

Note. *The two largest frequencies are in bold print.*

Table 6.9

Language Research Studies Classified by Substantive Concern and Conclusion Validity Rating

N= 59

Substantive Language Concern	Conclusion Validity					
	High		Moderate		Low	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
Autism/ Pervasive Developmental Disorder	1	1.7	1	1.7	1	1.7
Brain activity	--	--	--	--	1	1.7
Child language disorders	2	3.4	4	6.8	16	27.1
Child language normative	--	--	1	1.7	2	3.4
Gestures	--	--	--	--	1	1.7
Multilingual / multicultural issues	1	1.7	--	--	8	13.6
Phonology	--	--	1	1.7	4	6.8
Assessment	--	--	1	1.7	--	--
AAC	--	--	--	--	4	6.8
Adult language	--	--	--	--	5	8.4
Memory	--	--	1	1.7	1	1.7
Morphology	--	--	3	5.0	--	--
Total	**4**	**6.8**	**12**	**20.3**	**43**	**72.9**

Table 6.10

Language Research Studies Classified by Population Validity Rating and Conclusion Validity Rating

N= 59

Population Validity Rating	Conclusion Validity Rating					
	High		Moderate		Low	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
High Population Validity	3	5.1	2	3.4	--	--
Moderate Population Validity	--	--	2	3.4	5	8.5
Low Population Validity	1	1.7	8	13.5	38	64.4
Total	**4**	**6.8**	**12**	**20.3**	**43**	**72.9**

First, the 59 conclusion validity ratings shown in Table 6.7 are distributed across three ordinal levels—high, moderate, and low. Second, few of the language research studies (4 of 59 studies or 6.8%) were judged to have high conclusion validity.

Third, approximately one-fifth of the language research studies (12 of 59 studies or 20.3%) were considered to have moderate conclusion validity.

Fourth, almost three-fourths of the language research studies (43 of 59 studies or 72.9%) were considered to have low conclusion validity.

This univariate distribution of 59 conclusion validity ratings provides one of the two marginal distributions for three separate bivariate distributions introduced below as a means to link conclusion validity with each of the other three variables in the research synthesis. These variables (in the order introduced below) are research design, substantive concern addressed in language research studies, and population validity.

Table 6.8 provides a bivariate distribution that classifies each of the 59 language research studies in terms of both research design and conclusion validity rating. Four interesting outcomes emerge.

First, only one of the four language research studies awarded a high conclusion validity rating was an experimental study. Accordingly, a closer examination of the upper panel of Table 6.8 allows one to conclude that just one of 14 language research studies that used an experimental research design (7.1% of all 14 studies) had a high conclusion validity rating.

Second, the other three language research studies that had a high conclusion validity rating were observational studies. A closer look at the lower panel of Table 6.8 indicates that just three of 45 language research studies that used an observational research design (6.7% of all 45 studies) had a high conclusion validity rating.

Third, a comparison of the first and second elaborations above yields two interesting outcomes. Specifically, language research studies that use either an experimental or observational research design have an almost equal likelihood (approximately 7.0% percent possibility) of receiving a high conclusion validity rating. Also, the four language research studies having a high conclusion validity rating are distributed over just three of the ten research design categories specified in Table 6.8.

Fourth, a large majority of language research studies (43 of 59 studies or 72.9%) were given low conclusion validity ratings. These 43 studies were distributed over all ten research design categories in Table. 6.9.

Given **(a)** the equal proportion of experimental and observational studies having a high conclusion validity rating, **(b)** the distribution of 43 low conclusion validity rating studies distributed across all ten research designs used in this research synthesis population, and **(c)** the fact that a low conclusion validity rating was the most likely rating for eight of the ten research design categories, it is quite clear that there is no overall meaningful relationship between the three

levels of conclusion validity and the ten research design categories uncovered in the language research synthesis population.

Table 6.9 provides a bivariate distribution that classifies each of the 59 language research studies in terms of both the substantive concerns addressed in the language research study and the corresponding conclusion validity rating. Analysis of the data in this table yields five specific findings which are elaborated below.

First, the four language research studies in Table 6.9 judged to have high conclusion validity were distributed over three substantive language research concerns.

Second, the 12 language research studies in Table 6.9 judged to have moderate conclusion validity were distributed over seven of the 12 substantive language research concerns specified in this table.

Third, the large majority of 43 studies in Table 6.9 having low conclusion validity were distributed over ten of the 12 substantive language research concerns specified in this table.

Fourth, for eight of the 12 substantive language concern categories specified in Table 6.9, the most likely conclusion validity rating is a low rating.

Fifth, since a large percentage of language research studies (22 of 59 studies or 37.3%) address child language disorders, it was important to examine this group of studies separately. This independent analysis indicates that **(a)** the ratio of low conclusion validity studies to moderate conclusion validity studies is

four to one and **(b)** the ratio of low conclusion validity studies to high conclusion validity studies is eight to one. Thus, the overall trend given in the fourth elaboration is consistent with the two ratios given here for the 22 child language disorder studies.

Taken collectively, information provided in the five elaborations above yield ample evidence to support the claim that there is no overall meaningful relationship between the three levels of conclusion validity and the 12 substantive concern categories uncovered in the language research synthesis population.

Table 6.10 provides a bivariate distribution that classifies each of the 59 language research studies in terms of both the population validity rating and the corresponding conclusion validity rating. Examination of this table yields seven noteworthy elaborations.

Uncovering trends and relationships for the two variables of interest in Table 6.10 begin by noticing that 43 of the 59 language research studies (72.9%) reside on the principal diagonal of this table. With this in mind, the first three elaborations below interpret the three frequency counts on this diagonal.

First, only three of the 43 diagonal counts reside in the upper left position of the principal diagonal. Accordingly, just three of 59 studies in the language research synthesis population (5.1%) have both a **methods section narrative** that provides an accurate specification of all essential sampling concerns and a

results section narrative that clearly documents the extent to which the actual sampling design characteristics should qualify or constrain study findings.

Second, just two of the 43 diagonal counts reside in the middle position of the principal diagonal. Thus, just two of the 59 studies in the language research synthesis population (3.4%) have a **methods section narrative** that provides some useful information on sampling concerns and a **results section narrative** that shares some useful comments on how the study findings should be tempered by the limitations of the sampling design.

Third, the remaining 38 diagonal counts reside in the lower right position of the principal diagonal. Accordingly, 38 of the 59 studies in the language research synthesis population (64.4%) have neither an adequate methods section narrative nor an adequate results section narrative.

Four off-diagonal entries in Table 6.10 are used to provide the evidence needed to construct the four final elaborations that follow.

Fourth, nine counts in Table 6.10 reside below the principal diagonal. Eight of these nine counts reside in a single cell. This table entry indicates that eight of the 59 studies in the language research synthesis population (13.6%) have a low population validity rating but a moderate conclusion validity rating. This is the case where eight of these studies failed to adequately address basic sampling concerns in the methods section narrative, but the narrative in the results section did offer some useful information regarding how the study

findings should be tempered due to limitations associated with the actual sampling design.

Fifth, one count in Table 6.10 resides in the remaining cell below the principal diagonal. This indicates that this single study (1 of 59 or 1.7%) has a low population validity rating but a high conclusion validity rating. This means that an accurate description of the actual study sample and a clear specification of the corresponding target population were not discussed in the methods section. However, because the results section narrative offered clear restrictions of the study findings, a high conclusion validity rating was afforded the study.

Sixth, two counts in Table 6.10 reside in a single cell above the principal diagonal. This table entry indicates that two of the 59 studies in the language research synthesis population (3.4%) have a high population validity rating and a moderate conclusion validity rating. Thus, each of these two studies provides a methods section narrative that has an accurate identification of all essential sampling concerns and a results section narrative that has useful information indicating how the study findings should be tempered by the limitations of the sampling design.

Seventh, Table 6.10 shows the five remaining studies of the 59 studies in the language research synthesis population (8.5%) contained in a single cell above the principal diagonal. This cell is characterized as having moderate population validity and low conclusion validity. This indicates that there was

some description of sampling concerns in the methods section but findings were not tempered in the discussion of results.

The bivariate distribution given in Table 6.10 reveals a meaningful relationship between population validity and conclusion validity ratings. Thinking in terms of a prediction model, population validity ratings provide an accurate prediction for 43 of the 59 studies that fall in the principal diagonal in the language research synthesis population.

In more specific terms, this prediction model declares that population validity ratings are identical to conclusion validity ratings. For example, 38 of the 47 studies having a low population validity rating also have a low conclusion validity rating. Similarly, two of seven studies having a moderate population validity rating also have a moderate conclusion validity rating. Likewise, three of the five studies having a high population validity rating also have a high conclusion validity rating.

Since 43 of 59 studies have identical ratings for both population and conclusion validity, the accuracy of this prediction model is 72.9 percent.

CHAPTER VII

PHASE SIX

This chapter explores trends that have been exposed across the three preceding chapters. To provide a parallel to phases three (hearing), four (speech) and five (language), these trends are developed in four distinct parts. Accordingly, these four parts address trends across substantive concerns, research designs, population validity, and conclusion validity.

Substantive Concerns

The first part of the trend analysis looks at substantive concerns. The substantive concerns that have been explored in the three preceding chapters are specific to the three research domains of hearing, speech, and language. As such, it would be unnecessary to identify trends across the substantive concerns of the three research domains. Accordingly, no additional exploration is needed.

Research Designs

The second part of the trend analysis provides information about the research designs used in the overall research synthesis population of 105 studies. The information contained in Table 7.1 uses the 16 category research design classification system developed in the theoretical framework and defined in Table 2.9. When analyzing Table 7.1, six trends are worth noting.

Table 7.1

Trends in Research Designs for Hearing, Speech, and Language Research Studies

N=105

Research Design	Hearing		Speech		Language		Total	
	Frequency	Percent	Frequency	Percent	Frequency	Percent	Frequency	Percent
Experimental								
Experimental Study I	--	--	--	--	--	--	--	--
Experimental Study II	2	8.6	--	--	3	5.1	5	4.8
Intervention Evaluation Research I	--	--	--	--	1	1.7	1	0.9
Intervention Evaluation Research II	--	--	--	--	2	3.4	2	1.9
Intervention Evaluation Research III	--	--	1	4.4	4	6.8	5	4.8
Diagnostic Instrument Development I	--	--	--	--	--	--	--	--
Diagnostic Instrument Development II	--	--	--	--	4	6.8	4	3.8
Observational								
Cross-sectional Descriptive	**9**	**39.1**	**13**	**56.4**	**17**	**28.8**	**39**	**37.1**
Cross-sectional Predictive	1	4.4	--	--	1	1.7	2	1.9
Cross-sectional Explanatory	**8**	**34.7**	**6**	**26.0**	**18**	**30.5**	**32**	**30.5**
Longitudinal Descriptive	--	--	1	4.4	8	13.5	9	8.7
Longitudinal Predictive	1	4.4	--	--	1	1.7	2	1.9
Longitudinal Explanatory	1	4.4	1	4.4	--	--	2	1.9
Retrospective Descriptive	1	4.4	--	--	--	--	1	0.9
Retrospective Predictive	--	--	--	--	--	--	--	--
Retrospective Explanatory	--	--	1	4.4	--	--	1	0.9
Total	**23**	**100.0**	**23**	**100.0**	**59**	**100.0**	**105**	**100.0**

Note. *Two largest frequencies for each substantive research topic in observational studies are in bold.*

First, data in the top panel of Table 7.1 indicate that overall only 17 of the 105 studies (16.2%) used experimental research designs.

Second, data in the top panel of Table 7.1 indicate that language research studies were more likely to use experimental research designs than speech or hearing research studies. Specifically, 23.7 percent of language research studies (14 of 59 studies) used experimental research designs; whereas, experimental research designs were used in 8.7 percent of hearing research studies (2 of 23 studies) and 4.3% of speech research studies (1 of 23).

Third, data in the bottom panel of Table 7.1 indicate that 88 of the 105 studies in the overall research synthesis population (83.8%) used observational (nonexperimental) research designs.

Fourth, information in the bottom panel of 7.1 shows across all three substantive research domains, a large majority of studies used an observational research design. Specifically, 95.7 percent of speech research studies (22 of 23 studies), 91.3 percent of hearing research studies (21 of 23 studies) and 76.3% of language research studies (45 of 59 studies) used an observational research design.

Fifth, across all three substantive research domains as shown in the bottom panel of Table 7.1, a large majority of observational (nonexperimental) studies used cross-sectional rather than either longitudinal or retrospective research designs. Specifically, using the number of observational studies in

each domain as the base, 90.5% of speech research studies (19 of 21 studies), 85.7% of hearing research studies (18 of 21 studies), and 90.0% of language research studies (36 of 45 studies) used these two cross-sectional research designs.

Sixth, another look at the bottom panel of Table 7.1 indicates that the two most common research designs used across the three substantive research domains are cross-sectional descriptive (39 of 105 studies or 37.1%) and cross-sectional explanatory (32 of 105 studies or 30.5%) research designs. Accordingly, these two research designs are used in every two out of three studies in the overall research synthesis population.

Population Validity

The third part of the trend analysis provides information about the population validity assessments for the research synthesis population of 105 studies. Summary of findings from the three previous phases are reported in Tables 7.2 and 7.3.

A comparative analysis of the information given in Table 7.2 yields a single striking trend; namely, a large majority of studies in each of the substantive research domains received a low rather than either a high or moderate population validity rating. In more specific terms, 89 of the 105 studies in the overall research synthesis population (84.8%) had low population validity ratings.

Table 7.2

Crosstabs for Substantive Research Topic by

Population Validity Rating

N=105

Substantive Research Topic	Population Validity					
	High		Moderate		Low	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
Hearing	--	--	2	1.9	21	20.0
Speech	--	--	2	1.9	21	20.0
Language	5	4.8	7	6.6	47	44.8
Total	**5**	**4.8**	**11**	**10.4**	**89**	**84.8**

Note. *Frequency and percent values are for individual cells. Thus, the nine crosstab entries in this table taken collectively will yield 100 percent.*

Table 7.3

Crosstabs for Type of Research Design by Substantive Research Topic by Population Validity Rating

N=105

	Population Validity Rating						**Total**	
	High		Moderate		Low			
	Frequency	Percent	Frequency	Percent	Frequency	Percent	Frequency	Percent
Experimental								
Hearing	--	--	--	--	2	100.0	**2**	**100.0**
Speech	--	--	--	--	1	100.0	**1**	**100.0**
Language	--	--	1	7.1	13	92.9	**14**	**100.0**
All	--	--	1	5.9	16	94.1	**17**	**100.0**
Observational								
Hearing	--	--	2	10.5	19	90.5	**21**	**100.0**
Speech	--	--	2	9.1	20	90.9	**22**	**100.0**
Language	5	11.1	6	13.3	34	75.6	**45**	**100.0**
All	5	5.7	10	11.4	73	82.9	**88**	**100.0**
Total	**5**	**4.8**	**11**	**10.4**	**89**	**84.8**	**105**	**100.0**

The individual percents of low population validity ratings for each of the three substantive research domains are provided in Table 7.2. Specifically, 79.7 percent of language research studies (47 of 59 studies), 91.3% of hearing research studies (21 of 23 studies), and 91.3% of speech research studies (21 of 23 studies) had low population validity ratings. Clearly, these percents support the use of the descriptor “large majority” in the striking trend statement offered in the previous paragraph.

This trend suggests that only the remaining 16 of the 105 studies in the overall research synthesis population (15.2%) provided at least an adequate description of essential sample concerns in the research methods section of their journal article.

Table 7.3 provides a cumulative summary of the type of research design used (experimental or observational) by substantive research topic (hearing, speech, or language) by the population validity rating (high, moderate, or low).

Overall, the data included in Table 7.3 show that across all three substantive research topics using either type of research design, only 16 studies (16 of 105 of 15.2%) received either a high or moderate rating of population validity.

The top panel of Table 7.3 looks at the 17 studies using experimental research design. Of the 17 studies only one study (1 of 17 or 5.9%) using an experimental research design was found to have moderate population validity

rating. The large majority (16 of 17 studies or 94.1%) of experimental studies were rated as having low population validity.

The second panel of Table 7.3 looks at the 88 studies using observational research designs. Studies using observational research designs were rated either as high, moderate, and low population validity. However, the large majority (73 of 88 studies or 82.9%) had a rating of low population validity.

Conclusion Validity

The fourth part of the trend analysis provides information about the conclusion validity assessment of the research synthesis population of 105 studies. A summary of these trend analysis findings is given in the next four tables.

Information given in Table 7.4 reveals that the large majority (79 of 105 studies or 75.3%) of the research synthesis population of studies received low conclusion validity ratings. Thinking in terms of the three substantive research topics, 91.3 percent of hearing research studies (21 of 23 studies), 65.2 percent of speech research studies (15 of 23 studies), and 72.9 percent of language research studies (43 of 59 studies) had low conclusion validity ratings. These three proportions indicate that the overall trend of low conclusion validity ratings also holds across all three substantive research topics.

In comparison, about one-fourth of the research synthesis population of studies (26 of 105 studies or 24.7%) were rated as having high or moderate

Table 7.4

Substantive Research Topic by Conclusion Validity Rating

N=105

Substantive Research Topic	Conclusion Validity					
	High		Moderate		Low	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
Hearing	--	--	2	1.9	21	20.0
Speech	--	--	8	7.6	15	14.3
Language	4	3.8	12	11.4	43	41.0
Total	**4**	**3.8**	**22**	**20.9**	**79**	**75.3**

Note. *Frequency and percent values are for individual cells. Thus, the nine crosstab entries in this table taken collectively will yield 100 percent*.

conclusion validity. This trend suggests that the results section narrative of only 26 studies in the overall research synthesis population of 105 studies provided any relevant information regarding how the actual sampling design should limit or constrain the research findings presented for the study.

Table 7.5 provides a cumulative summary of the type of research design used (experimental or observational) across the substantive research topics (hearing, speech, or language). Combining these two variables yields six independent types of studies. Specifically, there are three types of experimental studies (experimental hearing, experimental speech, and experimental language) and there are three types of observational studies (observational hearing, observational speech, and observational language).

The bottom line in Table 7.5 shows that only 26 of 105 studies (24.8%) received a high or moderate conclusion validity rating. In contrast, the large majority (79 of 105 studies or 75.2%) of the research synthesis study population were identified as having a low conclusion validity rating.

The “all” entry in the top panel of Table 7.5 indicates that when taken collectively a large majority of the experimental studies (11 of 17 studies or 64.7%) distributed across the three substantive research topics (hearing, speech, or language) received low conclusion validity ratings. Thus, the overall trend for conclusion validity stated in the previous paragraph also holds for these 17 experimental studies.

Table 7.5

Crosstabs for Type of Research Design by Substantive Research Topic by Conclusion Validity Rating

N=105

	Conclusion Validity Rating						**Total**	
	High		Moderate		Low			
	Frequency	Percent	Frequency	Percent	Frequency	Percent	Frequency	Percent
Experimental								
Hearing	--	--	--	--	2	100.0	**2**	**100.0**
Speech	--	--	1	100.0	--	--	**1**	**100.0**
Language	1	7.1	4	28.6	9	64.3	**14**	**100.0**
All	1	5.9	5	29.4	11	64.7	**17**	**100.0**
Observational								
Hearing	--	--	2	9.5	19	90.5	**21**	**100.0**
Speech	--	--	7	31.8	15	68.2	**22**	**100.0**
Language	3	6.7	8	17.8	34	75.5	**45**	**100.0**
All	3	3.4	17	19.3	68	77.3	**88**	**100.0**
Total	**4**	**3.8**	**22**	**21.0**	**79**	**75.2**	**105**	**100.0**

The “all” entry in the bottom panel of Table 7.5 indicates that when taken collectively a large majority of observational studies (68 of 88 studies or 77.3%) distributed across the three substantive research topics (hearing, speech, or language) received low conclusion validity ratings. Thus, the overall trend for conclusion validity stated above also holds for these 88 observational studies.

A final way to analyze the disaggregated data for the 105 studies is to think of the six independent types of studies referenced in Table 7.5 as a single inventory having six groups of studies. Specifics for this trend analysis are offered below.

When taken collectively, five out of the six independent types of studies in Table 7.5 exhibit a large majority of low conclusion validity ratings. Specifically, 100 percent of experimental hearing studies (2 of 2 studies), 64.3 percent of experimental language studies (9 of 14 studies), 90.5 percent of observational hearing studies (19 of 21 studies), 68.2 percent of observational speech studies (15 of 22 studies), and 75.5 percent of observational language studies (34 of 45 studies) had low conclusion validity ratings.

Only one of the six independent types of studies (experimental speech) in Table 7.5 does not follow the trend elaborated in the preceding paragraph. Specifically, 100 percent of experimental speech studies (1 of 1 study) had a moderate rather than a low conclusion validity rating. However, this single case yields almost no direct influence on the findings elaborated in the previous paragraph.

Given all the individual Table 7.5 analyses offered above, a single trend emerges; namely, the overall trend generated for the single set of 105 studies holds for all disaggregated groups (all 17 experimental studies, all 88 observational studies, and all studies specified as six independent types of studies). Accordingly, both the variable substantive research topic (with three levels) and the variable population validity rating (with three levels) are not related to the third variable indicating the conclusion validity rating.

This trend suggests that the results section narrative of only 26 studies in the overall research synthesis population of 105 studies provided any relevant information regarding how the actual sampling design should limit or constrain the research findings presented for the study.

Table 7.6 provides another multivariate perspective. In this case, the synthesis variables substantive research topic with three levels and population validity rating with three levels are combined to form nine independent groups. These nine groups of studies are organized into three major panels, with each panel focusing on just three of these nine independent groups.

The multivariate relationship of interest in Table 7.6 is completed by using these nine independent research synthesis study groups to disaggregate the univariate distribution of 105 conclusion validity ratings.

Using the “all” row entries of the first panel of Table 7.6 indicates that the most likely combined conclusion validity rating for all three groups of hearing studies (21 of 23 studies or 91.3%) is a low population validity rating.

Table 7.6

Crosstabs for Substantive Research Topic by Population Validity Rating by Conclusion Validity Rating

N=105

Substantive Research Topic	Population Validity Rating	Conclusion Validity Rating: High Frequency	High Percent	Moderate Frequency	Moderate Percent	Low Frequency	Low Percent	**Total** Frequency	**Total** Percent
Hearing	High	--	--	--	--	--	--	**--**	**--**
	Moderate	--	--	1	50.0	1	50.0	**2**	**100.0**
	Low	--	--	1	4.8	20	95.2	**21**	**100.0**
	All	--	--	2	8.7	21	91.3	**23**	**100.0**
Speech	High	--	--	--	--	--	--	**--**	**--**
	Moderate	--	--	2	100.0	--	--	**2**	**100.0**
	Low	--	--	6	28.6	15	71.4	**21**	**100.0**
	All	--	--	8	34.8	15	65.2	**23**	**100.0**
Language	High	3	60.0	2	40.0	--	--	**5**	**100.0**
	Moderate	--	--	2	28.6	5	71.4	**7**	**100.0**
	Low	1	2.1	8	17.0	38	80.9	**47**	**100.0**
	All	4	6.8	12	20.3	43	72.9		
	Total	**4**	**3.8**	**22**	**21.0**	**79**	**75.2**	**105**	**100.0**

Similarly, using the “all” row entries of the second panel of Table 7.6 indicates that the most likely combined conclusion validity rating for all three groups of speech studies (15 of 23 studies or 65.2%) is a low population validity rating.

Finally, the “all” row entries in the third panel of Table 7.6 also reveals that the most likely combined conclusion rating for all three groups of language studies (43 of 59 studies or 72.9%) is a low population validity rating.

The identical likelihood results for each of the three panels in Table 7.6 suggests that the categorical variable having nine independent study group categories is not related to the third research synthesis variable reflecting conclusion validity ratings.

An additional insightful way to analyze the disaggregated (multivariate) data in Table 7.6 for the 105 studies is to analyze nine independent types of studies where each of the nine types of studies has its own conclusion validity rating distribution.

This additional analysis indicates that four of the nine independent categories in Table 7.6 contain 96 of the 105 research synthesis studies (91.4%). In each of these four categories, the most likely conclusion validity rating is a low rating. Specifically, 95.2 percent of the hearing research studies having low population validity (20 of 21 studies), 71.4 percent of the speech research studies having low population validity (15 of 21 studies), 71.4 percent of the language research studies having moderate population validity (5 of 7

studies), and 80.9 percent of the language research studies having low population validity (38 of 47 studies) had low conclusion validity ratings.

Review of the remaining five of nine independent types of studies does little to alter the overall pattern uncovered in the previous paragraph. Specific reasons follow.

Two of the nine Table 7.6 categories have no studies, and therefore make no contribution to the overall pattern.

The three remaining Table 7.6 categories account for just 9 of the 105 studies in this research synthesis. Analyzing each of these categories separately yields these proportions. In row two (hearing, moderate) one of the two studies has a low conclusion validity rating. In row five (speech, moderate), neither of the two studies in this category has a low conclusion validity rating. In row seven (language, high), none of the five studies have a low conclusion validity rating.

This more detailed disaggregate analysis involving nine independent research synthesis study groups lends more pervasive evidence to support the position that these nine types of studies viewed as a single categorical variable are unrelated to the overall distribution of 105 conclusion validity ratings.

The final multivariate perspective developed in the sixth phase of the research synthesis uses a prediction model format. The predictor (independent) variable is the population validity rating. The outcome (independent) variable is the conclusion validity rating. The moderator variable of interest in this prediction model framework is the substantive research topic.

The predictor model findings are developed in two steps. The first step (see Table 7.7) is used to specify the overall prediction using all 105 studies. The second step (see Table 7.8) uses the moderator variable (substantive research topic) to specify three independent predictor models. These models allow one to decide if in fact the moderator variable has unique explanatory power. This power can occur if and only if the three independent prediction models have substantially different prediction outcomes.

Table 7.7 provides an overall bivariate distribution that classifies all of the 105 studies in the research synthesis population in terms of both the population validity rating and the corresponding conclusion validity rating. Examination of this table yields six noteworthy elaborations.

First, the upper left diagonal entry in Table 7.7 reveals that three of the 105 studies in the research synthesis population were found to have simultaneously both high population validity and high conclusion validity ratings.

The logical interpretation for this upper left diagonal entry implies that it should be awarded only when a study has manifested two established research characteristics. Specifically, (a) the methods section narrative for the study provides an accurate specification of essential sampling concerns (including an explicit specification of the target population, an accurate description of the actual study sample or study participants, and a clear rationale for the study's sample size) and (b) the results section narrative clearly documents the extent to which the actual sampling design characteristics should qualify or constrain

Table 7.7

The Bivariate Distribution that Describes the Independent and Dependent Variables Used in the Overall Prediction Model

N=105

Population Validity Rating	Conclusion Validity Rating			
	High	Moderate	Low	
High	**3**	2	--	Since **81** of **105** studies in the research synthesis population have identical ratings for both population validity and conclusion validity the accuracy of prediction is **77.1 percent**.
Moderate	--	**5**	6	
Low	1	15	**73**	

Table 7.8

The Three Bivariate Distributions that Describe the Prediction Models

Where Substantive Research Topic is Used as a Moderator Variable

Substantive Research Topic of Hearing				Since **21** of **23** hearing studies have identical ratings for both population validity and conclusion validity the accuracy of prediction for hearing studies is **91.3 percent**.
Population Validity Rating	Conclusion Validity Rating			
	High	Moderate	Low	
High	--	--	--	
Moderate	--	**1**	1	
Low	--	1	**20**	
Substantive Research Topic of Speech				Since **17** of **23** speech studies have identical ratings for both population validity and conclusion validity the accuracy of prediction for hearing studies is **73.9 percent**.
Population Validity Rating	Conclusion Validity Rating			
	High	Moderate	Low	
High	--	--	--	
Moderate	--	**2**	--	
Low	--	6	**15**	
Substantive Research Topic of Language				Since **43** of **59** language studies have identical ratings for both population validity and conclusion validity the accuracy of prediction for hearing studies is **72.9 percent**.
Population Validity Rating	Conclusion Validity Rating			
	High	Moderate	Low	
High	**3**	2	--	
Moderate	--	**2**	5	
Low	1	8	**38**	

study findings. Given the overall entry provided in Table 7.7 for studies across all three substantive research topics, only three studies in this research synthesis contained these two research characteristics.

Second, the middle principal diagonal entry in Table 7.7 documents that five studies had both moderate population validity ratings and moderate conclusion validity ratings. Accordingly, the narrative in these five studies provided some useful information on sampling concerns and, equally important, shared some relevant comments on how the study findings should be tempered by the actual limitations of the sampling design.

Third, the lower right entry of the principal diagonal in Table 7.7 indicates that 73 studies in the research synthesis population of 105 studies were given both low population validity ratings and low conclusion validity ratings. This means that in each of these 73 studies, neither basic sampling concerns nor the constraints to be placed on study findings due to the actual sampling design characteristics were adequately discussed in the study narrative.

Fourth, 16 cases (16 of 105 studies or 15.3%) fell below the principal diagonal in Table 7.7. These two cell locations having nonzero count reveal that 16 studies were evaluated as having lower population validity ratings than their corresponding conclusion validity ratings. Accordingly, these cells indicate that each of these 16 studies failed to adequately address basic sampling concerns in the methods section narrative, but the narrative in the results section did offer

relevant information regarding how the study findings should be tempered due to limitations associated with the actual sampling design.

Fifth, the eight cases (8 of 105 studies or 7.6%) above the principal diagonals in Table 7.7 were evaluated as having a higher population validity rating than the corresponding conclusion validity rating. These cell locations indicate that each of these eight cases included some description of sampling concerns in the methods section; however, findings were not clarified in the discussion of results.

Finally, the information given in Table 7.7 reveals a strong overall relationship between population validity and conclusion validity ratings across the three substantive concerns. Thinking in terms of a prediction model, population validity ratings provide an accurate prediction for 81 of the 105 studies (77.1%) in the total research synthesis population.

In more specific terms, this prediction model declares that population validity ratings are identical to conclusion validity ratings. For example, 73 of the 89 studies having a low population validity rating also have a low conclusion validity rating. Additionally, five of the 11 studies having a moderate population validity rating also have a moderate conclusion validity rating. Likewise, three of the five studies having a high population validity rating also have a high conclusion validity rating.

Since 81 of the 105 studies have identical ratings for both population and conclusion validity across the three substantive concerns, the accuracy of this prediction model is 77.1 percent.

Table 7.8 provides the three bivariate distributions that yield the information needed to generate three disaggregated but independent prediction models. For example, the first panel provides a bivariate crosstab table for the 23 research synthesis studies addressing hearing research topics. In this case, the accuracy of the prediction model is 91.3 percent since 21 of the 23 studies have identical ratings for population and conclusion validity.

Similarly, the second panel of Table 7.8 provides a bivariate crosstab table for 23 studies addressing speech research concerns. Here the accuracy of this prediction model is 73.9 percent.

Likewise, the third panel of this table documents the bivariate relationship for the 59 language research studies. In this case, the accuracy of the prediction model is 72.9 percent.

Since all three of the disaggregated prediction models indicate that a substantive majority of studies share the same prediction status (i.e., inferring that population validity ratings are an accurate predictor of conclusion validity ratings) of between 72.9 percent and 91.3 percent, the moderator variable substantive research topic does not have unique explanatory power. Thus, the overall prediction generate in step one holds also for the three subgroups of studies created by focusing on the different substantive research topics.

CHAPTER VIII

PHASE SEVEN

This chapter provides information for phase seven which is dedicated to making recommendations for conducting future research, to making recommendations for practice, and a recommendation for training. To reflect these three objectives, the chapter is divided into three parts.

Recommendations for Conducting Future Research Synthesis Studies

An inventory of five recommendations for maintaining this research agenda are discussed below.

Replication

This scholarly endeavor is a preliminary analysis of published research in the *Journal of Speech, Language, and Hearing Research* (*JSLHR*). Not only is this the first comprehensive evaluation of the articles published in the *JSLHR*, but this research synthesis uses three classification systems that were developed specifically to fulfill the objectives of this study.

Recommendation One. ***Another year of the JSLHR should be assessed to see these research synthesis findings can be replicated.***

Replication of this study is feasible because the three coding procedures defined and used in this study are fully defined. Specifically, coding procedures for (a) research designs are established in Appendix H, (b) population validity are provided in Appendix I, and (c) conclusion validity are specified in Appendix J.

Evaluation of Statistical Evidence

This scholarly endeavor focused on three synthesis domains: research designs, population validity, and conclusion validity. Statistical methods were only given a cursory rather than an explicitly detailed look when evaluating both population validity and conclusion validity. Clearly defining and evaluating statistical methods used and documenting the statistical data interpretations reported would provide additional relevant information about studies published in the *JSLHR*.

Recommendation Two. ***A parallel study that focuses explicitly on evaluating both the statistical methods used and the interpretation of statistical data reported in the JSLHR could be done.***

Using a similar theoretical framework presented in this methodological research synthesis, a classification system could be developed that would allow researchers to make an informed decision about the appropriate selection of

statistical methods and findings that are reported within studies. This information would provide additional facts about conclusions drawn within a study.

Analysis of Trends

Trends within and across the three topics of hearing, speech, and language have been identified for Volume 47 of the *JSLHR*. It is important to determine if changes in identified trends have occurred in subsequent years. Trends provide information about topics of research interest, and more importantly, information about validity of conclusions drawn in studies.

Recommendation Three. ***A five year scope of the JSLHR can be reviewed for additional information about trends both within and across the three major substantive topics of hearing, speech, and language.***

Three classification systems were developed in the scope of this study. Specifically, a classification system for research designs, a classification system to determine population validity, and a classification system to make a determination about conclusion validity were reported. These classification systems can be used to evaluate articles and studies in subsequent years. This information will provide longitudinal information about the fields of speech pathology and audiology.

Quantitative Validity Assessments

The assessment of both population validity and conclusion validity in this study are based on informed judgments. The process of making determinations, even though well-defined and reliable (based upon the high rate of agreement among raters), are derived from opinions and perceptions. It would be beneficial to create a more accurate method of making judgments about population validity and conclusion validity.

Recommendation Four. ***Create more accurate quantitative methods for making assessments about population validity and conclusion validity based on scales rather than informed judgments.***

Eight essential questions to inform decisions about population validity are presented in Table 2.10 and four evaluation components to assess conclusion validity are provided in Table 2.11. This information can be used as the foundation for developing scales to make judgments about population validity and conclusion validity.

Transferability of Research Synthesis Methods

Other professions are asked to engage in evidence-based practices. Although the four previous recommendations have focused on the *JSLHR*, this research methodology could be used in other journals.

Recommendation Five. ***The research synthesis methodology developed in this study can be used to assess the research designs, ratings of population validity, and conclusion validity of articles published in other journals.***

The procedures to develop and validate the three classification systems used in this methodological research synthesis are presented in detail in the attached appendices. Specifically, Appendix H describes procedures used to determine research designs. Appendix I reports procedures to determine ratings of population validity in the research synthesis population of studies. Appendix J provides procedures to determine ratings of conclusion validity in the research synthesis population of studies. Using these procedures in other journals would make a determination of transferability of these research synthesis methods.

Recommendations for Practice

An inventory of three additional recommendations is advanced for improving practice is discussed below.

Professional Standards for Inquiry

The profession of speech-language pathology and audiology operates by both a Code of Ethics and a published set of statements about best practices for

research. As such, practitioners are responsible for not engaging in harmful practices and, equally important, researchers are responsible for ensuring that inquiries actually use best (established) research methods. However, if these procedures are not clearly defined and reported, consumers of research are unable to determine the validity and transferability of reported research findings. With these ideas in mind, the following recommendation has implications for researchers, consumers of research, and editors.

Recommendation Six. ***Articles should provide transparent, complete, and accurate reporting of all ethical and research procedures and results. From a best practice perspective, these procedures include defining the target population, specifying the actual sampling design, and describing the data analysis strategy that yields study findings.***

The *JSLHR* and ASHA have published guidelines for transparent reporting of procedures and results in articles. Authors should adhere to these guidelines, consumers of research should demand transparent reporting and easy-to-read articles, and editors and reviewers should be more rigorous in applying standards of clear research. In doing so, transference of findings can be made which will benefit both patients and practice.

Relevant Operationalizations

Authors frequently use nomenclatures in narrative. These terms, however, were often not fully defined or explained. Often, these terms of interest were abbreviated and then combined with other abbreviated terms or multiple abbreviated terms used in the same sentence. This practice creates confusion for consumers of research.

Recommendation Seven. ***Nomenclatures used in hearing, speech, and language research should be fully defined and explained. If abbreviated terms are combined or multiple abbreviated terms used in the same context, consideration should be given to providing clear operationalizations to clarify all theoretical concepts used in a research study.***

Fully defining nomenclatures is feasible because researchers should be able to clearly describe the variables of interest. In doing so, consumers of research will be able to determine if their definition is in agreement with the researcher's definition and clinical needs.

Statistical Conclusion Validity

Both the proposed statistical methods and their corresponding results were often not fully explained or provided in articles. Often the wrong statistical analysis was used with no clear reasoning or rationale was given for departure from established procedures for a specific research design. The most common error was the incorrect use of inferential statistical methods for nonprobability sampling designs, especially those designs that were best described in the theoretical framework as convenience or availability.

Recommendation Eight. ***Reasoning for the proposed statistical analysis should be fully explained with reference to the sample used in the study. This explanation should include how the assumptions associated with the proposed statistical analysis have been met.***

Reporting of proposed statistical analysis is possible because the statistical analysis should be developed as producers of research develop the study. By providing reference to the sample, consumers of research will be able to link sample to the target population. Disclosure of how the assumptions associated with the proposed statistical analysis have been met provides transparent reporting demanded by consumers of research.

Recommendation for Training

A final recommendation addresses a need in doctoral training programs based on the findings revealed in this study.

Training Needs

This study revealed that the majority of studies (88 of 105 studies or 83.8%) in this research synthesis population of 105 studies used an observational research design. The remaining 17 studies (16.2%) used an experimental research design.

Moreover, among the nine research design categories for observational studies (see Table 7.1), approximately two out of every three observational studies fall into just two of these nine categories. Specifically, these two observational research designs were cross-sectional descriptive and cross-sectional explanatory.

These synthesis trends about research designs used in the 105 studies in the research synthesis population give rise to a specific recommendation for training in research methods.

Recommendation Nine. ***Doctoral training programs should provide formal training for developing research skill to conduct observational studies.***

Both my personal experience and my familiarity with behavioral science research handbooks suggest that the major portion of doctoral level research methods training is usually dedicated to learning to conduct experimental studies. Far less time appears to be devoted to conducting observational studies. This recommendation is an effort to advance the idea that research training should emphasize training in both experimental and observational research methods.

Advanced training programs provide courses that address both quantitative research methods and qualitative research methods. It is important for both producers of research and consumers of research to fully understand observational studies. Specifically, cross-sectional designs with descriptive and explanatory intents should be emphasized.

Additionally, the methodologies used in evaluating trends in data revealed in research syntheses are unique to these types of studies. It is important for both producers of research and consumers of research to be able to accurately evaluate the information contained in the data.

CHAPTER IX

SUMMARY

This final chapter, as designated in the theoretical framework, is dedicated to summarizing the purpose, design, conclusions, and recommendations that have emerged from this methodological research synthesis. This chapter is organized into six parts. The first two parts review the purpose of the inquiry and the design of the inquiry, respectively. The third part provides a summary of the conclusions provided in each of the three phases dedicated to the substantive research topics of hearing, speech, and language and phase six which examines trends. The fourth part of this chapter reviews the set of eight recommendations elaborated in phase seven. The fifth part sets forth two limitations of this scholarly endeavor. The sixth and final part of this chapter provides some final thoughts.

Purpose of the Inquiry

The first part of this chapter revisits the purpose of this inquiry. As established in Chapter I, the purpose of this inquiry was to conduct a methodological research synthesis in order to determine both population validity and conclusion validity of empirical research studies reported in Volume 47 of the *Journal of Speech, Language and Hearing Research* (*JSLHR*).

The specific intent of this methodological research synthesis was to conduct a critical review of (a) the accuracy and completeness of the sampling

designs reported in the *JSLHR* articles, and (b) the extent to which the actual characteristics of these sampling designs were reflected in discussions of the rationale for selecting the data analysis procedures, the application of these procedures, and the narrative report of the information dealing with the findings, conclusions, implications for practice, and recommendations for future research.

Design of the Inquiry

The second part of this chapter revisits the design of the inquiry. As established in Chapter I, the intent of this methodological research synthesis was accomplished by implementing a research design.

This inquiry was developed in seven sequential phases. Specifically, phase one developed the theoretical framework that guided this inquiry in three parts. These three parts (a) developed a category system which classified empirical studies according to purpose and research design, (b) created a set of questions that provided an accurate description of sampling designs used in the empirical studies, and (c) framed a set of questions that assessed the impact of sampling design on conclusion validity.

Phase two described the research synthesis population of both articles and studies along three dimensions: (a) substantive concerns addressed in each study; (b) research design used; and (c) country of origin where each study was conducted.

Phases three, four, and five were dedicated to three parallel independent research syntheses for each of the substantive research topics of hearing, speech, and language. Additionally, phase six explored trends across the three substantive research domains.

Phase seven provided an inventory of eight recommendations. These eight recommendations were developed as an integral part of this research synthesis.

Conclusions

The third part of this chapter summarizes the conclusions of this scholarly endeavor. Conclusions have been presented in the three chapters associated with the three substantive topics of hearing, speech, and language and in the chapter that addressed trends across the three substantive topics. To summarize these sections four comments are offered. Two additional insights will conclude this section.

First, the overall assessment of population validity across all three substantive research topics reveals that only 16 studies (16 of 105 of 15.2%) received either a high or moderate rating of population validity. Moreover, 89 of the 105 studies in the overall research synthesis population (84.8%) had low population validity ratings.

Second, the overall assessment of conclusion validity across all three substantive research topics reveals that 26 of 105 studies or 24.7 percent were

found to have high or moderate conclusion validity ratings. Additionally, the large majority (79 of 105 studies or 75.3%) of the research synthesis population of studies received low conclusion validity ratings.

Third, when the corresponding population validity ratings and conclusion validity ratings were evaluated, 79 of the 105 studies in the research synthesis population were found to have low ratings in both areas.

Fourth, thinking in terms of a prediction model, population validity ratings provide an accurate prediction for 81 of the 105 studies (77.1%) in the total research synthesis population.

Two additional thoughts are offered as concluding remarks for this section.

First, population validity and conclusion validity are intricately intertwined. However, this relationship between the sample or participants in a study and conclusions drawn has historically not been addressed. Rather, as is often the case, parts (such as the sample or statistical procedures or conclusions) are defined and developed as separate entities. But rarely are these two parts (population and conclusions) looked at as a whole. To advance the concept of scientifically based research and evidence-based practice, it is imperative that this relationship between population validity and conclusion validity be acknowledged and evaluated.

Second, summary publication statistics for 2004 offer acceptance and rejection rates in the *Journal of Speech, Language, and Hearing Research*

(2005). These 2004 statistics indicate that 65 percent of hearing research articles submitted for publication were accepted, 27 percent of speech research articles submitted for publication were accepted, and 51 percent of language research articles submitted for publication were accepted. Given these acceptance rates and the low population validity and low conclusion validity ratings (81 of 105 studies or 77.1%) reported in Table 7.7, serious concerns exist in this author's mind about the quality of research conducted and subsequently reported in the *JSLHR*.

Recommendations

The fourth part of this chapter summarizes the inventory of eight recommendations dedicated to extending methodological research synthesis efforts initiated in this study and for clinical practice.

The first five recommendations provide recommendations for conducting future research synthesis studies. These five recommendations specifically address replication, evaluation of statistical evidence, analysis of trends, quantitative validity assessments, and transferability of research synthesis methods.

The last three recommendations offer considerations for practice. These recommendations address professional standards for inquiry, relevant operationalizations to clarify the theoretical concepts used in a research study, and statistical conclusion validity.

Limitations

The fifth part of this chapter addresses limitations of this scholarly endeavor. Two main limitations exist.

First, and most importantly, this study is an initial attempt to uncover reporting trends within the *JSLHR*. With that said, findings reported in this study apply **only** to this specific journal (*JSLHR*), to this specific volume (Volume 47), and to this specific year (2004).

Second, the three classification systems that were developed within this study are initial attempts to identify research designs, to establish population validity, and to establish conclusion validity. This initial attempt is heuristic and as a result a broad picture emerges rather than a well-defined picture.

For example, categories of experimental studies within the research design classification system are fairly well defined. However, descriptions of observational studies are more open to individual interpretation. Although inter-rater reliability was high within this study, the lack of well-defined criteria for descriptive studies, predictive studies, and explanatory studies makes these categories suspect.

Final Thoughts

Four final thoughts conclude this scholarly endeavor.

First, a research synthesis is time consuming and laborious. I have a keen appreciation of the effort required in producing a research synthesis. With that said, a synthesis is not a one-person show. Support for developing and implementing the coding system crucial to the success of the research synthesis process.

Second, numerous items related to inaccuracies and ambiguities in the articles were noted while reading the research synthesis population of articles. These interesting items are chronicled in Appendix N. In themselves, these inaccuracies and ambiguities provide supporting evidence about low population validity and conclusion validity ratings that were achieved.

Third, even though the effort required to complete a research synthesis can at times be described as torturous, the findings far outweigh effort. Research synthesis can provide insights into practice which may have gone unheeded or attempt to discover what is known. The completion of this research synthesis supports the ethos of honesty, openness, and reflection needed to advance scientific culture and sound clinical practice.

The fourth and final thought exceeds the boundaries of this methodological synthesis. Most of the required statistics courses in doctoral training programs center on experimental designs and parametric methods. However, the results of this study indicate that many observational studies are published in professional journals. As such, it is important that techniques for

synthesizing data and technical writing skills be included in advanced training programs.

REFERENCES

Agresti, A., & Finlay, B. (1999). *Statistical methods for the social sciences* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Alreck, P. L., & Settle, R. B. (1995). *The survey research handbook: Guidelines and strategies for conducting a survey* (2nd ed.). New York: McGraw-Hill.

American Speech-Language-Hearing Association. (2004). *Evidence-based practice in communication disorders: An introduction [Technical report].* Rockville, MD: Author. Available at http://www.asha.org/members/deskref-journals/deskref/default

American Speech-Language-Hearing Association. (2005). *Evidence-based practice in communication disorders [Position statement].* Rockville, MD: Author. Available at http://www.asha.org/members/deskrefjournals/deskref/default

American Speech-Language-Hearing Association. (2006a). *Highlights and trends: ASHA member counts.* Rockville, MD: Author. Available at http://www.asha.org/about/membership-certification/member-counts.htm

American Speech-Language-Hearing Association. (2006b). *Manuscript style and requirements.* Rockville, MD: Author. Available at http://www.asha.org /about/publications/journal-abstracts/submissions/prepare_ms.htm

Babbie, E. (1990). *Survey research methods* (2nd ed.). Belmont, CA: Wadsworth Publishing Company.

Best, J. W., & Kahn, J. V. (2003). *Research in education* (9th ed.). Boston: Pearson Education Inc.

Christensen, L. B. (1988). *Experimental methodology* (4th ed.). Boston, MA: Allyn and Bacon.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression / correlation analysis for the behavioral sciences* (3rd ed.). Hillsdale, NJ: Erlbaum.

Cooper, H. & Hedges, L. V. (Eds.) (1994). *The handbook of research synthesis.* New York: Russell Sage Foundation.

Cook, T. E., & Campbell, D. T. (1979). *Quasi-experimentation: Design & analysis issues for field settings.* Boston: Houghton Mifflin Company.

Cook, T. D., Cooper, H., Cordray, D. S., Hartmann, H., Hedges, L. V., Light, R. J., Louis, T. A., & Mosteller, F. (1992). *Meta-analysis for explanation: A casebook*. New York: Russell Sage Foundation.

Denzin, N. K., & Lincoln, Y. S. (Eds.). (2000). *Handbook of qualitative research* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Des Jarlais, D. C., Lyles, C., Crepaz, N., & the TREND Group. (2004). Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: The TREND statement. *American Journal of Public Health, 94*, 361-366.

Feldman, K. A. (1971). Using the work of others: Some observations on reviewing and integrating. *Sociology of Education, 44,* 86-102.

Feuer, M. J., Towne, L., & Shavelson, R. J. (2002). Scientific culture and educational research. *Educational Researcher, 31*, 4-14.

Gall, J. P., Gall, M. D., & Borg, W. R. (1999). *Applying educational research: A practical guide*. New York: Addison Wesley Longman, Inc.

Gall, J. P., Gall, M. D., & Borg, W. R. (2005). *Applying educational research: A practical guide* (5th ed.). Boston: Pearson Education, Inc.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher, 5*, 3-8.

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2003). *Applied statistics for the behavioral science* (5th ed.). Boston: Houghton Mifflin.

Holt, D. (1988). Missing data and nonresponse. In J. P. Keeves (Ed.). *Educational research, methodology, and measurement: An international handbook* (pp. 685-689). New York: Pergamon Press.

Hultsch, D. F., MacDonald, S. W. S., Hunter, M. A., Maitland, S. B., & Dixon, R. A. (2002). Sampling and generalisability in developmental research: Comparison of random and convenience samples of older adults. *International Journal of Behavioral Development, 26*, 345-359.

Johnson, B. (2001). Toward a new classification of nonexperimental quantitative research. *Educational Researcher, 30*, 3-13.

Justice, L. M., & Fey, M. E. (2004, Sept. 21). Evidence-based practice in schools: Integrating craft and theory with science and data. *The ASHA Leader*, pp. 4-5, 30-32.

Kier, F. J. (1999). Shortcomings of present practice in applied psychology research: A meta-analytic inquiry. Unpublished doctoral dissertation, Texas A&M University, College Station.

Kish, L. (1965). *Survey sampling.* New York: John Wiley & Sons.

Koehler, H. J. (2000). A comparative analysis of results based on using the vote count method and meta-analysis as two alternative strategies for integrating the findings of individual research studies that investigate gender differences in identity formation of college students. Unpublished doctoral dissertation, Texas A&M University, College Station.

Krathwohl, D. R. (1993). *Methods of educational and social science research: An integrated approach.* New York: Longman.

Light, R. J., & Pillemer, D. B. (1984). *Summing up: The science of reviewing research.* Cambridge, MA: Harvard University Press.

Lincoln, Y. S., & Guba, E. G. (2000). Paradigmatic controversies, contradictions, and emerging confluences. In N.K. Denzin & Y. S. Lincoln (Eds.). *Handbook of qualitative research* (2nd ed., pp. 163-188). Thousand Oaks, CA: Sage Publications.

Lipsey, M. W. (1994). Identifying potentially interesting variables and analysis opportunities. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis*. New York: Russell Sage Foundation

Lohr, S. L. (1999). *Sampling: Design and analysis*. Pacific Grove, CA: Duxbury Press.

McMillan, J. H., & Schumacher, S. (2001). *Research in education: A conceptual introduction* (5th ed.), New York: Addison Wesley Longman, Inc.

McNamara, J. F. (1994). *Survey and experiments in education research.* Lancaster, PA: Technomic Publishing Company.

McNamara, J. F. (1997). Meta-analysis: A strategy to improve integrative research reviews. *International Journal of Educational Reform, 6*, 490-495.

McNamara, J. F. (2003). Getting good results from survey research: Part one. *International Journal of Educational Reform, 12*, 275-288.

Meehl, P. E. (1997). Credentialed persons, credentialed knowledge. *Clinical Psychology: Science and Practice, 4*, 91-98.

Mertens, D. M. (2005). *Research and evaluation in educational psychology: Integrating diversity with quantitative, qualitative, and mixed methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Meyer, G. J. (2003). Guidelines for reporting information in studies of diagnostic test accuracy: The STARD initiative. *Journal of Personality Assessment, 81*, 191-193.

Moher, D., Schulz, K. F., & Altman, D. (2001). The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomized trials. *Journal of the American Medical Association, 285*, 1987-1991.

Moore, G. W. (1983). *Developing and evaluating educational research*. Boston: Little, Brown and Company.

Moore, J. D. (2001). Moving beyond point estimates in counseling psychology research: Implications for future research and meta-analysis inquiries. Unpublished doctoral dissertation, Texas A&M University, College Station.

No Child Left Behind Act of 2001, Pub. L. No. 107-110.

Ottenbacher, K. J. (1989). Statistical conclusion validity of early intervention research with handicapped children. *Exceptional Children, 55*, 534-540.

Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2nd ed.). Newbury Park, CA: Sage.

Patton, M. L. (2001). *Questionnaire research: A practical guide* (2nd ed.). Los Angeles, CA: Pyrczak Publishing.

Pressman, J. D. (1998). *Last resort: Psychosurgery and the limits of medicine*. Cambridge: Cambridge University Press.

Rosier, M. J. (1988). Survey research methods. In J. P. Keeves (Ed.). *Educational research, methodology, and measurement: An international handbook* (pp. 107-113). New York: Pergamon Press.

Ross, K. N. (1988). Sampling. In J. P. Keeves (Ed.). *Educational research, methodology, and measurement: An international handbook* (pp. 527-537). New York: Pergamon Press.

Rubin, A., & Babbie, E. (2001). *Research methods for social work* (4th ed.). Belmont, CA: Wadsworth/Thomson Learning.

Sackett, D. L., Haynes, R. B., Guyatt, G. H., & Tugwell, P. (1991). *Clinical epidemiology: A basic science for clinical medicine* (2nd ed.). Boston: Little, Brown.

Sackett, D. L., Rosenberg, W. M. C., Gray, J. A. M., Hanes, R. B., & Richardson, W. S. (1996). Evidence-based medicine: What it is and what it isn't. *British Medical Journal, 312,* 71-72.

Scheaffer, R. L., Mendenhall, W., & Ott, L. (1996). *Elementary survey sampling* (6th ed.). Pacific Grove, CA: Duxbury Press.

Sheskin, D. J. (2004). *Handbook of parametric and nonparametric statistical procedures* (3rd ed.). Boca Raton, FL: CRC Press.

STARD Group. (2006). *The STARD initiative—towards complete and accurate reporting of studies on diagnostic accuracy.* Amsterdam, The Netherlands: Author. Available at http://www.consortstatement.org/ stardstatement.htm

Summary Publication Statistics for 2004 (2005). *Journal of Speech, Language, and Hearing Research*, *48*, 1253.

Tate, R. (1988). Experimental studies. In J.P. Keeves (Ed.). *Educational research, methodology, and measurement: An international handbook* (pp. 93-101). New York: Pergamon Press.

Thompson, D. P. (1993). Job satisfaction: A synthesis of research in the *Educational Administration Quarterly*. Unpublished doctoral dissertation, Texas A&M University, College Station.

Van Riper, C. (1965). *Speech correction: Principles and methods* (4th ed.). Englewood Cliffs, NJ: Prentice-Hall, Inc.

Vogt, W. P. (1999). *Dictionary of statistics & methodology: A nontechnical guide for the social sciences* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Vogt, W. P. (2007). *Quantitative research methods for professionals*. Boston: Pearson Education, Inc.

Walker, D. A. & Burnhill, P. M. (1988). Survey studies, cross-sectional. In J. P. Keeves, (Ed.). *Educational research, methodology, and measurement: An international handbook* (pp. 101-107). New York: Pergamon Press.

Wang, L. (1996). A typology and evaluation of the survey sample designs in the *Educational Administration Quarterly:* 1980-1995. Unpublished doctoral dissertation, Texas A&M University, College Station.

APPENDIX A

A BIBLIOGRAPHY OF THE ARTICLES IN VOLUME 47 OF THE *JOURNAL OF SPEECH, LANGUAGE, AND HEARING RESEARCH*

Article	Citation
1	Grose, J. H., Hall, J. W., & Buss, E. (2004). Duration discrimination in listeners with cochlear hearing loss: Effects of stimulus type and frequency. *Journal of Speech, Language, and Hearing Research, 47*, 5-12.
2	Hall, J. W., Buss, E., Grose, J. H., & Dev, M. B. (2004). Developmental effects in the masking-level difference. *Journal of Speech, Language, and Hearing Research, 47*, 13-20.
3	Bender, B. K., Cannito, M. P., Murry, T., & Woodson, G. E. (2004). Speech intelligibility in severe adductor spasmodic dysphonia. *Journal of Speech, Language, and Hearing Research, 47*, 21-32.
4	Daniels, S. K., Corey, D. M., Hadskey, L. D., Legendre, C., Priestly, D. H., Rosenbek, J. C., & Foundas, A. L. (2004). Mechanism of sequential swallowing during straw drinking in healthy young and older adults. *Journal of Speech, Language, and Hearing Research, 47*, 33-45.
5	Guenther, F. H., Nieto-Castanon, A., Ghosh, S. S., & Tourville, J. A. (2004). Representation of sound categories in auditory cortical maps. *Journal of Speech, Language, and Hearing Research, 47*, 46-57.
6	Munson, B. (2004). Variability in /s/ production in children and adults: Evidence from dynamic measures of spectral mean. *Journal of Speech, Language, and Hearing Research, 47*, 58-69.
7	Connaghan, K. P., Moore, C. A., & Higashakawa, M. (2004). Respiratory kinematics during vocalization and nonspeech respiration in children from 9 to 48 months. *Journal of Speech, Language, and Hearing Research, 47*, 70-84.
8	Tasko, S. M., & McClean, M. D. (2004). Variations in articulatory movement with changes in speech task. *Journal of Speech, Language, and Hearing Research, 47*, 85-100.
9	Howell, P. (2004). Cerebellar activity and stuttering: Comments on Max and Yudman (2003). *Journal of Speech, Language, and Hearing Research, 47*, 101-104.
10	Max, L., & Yudman, E. M. (2004). Understanding stuttering will require theoretical models that fit the data rather than attempts to make the data fit the preferred models: A response to Howell (2004). *Journal of Speech, Language, and Hearing Research, 47*, 105-113.

Article	Citation

11 Hoffman, L. M., & Gillam R. B. (2004). Verbal and spatial information processing constraints in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 47*, 114-125.

12 Thiemann, K. S., & Goldstein, H. (2004). Effects of peer training and written text cueing on social communication of school-age children with pervasive developmental disorder. *Journal of Speech, Language, and Hearing Research, 47*, 126-144.

13 Conti-Ramsden, G., & Botting, N. (2004). Social difficulties and victimization in children with SLI at 11 years of age. *Journal of Speech, Language, and Hearing Research, 47*, 145-161.

14 Hugdahl, K., Gundersen, H., Brekke, C., Thomsen, T., Rimol, L. M., Ersland, L., & Niemi, J. (2004). fMRI brain activation in a Finnish family with specific language impairment compared with a normal control group. *Journal of Speech, Language, and Hearing Research, 47*, 162-172.

15 Capone, N. D., & McGregor, K. K. (2004). Gesture development: A review for clinical and research practices. *Journal of Speech, Language, and Hearing Research, 47*, 173-186.

16 Maillart, C., Schelstraete, M-A., & Hupet, M. (2004). Phonological representations in children with SLI: A study of French. *Journal of Speech, Language, and Hearing Research, 47*, 187-198.

17 Nation, K., Clarke, P., Marshall C. M., & Durand, M. (2004). Hidden language impairments in children: Parallels between poor reading comprehension and specific language impairment? *Journal of Speech, Language, and Hearing Research, 47*, 199-211.

18 Coles-White, D. (2004). Negative concord in child African American English: Implications for specific language impairment. *Journal of Speech, Language, and Hearing Research, 47*, 212-222.

19 Roy, P., & Chiat, S. (2004). A prosodically controlled word and nonword repetition task for 2- to 4-year-olds: Evidence from typically developing children. *Journal of Speech, Language, and Hearing Research, 47*, 223-234.

20 Summers, V., & Molis, M. R. (2004). Speech recognition in fluctuating and continuous maskers: Effects of hearing loss and presentation level. *Journal of Speech, Language, and Hearing Research, 47*, 245-256.

Article	Citation

21 Lister, J., & Tarver, K. (2004). Effect of age on silent gap discrimination in synthetic speech stimuli. *Journal of Speech, Language, and Hearing Research, 47*, 257-268.

22 Rothpletz, A. M., Tharpe, A. M., & Grantham, D. W. (2004). The effect of asymmetrical signal degradation on binaural speech recognition in children and adults. *Journal of Speech, Language, and Hearing Research, 47*, 269-280.

23 Roy, N., Merrill, R. M., Thibeault, S., Parsa, R. A., Gray, S. D., & Smith, E. M. (2004). Prevalence of voice disorders in teachers and the general population. *Journal of Speech, Language, and Hearing Research, 47*, 281-293.

24 McClean, M. D., Tasko, S. M., & Runyan, C. M. (2004). Orofacial movements associated with fluent speech in persons who stutter. *Journal of Speech, Language, and Hearing Research, 47*, 294-303.

25 Massaro, D. W., & Light, J. (2004). Using visible speech to train perception and production of speech for individuals with hearing loss. *Journal of Speech, Language, and Hearing Research, 47*, 304-320.

26 Ingham, R. J., Fox, P. T., Ingham, J. C., Xiong, J., Zamarripa, F., Hardies, L. J., & Lancaster, J. L. (2004). Brain correlates of stuttering and syllable production: Gender comparison and replication. *Journal of Speech, Language, and Hearing Research, 47*, 321-341.

27 Steele, C. M., & Van Lieshout, P. H. H. M. (2004). Use of electromagnetic midsagittal articulography in the study of swallowing. *Journal of Speech, Language, and Hearing Research, 47*, 342-352.

28 Dworkin, J. (2004). Is voice amplification for teachers with dysphonia really beneficial? *Journal of Speech, Language, and Hearing Research, 47*, 353-357.

29 Roy, N. (2004). Replication, randomization, and clinical relevance: A response to Dworkin and colleagues (2004). *Journal of Speech, Language, and Hearing Research, 47*, 358-365.

30 Southwood, F., & Russell, A. F. (2004). Comparison of conversation, freeplay, and story generation as methods of language sample elicitation. *Journal of Speech, Language, and Hearing Research, 47*, 366-376.

Article	Citation

31 Nathan, L., Stackhouse, J., Goulandris, N., & Snowling, M. J. (2004). The development of early literacy skills among children with speech difficulties: A test of the "critical age hypothesis". *Journal of Speech, Language, and Hearing Research, 47*, 377-391.

32 Jarvis, L. H., Merriman, W. E., Barnett, M., Hanba, J., & Van Haitsma, K. S. (2004). Input that contradicts young children's strategy for mapping novel words affects their phonological and semantic interpretation of other novel words. *Journal of Speech, Language, and Hearing Research, 47*, 392-406.

33 Alt, M., Plante, E., & Creusere, M. (2004). Semantic features in fast-mapping: Performance of preschoolers with specific language impairment versus preschoolers with normal language. *Journal of Speech, Language, and Hearing Research, 47*, 407-420.

34 Edwards, J., Beckman, M. E., & Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language, and Hearing Research, 47*, 421-436.

35 Iacono, T., & Cupples, L. (2004). Assessment of phonemic awareness and word reading skills of people with complex communication needs. *Journal of Speech, Language, and Hearing Research, 47*, 437-449.

36 Craig, H. K., & Washington, J. A. (2004). Grade-related changes in the production of African American English. *Journal of Speech, Language, and Hearing Research, 47*, 450-463.

37 Dollaghan, C. A. (2004). Taxometric analysis of specific language impairment in 3- and 4-year-old children. *Journal of Speech, Language, and Hearing Research, 47*, 464-475.

38 Ertmer, D. J. (2004). How well can children recognize speech features in spectrograms? Comparisons by age and hearing status. *Journal of Speech, Language, and Hearing Research, 47*, 484-495.

39 Collison, E. A., Munson, B., & Carney, A. E. (2004). Relations among linguistic and cognitive skills and spoken word recognition in adults with cochlear implants. *Journal of Speech, Language, and Hearing Research, 47*, 496-508.

40 Connor, C. M., & Zwolan, T. A. (2004). Examining multiple sources of influence on the reading comprehension skills of children who use cochlear implants. *Journal of Speech, Language, and Hearing Research, 47*, 509-526.

Article	Citation

41 McArthur, G. M., & Bishop, D. V. M. (2004). Frequency discrimination deficits in people with specific language impairment: Reliability, validity, and linguistic correlates. *Journal of Speech, Language, and Hearing Research, 47*, 527-541.

42 Roy, N., Merrill, R. M., Thibeault, S., Gray, S. D., & Smith, E. M. (2004). Voice disorders in teachers and the general population: Effects on work performance, attendance, and future career choices. *Journal of Speech, Language, and Hearing Research, 47*, 542-551.

43 Anderson, J. D., & Conture, E. G. (2004). Sentence-structure priming in young children who do and do not stutter. *Journal of Speech, Language, and Hearing Research, 47*, 552-571.

44 Hiss, S. G., Strauss, M., Treole, K., Stuart, A., & Boutilier, S. (2004). Effects of age, gender, bolus volume, bolus viscosity, and gestation on swallowing apnea onset relative to lingual bolus propulsion onset in normal adults. *Journal of Speech, Language, and Hearing Research, 47*, 572-583.

45 Solomon, N. P., & Munson, B. (2004). The effect of jaw position on measures of tongue strength and endurance. *Journal of Speech, Language, and Hearing Research, 47*, 584-594.

46 Qi, C. H., & Kaiser, A. P. (2004). Problem behaviors of low-income children with language delays: An observation study. *Journal of Speech, Language, and Hearing Research, 47*, 595-609.

47 Silkes, J. P., McNeil, M. R., & Drton, M. (2004). Simulation of aphasic naming performance in non-brain-damaged adults. *Journal of Speech, Language, and Hearing Research, 47*, 610-623.

48 German, D. J., & Newman, R. S. (2004). The impact of lexical factors on children's word-finding errors. *Journal of Speech, Language, and Hearing Research, 47*, 624-636.

49 Fujiki, M., Spackman, M. P., Binton, B., & Hall, A. (2004). The relationship of language and emotion regulation skills to reticence in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 47*, 637-646.

50 Hart, K. I., Fujiki, M., Brinton, B., & Hart, C. H. (2004). The relationship between social behavior and severity of language impairment. *Journal of Speech, Language, and Hearing Research, 47*, 647-662.

Article	Citation

51 Brady, N. C., Marquis, J., Fleming, K., & McLean, L. (2004). Prelinguistic predictors of language growth in children with developmental disabilities. *Journal of Speech, Language, and Hearing Research, 47*, 663-677.

52 Crais, E., Douglas, D. D., & Campbell, C. C. (2004). The intersection of the development of gestures and intentionality. *Journal of Speech, Language, and Hearing Research, 47*, 678-694.

53 Meilijson, S. R., Kaher, A., & Elizur, A. (2004). Language performance in chronic schizophrenia: A pragmatic approach. *Journal of Speech, Language, and Hearing Research, 47*, 695-713.

54 Pankratz, M., Morrison, A., & Plante, E. (2004). Difference in standard scores of adults on the Peabody Picture Vocabulary Test (Revised and Third Edition). *Journal of Speech, Language, and Hearing Research, 47*, 714-718.

55 Markham, D., & Hazan, V. (2004). The effect of talker-and listener-related factors on intelligibility for a real-word, open-set perception test. *Journal of Speech, Language, and Hearing Research, 47*, 725-737.

56 Paatsch, L. E., Blamey, P. J., Sarant, J. Z., Martin, L. F. A., & Bow, C. P. (2004). Separating contributions of hearing, lexical knowledge, and speech production to speech-perception scores in children with hearing impairments. *Journal of Speech, Language, and Hearing Research, 47*, 738-750.

57 Heath, S. M., & Hogben, J. H. (2004). Cost-effective prediction of reading difficulties. *Journal of Speech, Language, and Hearing Research, 47*, 751-765.

58 Tjaden, K., & Wilding, G. E. (2004). Rate and loudness manipulations in dysarthria: Acoustic and perceptual findings. *Journal of Speech, Language, and Hearing Research, 47*, 766-783.

59 Zajac, D. J., & Weissler, M. C. (2004) Air pressure responses to sudden vocal tract pressure bleeds during production of stop consonants: New evidence of aeromechanical regulation. *Journal of Speech, Language, and Hearing Research, 47*, 784-801.

60 Deevy, P., & Leonard, L. B. (2004). The comprehension of wh- questions in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 47*, 802-815.

Article	Citation

61 Rice, M. L., Tomblin, J. B., Hoffman, L., Richman, W. A., & Marquis, J. (2004). Grammatical tense deficits in children with SLI and nonspecific language impairment: Relationships with nonverbal IQ over time. *Journal of Speech, Language, and Hearing Research, 47*, 816-834.

62 Kavé, G., & Levy, Y. (2004). Preserved morphological decomposition in persons with Alzheimer's Disease. *Journal of Speech, Language, and Hearing Research, 47*, 835-847.

63 Schlosser, R. W., & Blischak, D. M. (2004). Effects of speech and print feedback on spelling by children with autism. *Journal of Speech, Language, and Hearing Research, 47*, 848-862.

64 Gutiérrez-Clellen, V. F., Calderón, J., & Weismer, S. E. (2004). Verbal working memory in bilingual children. *Journal of Speech, Language, and Hearing Research, 47*, 863-876.

65 Windsor, J., & Kohnert, K. (2004). The search for common ground: Part I. Lexical performance by linguistically diverse learners. *Journal of Speech, Language, and Hearing Research, 47*, 877-890.

66 Kohnert, K., & Windsor, J. (2004). The search for common ground: Part II. Nonlinguistic performance by linguistically diverse learners. *Journal of Speech, Language, and Hearing Research, 47*, 891-903.

67 Van Hulle, C. A., Goldsmith, H. H., & Lemery, K. S. (2004). Genetic, environmental, and gender effects on individual differences in toddler expressive language. *Journal of Speech, Language, and Hearing Research, 47*, 904-912.

68 Pine, J. M., Joseph, K. L., & Conti-Ramsden, G. (2004). Do data from children with specific language impairment support the agreement/tense omission model? *Journal of Speech, Language, and Hearing Research, 47*, 913-923.

69 Law, J., Garrett, & Nye, C. (2004). The efficacy of treatment for children with developmental speech and language delay/disorder: A meta-analysis. *Journal of Speech, Language, and Hearing Research, 47*, 924-943.

70 Beverly, B. L., & Williams, C. C. (2004). Present tense be use in young children with specific language impairment: Less is more. *Journal of Speech, Language, and Hearing Research, 47*, 944-956.

Article	Citation

71 Roberts, R. A., & Lister, J. (2004). Effects of age and hearing loss on gap detection and the precedence effect: Broadband stimuli. *Journal of Speech, Language, and Hearing Research, 47*, 965-978.

72 Miller, P. (2004). Processing of written words by individuals with prelingual deafness. *Journal of Speech, Language, and Hearing Research, 47*, 979-989.

73 Miller, P. (2004). Processing of written word and nonword visual information by individuals with prelingual deafness. *Journal of Speech, Language, and Hearing Research, 47*, 990-1000.

74 Nabelek, A. K., Tampas, J. W., & Burchfield, S. B. (2004). Comparison of speech perception in background noise with acceptance of background noise in aided and unaided conditions. *Journal of Speech, Language, and Hearing Research, 47*, 1001-1011.

75 Hedrick, M. S., & Nabelek, A. K. (2004). Effect of F2 intensity on identity of /u/ in degraded listening conditions. *Journal of Speech, Language, and Hearing Research, 47*, 1012-1021.

76 Hill, P. R., Hartley, D. E. H., Glasberg, B. R., Moore, B. C. J., & Moore, D. R. (2004). Auditory processing efficiency and temporal resolution in children and adults. *Journal of Speech, Language, and Hearing Research, 47*, 1022-1029.

77 Blomert, L., Mitterer, H., & Paffen, C. (2004). In search of the auditory, phonetic, and/or phonological problems in dyslexia: Context effects in speech perception. *Journal of Speech, Language, and Hearing Research, 47*, 1030-1047.

78 Munson, B., & Solomon, N. P. (2004). The effect of phonological neighborhood density on vowel articulation. *Journal of Speech, Language, and Hearing Research, 47*, 1048-1058.

79 Ménard, L., Schwartz, J. L., & Boë, L. J. (2004). Role of vocal tract morphology in speech development: Perceptual targets and sensorimotor maps for synthesized French vowels from birth to adulthood. *Journal of Speech, Language, and Hearing Research, 47*, 1059-1080.

Article	Citation

80 O'Brian, S., Packman, A., Onslow, M., & O'Brian, N. (2004). Measurement of stuttering in adults: Comparison of stuttering-rate and severity-scaling methods. *Journal of Speech, Language, and Hearing Research, 47*, 1081-1087.

81 Goffman, L. (2004). Kinematic differentiation of prosodic categories in normal and disordered language development. *Journal of Speech, Language, and Hearing Research, 47*, 1088-1102.

82 Dillon, C. M., Burkholder, R. A., Cleary, M., & Pisoni, D. B. (2004). Nonword repetition by children with cochlear implants: Accuracy ratings from normal-hearing listeners. *Journal of Speech, Language, and Hearing Research, 47*, 1103-1116.

83 Gray, S. (2004). Word learning by preschoolers with specific language impairment: Predictors and poor learners. *Journal of Speech, Language, and Hearing Research, 47*, 1117-1132.

84 Drager, K. D. R., Light, J. C., Carlson, R., D'Silva, K., Larsson, B., Pitkin, L., & Stopper, G. (2004). Learning of dynamic display AAC technologies by typically developing 3-year-olds: Effect of different layouts and menu approaches. *Journal of Speech, Language, and Hearing Research, 47*, 1133-1148.

85 Chapman, S. B., Weiner, M. F., Rackley, A., Hynan, L. S., & Zientz, J. (2004). Effects of cognitive-communication stimulation for Alzheimer's Disease patients treated with Donepezil. *Journal of Speech, Language, and Hearing Research, 47*, 1149-1163.

86 Wong, A. M.-Y., Au, C. W.-S., & Stokes, S. F. (2004). Three measures of language production for Cantonese-speaking school-age children in a story-retelling task. *Journal of Speech, Language, and Hearing Research, 47*, 1164-1178.

87 Norbury, C. F. (2004). Factors supporting idiom comprehension in children with communication disorders. *Journal of Speech, Language, and Hearing Research, 47*, 1179-1193.

88 Storkel, H. L. (2004). The emerging lexicon of children with phonological delays: Phonotactic constraints and probability in acquisition. *Journal of Speech, Language, and Hearing Research, 47*, 1194-1212.

Article	Citation

89 Patterson, J. L. (2004). Comparing bilingual and monolingual toddlers' expressive vocabulary size: Revisiting Rescorla and Achenbach (2002). *Journal of Speech, Language, and Hearing Research, 47*, 1213-1215.

90 Rescorla, L. (2004). Comparing expressive vocabulary size in children from bilingual and monolingual home environments: A reply to Patterson's (2004) commentary. *Journal of Speech, Language, and Hearing Research, 47*, 1216-1217.

91 Editor. (2004). Summary publication statistics for 2003. *Journal of Speech, Language, and Hearing Research, 47*, 1225-1226.

92 Peng, S.-C., Spencer, L. J., & Tomblin, J. B. (2004). Speech intelligibility of pediatric cochlear implant recipients with 7 years of device experience. *Journal of Speech, Language, and Hearing Research, 47*, 1227-1236.

93 McAnally, K. I., Castles, A., & Bannister, S. (2004). Auditory temporal pattern discrimination and reading ability. *Journal of Speech, Language, and Hearing Research, 47*, 1237-1243.

94 Weber-Fox, C., Spencer, R. M. C., Spruill, J. E., & Smith, A. (2004). Phonologic processing in adults who stutter: Electrophysiological and behavioral evidence. *Journal of Speech, Language, and Hearing Research, 47*, 1244-1258.

95 Perkell, J. S., Matthies, M. L., Tiede, M., Lane, H., Zandipour, M., Marrone, N., Stockmann, E., & Guenther, F. H. (2004). The distinctness of speakers' /s/ - /ʃ/ contrast is related to their auditory discrimination and use of an articulatory saturation effect. *Journal of Speech, Language, and Hearing Research, 47*, 1259-1269.

96 Hoag, L. A., Bedrosian, J. L., McCoy, K. F., & Johnson, D. E. (2004). Trade-offs between informativeness and speed of message delivery in augmentative and alternative communication. *Journal of Speech, Language, and Hearing Research, 47*, 1270-1285.

97 Bird, E. K. –R., Chapman, R. S., & Schwartz, S. E. (2004). Fast mapping of words and story recall by individuals with Down Syndrome. *Journal of Speech, Language, and Hearing Research, 47*, 1286-1300.

Article	Citation

98 Fey, M. E., Catts, H. W., Proctor-Williams, K., Tomblin, J. B., & Zhang, X. (2004). Oral and written story composition skills of children with language impairment. *Journal of Speech, Language, and Hearing Research, 47*, 1301-1318.

99 Maslen, R. J. C., Theakston, A. L., Lieven, E. V. M., & Tomasello, M. (2004). A dense corpus study of past tense and plural overregularization in English. *Journal of Speech, Language, and Hearing Research, 47*, 1319-1333.

100 Brock, J., & Jarrold, C. (2004). Language influences on verbal short-term memory performance in Down Syndrome: Item and order recognition. *Journal of Speech, Language, and Hearing Research, 47*, 1334-1346.

101 Hayiou-Thomas, M. E., Bishop, D. V. M., & Plunkett, K. (2004). Simulating SLI: General cognitive processing stressors can produce a specific linguistic profile. *Journal of Speech, Language, and Hearing Research, 47*, 1347-1362.

102 Leonard, L. B., Camarata, S. M., Brown, B., & Camarata, M. N. (2004). Tense and agreement in the speech of children with specific language impairment: Patterns of generalization through intervention. *Journal of Speech, Language, and Hearing Research, 47*, 1363-1379.

103 Tompkins, C. A., Fassbinder, W., Blake, M. L., Baumgaertner, A., & Jayaram, N. (2004). Inference generation during text comprehension by adults with right hemisphere brain damage: Activation failure versus multiple activation. *Journal of Speech, Language, and Hearing Research, 47*, 1380-1395.

104 Klee, T., Stokes, S. F., Wong, A. M.-Y., Fletcher, P., & Gavin, W. J. (2004). Utterance length and lexical diversity in Cantonese-speaking children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research, 47*, 1396-1410.

105 van Daal, J., Verhoeven, L., & van Balkom, H. (2004). Subtypes of severe speech and language impairments: Psychometric evidence from 4-year-old children in the Netherlands. *Journal of Speech, Language, and Hearing Research, 47*, 1411-1423.

106 Fallon, K. A., Light, J., McNaughton, D., Drager, K., & Hammer, C. (2004). The effects of direct instruction on the single-word reading skills of children who require augmentative and alternative communication. *Journal of Speech, Language, and Hearing Research, 47*, 1424-1439.

Article	Citation

107 Wong, A. M.-Y., Leonard, L. B., Fletcher, P., & Stokes, S. F. (2004). Questions without movement: A study of Cantonese-speaking children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research, 47*, 1440-1453.

108 Storkel, H. L. (2004). Methods for minimizing the confounding effects of word length in the analysis of phonotactic probability and neighborhood density. *Journal of Speech, Language, and Hearing Research, 47*, 1454-1468.

109 Mackie, C., & Dockrell, J. E. (2004). The nature of written language deficits in children with SLI. *Journal of Speech, Language, and Hearing Research, 47*, 1469-1483.

APPENDIX B

TRANSFORMATION RECODE FROM RESEARCH SYNTHESIS POPULATION OF ARTICLES TO RESEARCH SYNTHESIS POPULATION OF STUDIES

This appendix provides information about the transition between the research synthesis population of articles and the research synthesis population of studies. This transition is explained in the four columns of Box B.1.

The first column provides the article log number for all articles referenced in the article citations located in Appendix A.

To respect articles with multiple studies, the second column provides the numbers for each study within the corresponding article. This column, then, constitutes the research synthesis study population as discussed in Chapter III.

The third column provides information about the substantive concerns of each article.

As explained in Chapter III, not all articles met the criteria for inclusion in the research synthesis population. As appropriate, column four provides the reason for exclusion from the research synthesis population.

Result of this transformation recode establishes 92 articles in the research synthesis population of articles with 17 articles eliminated from research synthesis population. Additionally, 105 studies were defined in the research synthesis population of studies.

Box B.1: Transformation Recode

Article Log Number	Study Number	Substantive Concern	Reason for Exclusion
1	1	Hearing	
2	2	Hearing	
3	3	Speech	
4	4	Speech	
5	83	Speech	
	84	Speech	
	85	Speech	
6	5	Speech	
7	6	Speech	
8	7	Speech	
9			Letter to the Editor
10			Response to Letter to the Editor
11	8	Language	
12	9	Language	
13	10	Language	
14	11	Language	
15			Literature Review
16	12	Language	
17	13	Language	
18	14	Language	
19	15	Language	
20	86	Hearing	
	87		
21	16	Hearing	
22	17	Hearing	
23	18	Speech	
24	19	Speech	
25	20	Speech	
26	21	Speech	
27			Research note
28			Letter to the Editor
29			Response to Letter to the Editor
30	22	Language	
31	23	Language	
32	88	Language	
	89		
	90		
33	24	Language	
34	25	Language	
35	26	Language	
36	27	Language	
37	91	Language	
	92		
38	28	Hearing	
39	29	Hearing	
40	30	Hearing	
41	31	Hearing	
42	32	Speech	
43	33	Speech	
44	34	Speech	
45			Research note
46	35	Language	
47	36	Language	

Box B.1 Transformation Recode (Continued)

Article Log Number	Study Number	Substantive Concern	Reason for Exclusion
48	37	Language	
49	38	Language	
50	39	Language	
51	40	Language	
52	41	Language	
53	42	Language	
54			Research note
55	43	Hearing	
56	44	Hearing	
57	45	Hearing	
58	46	Speech	
59	93	Speech	
	94		
60	47	Language	
61	95	Language	
	96		
62	97	Language	
	98		
63	48	Language	
64	49	Language	
65	50	Language	
66	51	Language	
67	52	Language	
68	53	Language	
69			Meta-analysis
70			Research note
71	54	Hearing	
72	55	Hearing	
73	56	Hearing	
74	57	Hearing	
75	58	Hearing	
76	59	Hearing	
77	99	Hearing	
	100		
	101		
78	102	Speech	
	103		
79	60	Speech	
80			Research note
81	61	Language	
82	62	Language	
83	63	Language	
84	64	Language	
85	65	Language	
86	66	Language	
87	67	Language	
88	68	Language	
89			Letter to the Editor
90			Response to Letter to the Editor
91			Summary statistics
92	69	Hearing	
93			Research note
94	70	Speech	

Box B.1 Transformation Recode (Continued)

Article Log Number	Study Number	Substantive Concern	Reason for Exclusion
95	71	Speech	
96	72	Language	
97	73	Language	
98	74	Language	
99	75	Language	
100	76	Language	
101	77	Language	
102	78	Language	
103	79	Language	
104	104	Language	
	105		
105	80	Language	
106	81	Language	
107	82	Language	
108			Research note
109			Research note
n= 92 articles in the research synthesis population of articles	n= 105 studies in the research synthesis population of studies		n= 17 articles eliminated from research synthesis population

APPENDIX C

CONSOLIDATED STANDARDS OF REPORTING TRIALS

(CONSORT)

Box C.1: CONSORT

PAPER SECTION and Topic	Item	Description	Reported on Page #
TITLE & ABSTRACT	1	How participants were allocated to interventions (*e.g.*, "random allocation", "randomized", or "randomly assigned").	
INTRODUCTION Background	2	Scientific background and explanation of rationale.	
METHODS Participants	3	Eligibility criteria for participants and the settings and locations where the data were collected.	
Interventions	4	Precise details of the interventions intended for each group and how and when they were actually administered.	
Objectives	5	Specific objectives and hypotheses.	
Outcomes	6	Clearly defined primary and secondary outcome measures and, when applicable, any methods used to enhance the quality of measurements (*e.g.*, multiple observations, training of assessors).	
Sample size	7	How sample size was determined and, when applicable, explanation of any interim analyses and stopping rules.	
Randomization -- Sequence generation	8	Method used to generate the random allocation sequence, including details of any restrictions (*e.g.*, blocking, stratification)	
Randomization -- Allocation concealment	9	Method used to implement the random allocation sequence (*e.g.*, numbered containers or central telephone), clarifying whether the sequence was concealed until interventions were assigned.	
Randomization -- Implementation	10	Who generated the allocation sequence, who enrolled participants, and who assigned participants to their groups.	
Blinding (masking)	11	Whether or not participants, those administering the interventions, and those assessing the outcomes were blinded to group assignment. When relevant, how the success of blinding was evaluated.	
Statistical methods	12	Statistical methods used to compare groups for primary outcome(s); Methods for additional analyses, such as subgroup analyses and adjusted analyses.	

Box C.1: CONSORT (Continued)

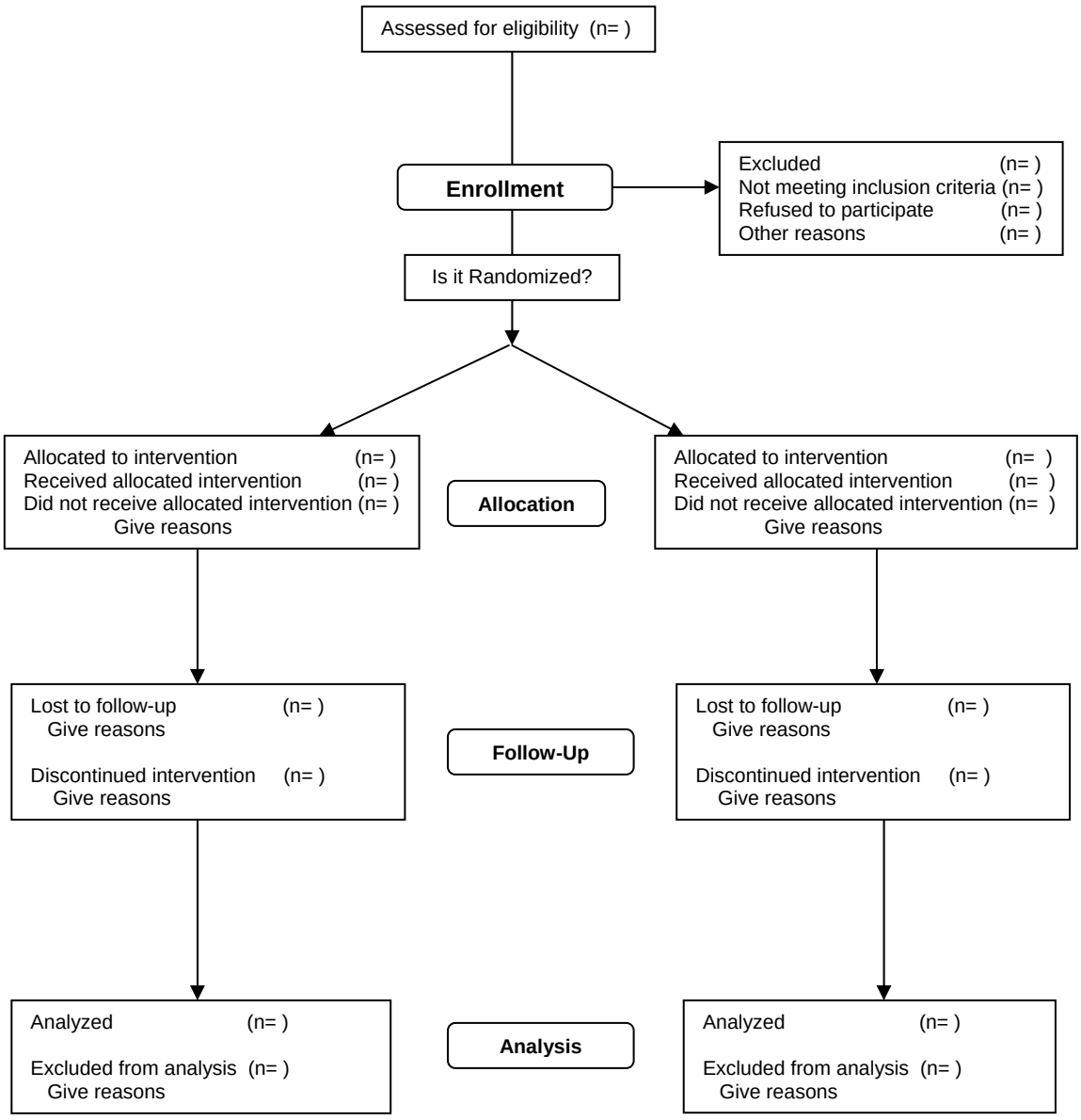
PAPER SECTION and Topic	Item	Description	Reported on Page #
RESULTS			
Participant flow	13	Flow of participants through each stage (a diagram is strongly recommended). Specifically, for each group report the numbers of participants randomly assigned, receiving intended treatment, completing the study protocol, and analyzed for the primary outcome. Describe protocol deviations from study as planned, together with reasons.	
Recruitment	14	Dates defining the periods of recruitment and follow-up.	
Baseline data	15	Baseline demographic and clinical characteristics of each group.	
Numbers analyzed	16	Number of participants (denominator) in each group included in each analysis and whether the analysis was by "intention-to-treat". State the results in absolute numbers when feasible (*e.g.*, 10/20, not 50%).	
Outcomes and estimation	17	For each primary and secondary outcome, a summary of results for each group, and the estimated effect size and its precision (*e.g.*, 95% confidence interval).	
Ancillary analyses	18	Address multiplicity by reporting any other analyses performed, including subgroup analyses and adjusted analyses, indicating those pre-specified and those exploratory.	
Adverse events	19	All important adverse events or side effects in each intervention group.	
DISCUSSION Interpretation	20	Interpretation of the results, taking into account study hypotheses, sources of potential bias or imprecision and the dangers associated with multiplicity of analyses and outcomes.	
Generalizability	21	Generalizability (external validity) of the trial findings.	
Overall evidence	22	General interpretation of the results in the context of current evidence.	

Note. From CONSORT Statement.

Retrieved from http://www.consort-statement.org/Statement/revisedstatement.htm

APPENDIX D

THE CONSORT E-FLOWCHART



Retrieved from http://www.consort-statement.org/Statement/revisedstatement.htm

APPENDIX E

TREND CHECKLIST OF ITEMS TO INCLUDE WHEN REPORTING NONRANDOMIZED TRIALS

Box E.1: TREND Checklist

Paper Section /Topic	Item No.	Descriptor	Examples From HIV Behavioral Prevention Research
Title and abstract	1	• Information on how units were allocated to interventions • Structured abstract recommended • **Information on target population or study sample**	Example (title): A nonrandomized trial of a clinic-based HIV counseling intervention for African American female drug users
Introduction Background	2	• Scientific background and explanation of rationale • **Theories used in designing behavioral interventions**	Example (theory used): the community-based AIDS intervention was based on social learning theory
Methods Participants	3	• Eligibility criteria for participants, **including criteria at different levels in recruitment/sampling plan (e.g., cities, clinics, subjects)** • Method of recruitment (e.g., referral, self-selection), including the **sampling method** if a systematic sampling plan was implemented • **Recruitment setting** • Settings and locations where the data were collected	Example (sampling method): using an alphanumeric sorted list of possible venues and times for identifying eligible subjects, every tenth venue-time unit was selected for the location and timing of recruitment Examples (recruitment setting): subjects were approached by peer opinion leaders during conversations at gay bars

Box E.1: TREND Checklist (Continued)

Paper Section /Topic	Item No.	Descriptor	Examples From HIV Behavioral Prevention Research
Interventions		• Details of the interventions intended for each study condition and how and when they were actually administered, specifically including:	
		Content: what was given?	Example (unit of delivery): the intervention was delivered to small groups of 5-8 subjects
		Delivery method: how was the content given?	
		Unit of delivery: how were subjects grouped during delivery?	
		Deliverer: who delivered the intervention?	
		Setting: where was the intervention delivered?	Examples (setting): the intervention was delivered in the bars; the intervention was delivered in the waiting rooms of sexually transmitted disease clinics
	4	Exposure quantity and duration: How many sessions or episodes or events were intended to be delivered? How long were they intended to last?	Examples (exposure quantity and duration): the intervention was delivered in five 1-hour sessions; the intervention consisted of standard HIV counseling and testing (pretest and posttest counseling sessions, each about 30 minutes)
		Time span: how long was it intended to take to deliver the intervention to each unit?	Examples (time span): each intervention session was to be delivered (in five 1-hour sessions) once a week for 5 weeks; the intervention was to be delivered over a 1-month period.
		Activities to increase compliance or adherence (e.g., incentives)	Example (activities to increase compliance or adherence): bus tokens and food stamps were provided
Objectives	5	• Specific objectives and hypotheses	

Box E.1: TREND Checklist (Continued)

Paper Section /Topic	Item No.	Descriptor	Examples From HIV Behavioral Prevention Research
Outcomes	6	• Clearly defined primary and secondary outcome measures • **Methods used to collect data** and any methods used to enhance the quality of measurements • Information on validated instruments such as psychometric and biometric properties	Examples (method used to collect data): self-report of behavioral data using a face-to-face interviewer-administered questionnaire; audio-computer-assisted self-administered instrument
Sample size	7	• How sample size was determined and, when applicable, explanation of any interim analyses and stopping rules	
Assignment method	8	• Unit of assignment (the unit being assigned to study condition, e.g., individual, group, community) • Method used to assign units to study conditions, including details of any restriction (e.g., blocking, stratification, minimization) • Inclusion of aspects employed to help minimize potential bias induced due to nonrandomization (e.g., matching)	Example 1 (assignment method): subjects were assigned to study conditions using an alternating sequence wherein every other individual enrolled (e.g., 1, 3, 5, etc.) was assigned to the intervention condition and the alternate subjects enrolled (e.g., 2, 4, 6, etc.) were assigned to the comparison condition Example 2 (assignment method): for odd weeks (e.g. 1, 3, 5), subjects attending the clinic on Monday, Wednesday, and Friday were assigned to the intervention condition and those attending the clinic on Tuesday and Thursday were assigned to the comparison condition; this assignment was reversed for even weeks
Blinding (masking)	9	• Whether or not participants, those administering the interventions, and those assessing the outcomes were blinded to study condition assignment; if so, statement regarding how the blinding was accomplished and how it was assessed	Example (blinding): the staff member performing the assessments was not involved in implementing any aspect of the intervention and knew the participants only by their study identifier number

Box E.1: TREND Checklist (Continued)

Paper Section /Topic	Item No.	Descriptor	Examples From HIV Behavioral Prevention Research
Unit of analysis	10	• Description of the smallest unit that is being analyzed to assess intervention effects (e.g., individual, group, or community) • If the unit of analysis differs from the unit of assignment, the analytical method used to account for this (e.g., adjusting the standard error estimates by the design effect or using multilevel analysis)	Example 1 (unit of analysis): since groups of individuals were assigned to study conditions, the analyses were performed at the group level, where mixed effects models were used to account for random subject effects within each group Example 2 (unit of analysis): since analyses were performed at the individual level and communities were randomized, a prior estimate of the intraclass correlation coefficient was used to adjust the standard error estimates before calculating confidence intervals
Statistical methods	11	• Statistical methods used to compare study groups for primary outcome(s), including complex methods for correlated data • Statistical methods used for additional analyses, such as subgroup analyses and adjusted analysis **• Methods for imputing missing data, if used** **• Statistical software or programs used**	
Results Participant flow	12	• Flow of participants through each stage of the study: enrollment, assignment, allocation and intervention exposure, follow-up, analysis (a diagram is strongly recommended) Enrollment: the numbers of participants screened for eligibility, found to be eligible or not eligible, declined to be enrolled, and enrolled in the study Assignment: the numbers of participants assigned to a study condition Allocation and intervention exposure: the number of participants assigned to each study condition and the number of participants who received each intervention	

Box E.1: TREND Checklist (Continued)

Paper Section /Topic	Item No.	Descriptor	Examples From HIV Behavioral Prevention Research
		Follow-up: the number of participants who completed the follow-up or did not complete the follow-up (i.e., lost to follow-up), by study condition Analysis: the number of participants included in or excluded from the main analysis, by study condition • Description of protocol deviations from study as planned, along with reasons	
Recruitment	13	• Dates defining the periods of recruitment and follow-up	
Baseline data		• Baseline demographic and clinical characteristics of participants in each study condition	
	14	• Baseline characteristics for each study condition relevant to specific disease prevention research • Baseline comparisons of those lost to follow-up and those retained, overall and by study condition • Comparison between study population at baseline and target population of interest	Example (baseline characteristics specific to HIV prevention research): HIV serostatus and HIV testing behavior
Baseline equivalence	15	• Data on study group equivalence at baseline and statistical methods used to control for baseline differences	Example (baseline equivalence): the intervention and comparison groups did not statistically differ with respect to demographic data (gender, age, race/ ethnicity; P > .05 for each), but the intervention group reported a significantly greater baseline frequency of injection drug use (P = .03); all regression analyses included baseline frequency of injection drug use as a covariate in the model

Box E.1: TREND Checklist (Continued)

Paper Section /Topic	Item No.	Descriptor	Examples From HIV Behavioral Prevention Research
Numbers analyzed		• Number of participants (denominator) included in each analysis for each study condition, particularly when the denominators change for different outcomes; statement of the results in absolute numbers when feasible	Example (number of participants included in the analysis): the analysis of condom use included only those who reported at the 6-month follow-up having had vaginal or anal sex in the past 3 months (75/125 for intervention group and 35/60 for standard group)
	16	• Indication of whether the analysis strategy was "intention to treat" or, **if not, description of how noncompliers were treated in the analyses**	Example ("intention to treat"): the primary analysis was intention to treat and included all subjects as assigned with available 9-month outcome data (125 of 176 assigned to the intervention and 110 of 164 assigned to the standard condition)
Outcomes and estimation		• For each primary and secondary outcome, a summary of results for each study condition, and the estimated effect size and a confidence interval to indicate the precision	
	17	• Inclusion of null and negative findings	
		• **Inclusion of results from testing prespecified causal pathways through which the intervention was intended to operate, if any**	
Ancillary analyses		• Summary of other analyses performed, including subgroup or restricted analyses, indicating which are prespecified or exploratory	Example (ancillary analyses): although the study was not powered for this hypothesis, an exploratory analysis shows that the intervention effect was greater among women than among men (although not statistically significant)
	18		

Box E.1: TREND Checklist (Continued)

Paper Section /Topic	Item No.	Descriptor	Examples From HIV Behavioral Prevention Research
Adverse events	19	• Summary of all important adverse events or unintended effects in each study condition (including summary measures, effect size estimates, and confidence intervals)	Example (adverse events): police cracked down on prostitution, which drove the target population, commercial sex workers, to areas outside the recruitment/sampling area
Discussion Interpretation	20	• Interpretation of the results, taking into account study hypotheses, sources of potential bias, imprecision of measures, multiplicative analyses, and other limitations or weaknesses of the study • Discussion of results taking into account the mechanism by which the intervention was intended to work (causal pathways) or alternative mechanisms or explanations • Discussion of the success of and barriers to implementing the intervention, fidelity of implementation • Discussion of research, programmatic, or policy Implications	
Generalizability	21	• Generalizability (external validity) of the trial findings, taking into account the study population, the characteristics of the intervention, length of follow-up, Incentives, compliance rates, specific sites/settings involved in the study, and other contextual issues	
Overall evidence	22	• General interpretation of the results in the context of current evidence and current theory	

Note. Masking (blinding) of participants or those administering the intervention may not be relevant or possible for many behavioral interventions. Theories used to design the interventions (see item 2) could also be reported as part of item 4. The comparison between study population at baseline and target population of interest (see item 14) could also be reported as part of item 21. Descriptors appearing in boldface are specifically added, modified, or further emphasized from the CONSORT statement. Boldface topic and descriptors are not included in the CONSORT statement but are relevant for behavioral interventions using nonrandomized experimental designs. The CONSORT statement (n11) or the explanation document for the CONSORT statement (n18) provides relevant examples for any topic or descriptor that is not in boldface. A structured format of the discussion is presented in Annals of Internal Medicine (information for authors: www.annals.org, accessed September 16, 2003).

APPENDIX F

THE STARD CHECKLIST FOR REPORTING INFORMATION IN DIAGNOSTIC ACCURACY STUDIES

Box F.1: STARD Checklist

Section and Topic	*Item*	*Description*	*On page #*
Title, abstract, and keywords	1	Identify the article as a study of diagnostic accuracy (recommend keyword for PsycINFO "diagnostic efficiency"; recommend MeSH heading for Medline "sensitivity and specificity")	
Introduction	2	State the research questions or aims, such as estimating diagnostic accuracy or comparing accuracy between tests or across participant groups	
Methods:			
Participants	3	Describe the study population: the inclusion and exclusion criteria and the settings and locations where the data were collected	
	4	Describe participant recruitment: was this based on presenting symptoms, results from previous tests, or the fact that the participants had received the index tests or the reference standard?	
	5	Describe participant sampling: was this a consecutive series of participants defined by selection criteria in items 3 and 4? If not, specify how participants were further selected	
	6	Describe data collection: was data collection planned before the index tests and reference standard were performed (prospective study) or after (retrospective study)?	
Test methods	7	Describe the reference standard and its rationale	
	8	Describe technical specifications of material and methods involved, including how and when measurements were taken, or cite references for index tests or reference standard, or both	
	9	Describe definition of and rationale for the units, cut-off points, or categories of the results of the index tests and the reference standard	
	10	Describe the number, training, and expertise of the persons executing and reading the index tests and the reference standard	

Box F.1: STARD Checklist (Continued)

Section and Topic	*Item*	*Description*	*On page #*
	11	Were the readers of the index tests and the reference standard blind (masked) to the results of the other test? Describe any other clinical information available to the readers of the index test.	
Statistical	12	Describe methods for calculating or comparing measures of diagnostic accuracy and the statistical methods used to quantify uncertainty (eg 95% confidence intervals)	
	13	Describe methods for calculating test reproducibility (e.g., interrater reliability), if done	
Results:			
Participants	14	Report when study was done, including beginning and ending dates of recruitment	
	15	Report clinical and demographic characteristics (e.g., age, sex, spectrum of presenting symptoms, comorbidity, current treatments, and recruitment center)	
	16	Report how many participants satisfying the criteria for inclusion did or did not undergo the index tests or the reference standard, or both; describe why participants failed to receive either test (a flow diagram is strongly recommended)	
Test results	17	Report time interval from index tests to reference standard, and any treatment administered between	
	18	Report distribution of severity of disease (define criteria) in those with the target condition and other diagnoses in participants without the target condition	
	19	Report a cross tabulation of the results of the index tests (including indeterminate and missing results) by the results of the reference standard; for continuous results, report the distribution of the test results by the results of the reference standard	
	20	Report any adverse events from performing the index test or the reference standard	

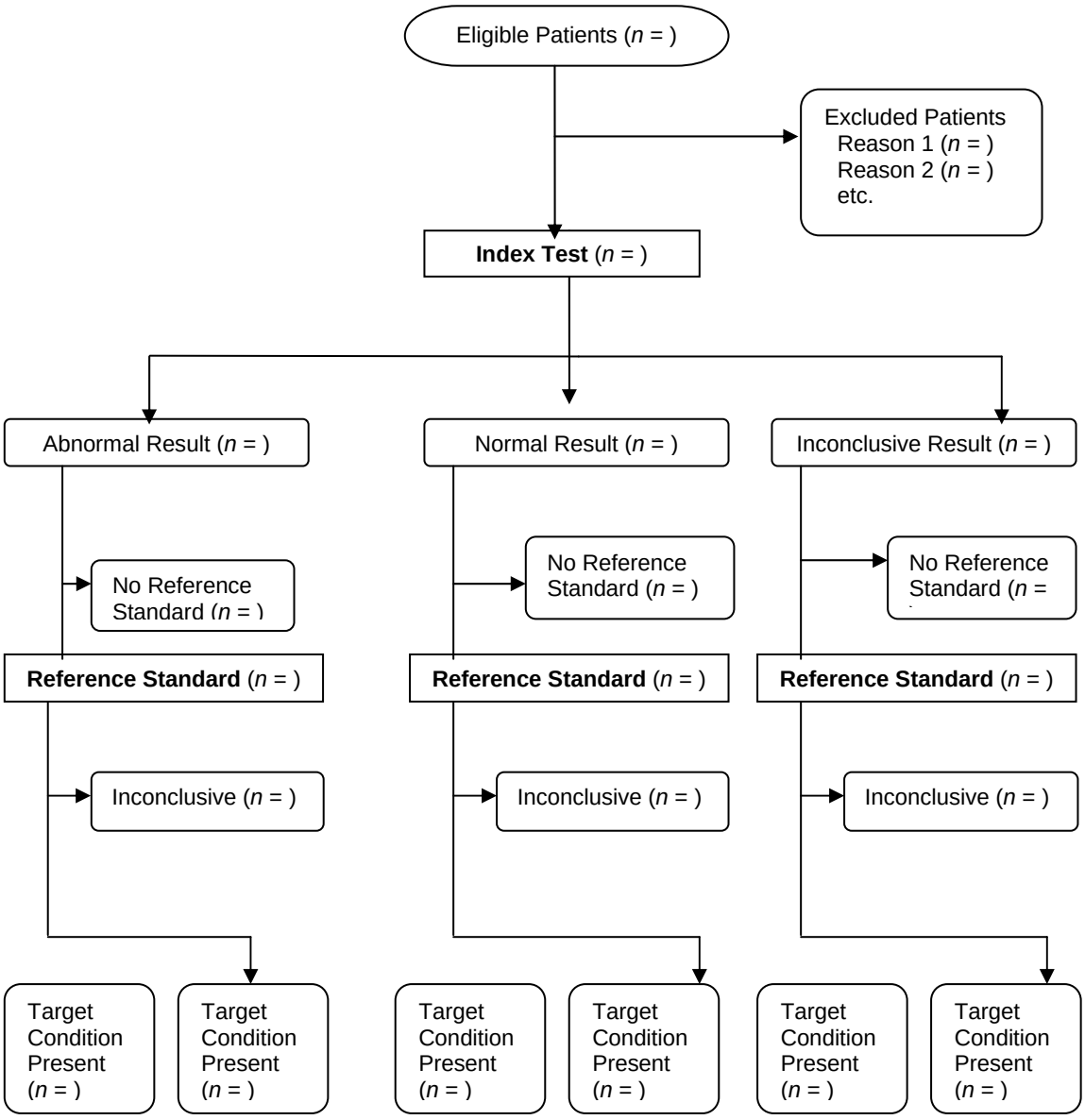
Box F.1: STARD Checklist (Continued)

Section and Topic	*Item*	*Description*	*On page #*
Estimates	21	Report estimates of diagnostic accuracy and measures of statistical uncertainty (e.g. 95% confidence intervals)	
	22	Report how indeterminate results, missing responses, and outliers of index tests were handled	
	23	Report estimates of variability of diagnostic accuracy between readers, centers, or subgroups of participants, if done	
	24	Report estimates of test reproducibility (e.g., interrater reliability), if done	
Discussion	25	Discuss the clinical applicability of the study findings	

Retrieved from http://www.consortstatement.org/ stardstatement.htm

APPENDIX G

STARD FLOWCHART



Retrieved from http://www.consortstatement.org/ stardstatement.htm

APPENDIX H

PROCEDURES USED TO DEVELOP AND VALIDATE THE RESEARCH DESIGN CLASSIFICATION SYSTEM

Elaborated below are the procedures used to develop and validate the research design classification system and the corresponding code numbers presented in Table 2.9. Idealistically this classification system was designed to produce **mutually exclusive** and **exhaustive** categories. With this ideal clearly in mind, the coding system was developed and refined in four sequential phases.

Phase One

Phase one was used to evaluate the extent to which the 16 categories in the classification system shown in Table 2.9 were unique and also adequate for placing each empirical study in a single category. (This and all subsequent tables are elaborated in chapter two of the dissertation.) This phase has two stages.

Stage One. Using all 16 empirical articles in the first issue of Volume 47 of the *JSLHR*, four coders independently placed each of these articles in one of eight categories elaborated in Table 2.6. Those articles that did not meet the established criteria of the first seven categories were placed in the eighth category identified as Not Elsewhere Classified (NEC).

However, an initial inspection of articles found that the assumption of one-to-one correspondence between article and study did not hold. One article (#5) was difficult to classify since it contained three independent studies rather than a single study. This article (along with all other articles having multiple studies in

subsequent issues) was removed from the initial coding task and placed in a separate code. To respect this difference between article population and study population, these recoded studies will be analyzed using the same classification task after all articles containing single studies. The removal of the article with multiple studies (#5) left 15 articles with single studies included in the stage one code.

Results of the stage one code yielded a high level of agreement (100%) on 14 articles out of 15 articles. The results of this activity are shown in Box H.1.

Box H.1: Issue 1 Stage 1

Article Number	Experimental Study I	Experimental Study II	Intervention Evaluation Research I	Intervention Evaluation Research II	Intervention Evaluation Research III	Diagnostic Instrument Development I	Diagnostic Instrument Development II	Not Elsewhere Classified
1								G, T, S, J
2								G, T, S, J
3				G, J				T,S,
4								G, T, S, J
6								G, T, S, J
7								G, T, S, J
8								G, T, S, J
11								G, T, S, J
12					G, T, S, J			
13								G, T, S, J
14								G, T, S, J
16								G, T, S, J
17								G, T, S, J
18								G, T, S, J
19							G, T, S, J	

Opinions of the four coders differed on the classification of one article (#3) in the stage one coding task. This article (#3) was reconciled using an independent reviewer. If needed, this same procedure will be used to reconcile future stage one coding differences. At this time, the reconciliation resulted in

the remaining article (#3) being included in the NEC category. As all articles were able to be placed in the provided categories, there was no need to add a new category. Thus, the end result was final agreement on 15 articles out of 15 articles. Accordingly, the need to satisfy both mutually exclusive and exhaustive was met.

An inspection of Box H.1 shows that all but two articles fell in the eighth category of NEC. These 13 articles contained in NEC provided the input for the stage two coding task using the classification system described in Table 2.8.

Stage Two. Using the 13 empirical studies contained in the NEC category from stage one, four coders independently placed each of these articles in one of the nine categories elaborated in Table 2.8. The stage two coding results are provided in Box H.2.

Box H.2: Issue 1 Stage 2

Article #	Cross-sectional Descriptive	Cross-sectional Predictive	Cross-sectional Explanatory	Longitudinal Descriptive	Longitudinal Predictive	Longitudinal Explanatory	Retrospective Descriptive	Retrospective Predictive	Retrospective Explanatory
1			G, T, S, J						
2			G, T, S, J						
3			G, T, S, J						
4			G, T, S, J						
6			G, T, S, J						
7				G, T, S, J					
8	G, T, J		S						
11			G, T, S, J						
13			S, J	G	T				
14	J		G, T, S						
16			G, T, S, J						
17			G, T, S, J						
18			G, T, S, J						

The stage two coding task resulted in a high level of agreement (100%) on ten articles out of 13 articles. Three of the four independent coders agreed on two of the remaining three articles (#8 and #14). Two coders were in agreement on the remaining article (#13).

To reconcile the classification of the three articles (#8, #13, and #14), an independent reviewer read the articles and cast the deciding (final) vote. This action resulted in article #8 classified as cross-sectional descriptive, article #13 classified as cross-sectional explanatory, and article #14 classified as cross-sectional descriptive. The independent reviewer provided a rationale for each final decision in a debriefing session attended by all independent coders.

The final result of the stage two coding task was 100% agreement. No additional categories were needed in the stage two classification system. Thus, the need to satisfy both the mutually exclusive criterion and the exhaustive criterion was met.

Phase Two

Phase two was a second attempt to replicate the phase one coding results including the ability of the classification system to provide both mutually exclusive and exhaustive categories. This phase replicates both the procedures and the criteria of mutually exclusive and exhaustive using Volume 47, Issue Two of the *JSLHR* with four independent coders.

Stage One. Using the 15 empirical articles in the second issue of Volume 47 of the *JSLHR*, four coders independently placed each of these articles in one of eight categories elaborated in Table 2.6. Those articles that did not meet the established criteria of the first seven categories were placed in the eighth category identified as NEC.

An initial inspection of articles found that the assumption of one-to-one correspondence between article and study did not hold for three articles. Two articles (# 20 and #37) contained two independent studies and one article (#32) contained three independent studies. These three articles (#20, #32, and #37) were removed from the stage one coding task and placed in a separate code. There were twelve articles left with single studies included in the stage one coding task.

Results of the stage one code yielded complete agreement (100%) on all of the articles. The results of this activity are shown in Box H.3. Again, the need to satisfy both mutually exclusive and exhaustive was met.

An inspection of Box H.3 shows that eleven of the twelve single-study articles fell in the NEC category. The one single-study article (#35) that did not fall in the NEC category explained the development of a diagnostic instrument without concurrent validity. The eleven articles that fell in the NEC category were retained for use in the stage two coding task.

Box H.3: Issue 2 Stage 1

Article Number	Experimental Study I	Experimental Study II	Intervention Evaluation Research I	Intervention Evaluation Research II	Intervention Evaluation Research III	Diagnostic Instrument Development I	Diagnostic Instrument Development II	Not Elsewhere Classified
21								G, D, S, J
22								G, D, T, J
23								G, D, T, J
24								G, D, S, J
25								G, S, T, J
26								G, S, T, J
30								G, S, T, J
31								G, D, T, J
33								G, D, T, J
34								G, D, S, J
35							G, S, D, J	
36								G, S, D, J

Stage Two. Using the eleven empirical studies contained in the NEC category from stage one, four coders independently placed them in one of the nine categories elaborated in Table 2.8. The stage two coding results are provided in Box H.4.

The stage two coding task resulted in a high level of agreement (100%) on nine articles out of eleven articles. The two remaining articles (#34 and #36) were evenly divided between the four independent coders.

To reconcile the classification of the two articles (#34 and #36), an independent reviewer read the articles and cast the deciding (final) vote. This action resulted in both articles being classified as cross-sectional explanatory. The independent reviewer provided a rationale for each final decision in a debriefing session attended by all independent coders.

The final result of the stage two coding task was 100% agreement. No additional categories were needed in the stage two classification system. Thus,

the need to satisfy both the mutually exclusive criterion and the exhaustive criterion was met.

Box H.4: Issue 2 Stage 2

Article #	Cross-sectional Descriptive	Cross-sectional Predictive	Cross-sectional Explanatory	Longitudinal Descriptive	Longitudinal Predictive	Longitudinal Explanatory	Retrospective Descriptive	Retrospective Predictive	Retrospective Explanatory
21			G, D, J, S						
22	G, D, T, J								
23	G, D, T, J								
24			G, D, S, J						
25						T, S, G, J			
26									S, G, T, J
30	S, G, T, J								
31					D, G, J, T				
33	D, T, G, J								
34	D, G		S, J						
36	G, D		S, J						

Thus, the phase one findings were replicated given that all articles fill in one category and no new categories were needed.

Phase Three

Given the results of the first two phases (satisfying the criteria of mutually exclusive and exhaustive) and the experience the coders gained by participating in the coding tasks, the phase three coding efforts continued using three rather than four coders. Initially, both issue three and issue four of Volume 47 contained 15 empirical articles each. Following the procedures established in phases one and two, the three coders independently classified each these articles.

Stage One. An initial inspection of articles found that the assumption of one-to-one correspondence between article and study did not hold for three articles. All three articles (#59, #61 and #62) contained two independent studies. These three articles were removed from the stage one coding task and placed in a separate code. One article (#69), a meta-analysis, was removed from the article population.

Using the 26 empirical articles in the third and fourth issue of Volume 47 of the *JSLHR*, three coders independently placed each of these articles in one of eight categories elaborated in Table 2.6. Those articles that did not meet the established criteria of the first seven categories were placed in the eighth category identified as NEC.

Results of the stage one code yielded 100% agreement. The results of this activity are shown in Box H.5. Again, the need to satisfy both mutually exclusive and exhaustive was met.

An inspection of Box H.5 shows that 24 or the 26 articles are classified as NEC. One article (#47) is an experimental study that has nonrandom selection of participants with random allocation to treatment. The remaining article (#63) is a single case study. The articles that fell in the NEC category were retained for use in the stage two coding task.

Box H.5: Issues 3 and 4, Stage 1

Article Number	Experimental Study I	Experimental Study II	Intervention Evaluation Research I	Intervention Evaluation Research II	Intervention Evaluation Research III	Diagnostic Instrument Development I	Diagnostic Instrument Development II	Not Elsewhere Classified
38								S, J, G
39								S, J, G
40								S, J, G
41								S, J, G
42								T, J, G
43								T, J, G
44								T, J, G
46								T, J, G
47		S, T, G						
48								T, S, G
49								T, S, G
50								T, S, G
51								D, J, G
52								T, D, G
53								D, S, G
55								J, D, G
56								J, D, G
57								J, D, G
58								S, T, G
60								S, T, G
63					T, S, G			
64								S, T, G
65								T, D, G
66								T, D, G
67								J, D, G
68								J, D, G

Stage Two. Using the empirical studies contained in the NEC category from stage one, three coders independently placed them in one of the nine categories elaborated in Table 2.8. The stage two coding results are provided in Box H.6.

The stage two coding task resulted in a high level of agreement (100%) on 20 articles out of 24 articles retained from the stage two code.

Box H.6: Issues 3 and 4, Stage 2

Article #	Cross-sectional Descriptive	Cross-sectional Predictive	Cross-sectional Explanatory	Longitudinal Descriptive	Longitudinal Predictive	Longitudinal Explanatory	Retrospective Descriptive	Retrospective Predictive	Retrospective Explanatory
38	S, J, G								
39	S, J, G								
40						S, J, G			
41							S,J,G		
42	T, J, G								
43	T, J, G								
44	T		J,G						
46	T, J, G								
48	S, T, G								
49		S, T, G							
50			T, S, G						
51				D	J, G				
52				T, D, G					
53				S, D, G					
55	J, D, G								
56		J, D, G							
57					J, D, G				
58	S, T, G								
60			S, T, G						
64			S, T, G						
65	T, D, G								
66	T, D, G								
67	D		J, G						
68				D	J, G				

The stage two coding task resulted in a high level of agreement (100%) on 20 articles out of 24 articles. Two of the three independent coders agreed on four of the remaining articles (#44, #51, #67, and #68).

To reconcile the classification of the four articles (#44, #51, #67, and #68), an independent reviewer read the articles and cast the deciding (final) vote. This action resulted in two articles (#44 and #67) being classified as cross-sectional descriptive. The two remaining articles fell under the category of longitudinal descriptive. The independent reviewer provided a rationale for each final decision in a debriefing session attended by all independent coders.

The final result of the stage two coding task was 100% agreement. No additional categories were needed in the stage two classification system. Thus, the need to satisfy both mutually exclusive and exhaustive was met.

Phase Four

Given that the criteria of mutually exclusive and exhaustive category requirements continued to be met and the additional experience coders gained by participating in the phase three coding tasks, the phase four coding efforts continued using two coders. Issues five and six plus the articles having multiple studies were included in this phase. Initially, issue five contained 17 empirical articles, issue six contained 15 empirical articles, and ten empirical articles were in the recode for articles containing more than one study. Following the procedures established in the first three phases, the two coders independently classified each of these articles.

Stage One. An initial inspection of articles found that the assumption of one-to-one correspondence between article and study did not hold for three articles. Two articles (#78 and #104) contained two independent studies and one article (#77) contained three independent studies. These three articles were removed from the stage one coding task and placed in a separate coding task. The ten recoded articles consisted of seven articles that contained two studies each (#20, #37, #59, #61, #62, #78, and #104) and three articles that contained three studies each (#5, #32, and #77).

Using the 29 single-study empirical articles in the fifth and sixth issue of Volume 47 of the *JSLHR*, and the 23 recoded articles provided 52 studies. Two coders independently placed each of these studies in one of eight categories elaborated in Table 2.6. Those studies that did not meet the established criteria of the first seven categories were placed in the eighth category identified as NEC.

Results of the stage one code yielded complete agreement (100%) between the two coders. The results of this activity are shown in Box H.7. Again, the need to satisfy both mutually exclusive and exhaustive was met.

An inspection of Box H.7 shows that there were three single case studies (Intervention Evaluation Research III). Moreover, 40 of the studies fell in the NEC category. These studies were retained for use in the stage two coding task.

Stage Two. Using the 40 empirical studies contained in the NEC category from stage one, two coders independently placed them in one of the nine categories elaborated in Table 2.8. The stage two coding results are provided in Box H.8. There was 100% agreement between the two independent coders. Again the need to satisfy both mutually exclusive and exhaustive was met.

Box H.7: Issues 5 and 6 and Articles with Multiple Studies, Stage 1

Article Number	Experimental Study I	Experimental Study II	Intervention Evaluation Research I	Intervention Evaluation Research II	Intervention Evaluation Research III	Diagnostic Instrument Development I	Diagnostic Instrument Development II	Not Elsewhere Classified
71								T, G
72		T, G						
73		T, G						
74								S, G
75								S, G
76								S, G
79								S, G
81								J, G
82								J, G
83								J, G
84				J, G				
85								D, G
86								D, G
87								D, G
88								D, G
92								S, G
94								S, G
95								S, G
96				T, G				
97								T, G
98								T, G
99					D, G			
100								D, G
101								D, G
102								D, G
103								J, G
105								J, G
106					J, G			
107								J, G
5.1								S, G
5.2								S, G
5.3								S, G
20.1								S, G
20.2								S, G
32.1		T, G						
32.2		T, G						
32.3								T, G
37.1							T, G	
37.2							T, G	
59.1								D, G
59.2					D, G			
61.1			D, G					
61.2								D, G
62.1								D, G
62.2								D, G
77.1								J, G
77.2								J, G
77.3								J, G
78.1								J, G
78.2								J, G
104.1								J, G
104.2								J, G

Box H.8: Issues 5 and 6 and Articles with Multiple Studies, Stage 2

Article Number	Cross-sectional Descriptive	Cross-sectional Predictive	Cross-sectional Explanatory	Longitudinal Descriptive	Longitudinal Predictive	Longitudinal Explanatory	Retrospective Descriptive	Retrospective Predictive	Retrospective Explanatory
71			T, G						
74	S, G								
75	S, G								
76			S, G						
79	S, G								
81	J, G								
82			J, G						
83			J, G						
85				D, G					
86	D, G								
87	D, G								
88	D, G								
92	S, G								
94	S, G								
95	S, G								
97	T, G								
98				T, G					
100	D, G								
101	D, G								
102				D, G					
103			J, G						
105			J, G						
107			J, G						
5.1	S, G								
5.2	S, G								
5.3	S, G								
20.1	S, G								
20.2	S, G								
32.3			T, G						
59.1	D, G								
61.2				D, G					
62.1	D, G								
62.2	D, G								
77.1			J, G						
77.2			J, G						
77.3			J, G						
78.1			J, G						
78.2			J, G						
104.1			J, G						
104.2			J, G						

Summary

This phase provides a summative report of both the stage one and stage two coding results.

In the stage one coding task, the coders independently classified the 105 studies in the research synthesis study population. A high level of agreement (100%) was met on 104 of the total 105 studies. This yielded an interrater reliability of 99% on the stage one coding.

An inspection of Box H.9 found that of the 105 studies that made up the stage one coding, 88 studies (84%) fell in the Not Elsewhere Classified category. These 88 studies were retained for the stage two coding. Two categories, Experimental Study I and Diagnostic Instrument Development I, held no studies.

The 88 studies retained from the stage one coding provided the basis of the stage two coding. The coders reached agreement on 79 of the 88 studies in the stage two coding task. This high level of agreement translates to an interrater reliability of 90%.

Results of the stage two coding task as reported in Box H.10, found that only two categories (cross-sectional descriptive and cross-sectional explanatory) contained 80% (71 out of 88) of the studies included in this coding task. Interestingly, only one category (retrospective-predictive) held no studies.

The purpose of this exercise was to develop and validate the research design classification system of empirical studies developed in Phase II of this

research synthesis. Taken collectively, the research design classification system was adequate to capture the research synthesis population of 105 studies. Additionally, all of the studies fell in one and only one of the categories in the research design classification system. Thus, this classification system fulfilled the expectations to produce **mutually exclusive** and **exhaustive** categories.

Box H.9: Stage 1 Reconciled Codes

Article Log Number	Study Number	Experimental Study I	Experimental Study II	Intervention Evaluation Research I	Intervention Evaluation Research II	Intervention Evaluation Research III	Diagnostic Instrument Development I	Diagnostic Instrument Development II	Not Elsewhere Classified
1	1								G, T, S, J
2	2								G, T, S, J
3	3								rectified
4	4								G, T, S, J
6	5								G, T, S, J
7	6								G, T, S, J
8	7								G, T, S, J
11	8								G, T, S, J
12	9					G, T, S, J			
13	10								G, T, S, J
14	11								G, T, S, J
16	12								G, T, S, J
17	13								G, T, S, J
18	14								G, T, S, J
19	15							G, T, S, J	
21	16								G, D, S, J
22	17								G, D, T, J
23	18								G, D, T, J
24	19								G, D, S, J
25	20								G, S, T, J
26	21								G, S, T, J
30	22								G, S, T, J
31	23								G, D, T, J
33	24								G, D, T, J
34	25								G, D, S, J
35	26							G, S, D, J	
36	27								G, S, D, J
38	28								S, J, G
39	29								S, J, G
40	30								S, J, G
41	31								S, J, G
42	32								T, J, G
43	33								T, J, G
44	34								T, J, G
46	35								T, J, G
47	36		S, T, G						
48	37								T, S, G
49	38								T, S, G
50	39								T, S, G
51	40								D, J, G
52	41								T, D, G
53	42								D, S, G
55	43								J, D, G
56	44								J, D, G
57	45								J, D, G
58	46								S, T, G
60	47								S, T, G
63	48					T, S, G			
64	49								S, T, G
65	50								T, D, G
66	51								T, D, G
67	52								J, D, G
68	53								J, D, G

Box H.9: Stage 1 Reconciled Codes continued

Article Log Number	Study Number	Experimental Study I	Experimental Study II	Intervention Evaluation Research I	Intervention Evaluation Research II	Intervention Evaluation Research III	Diagnostic Instrument Development I	Diagnostic Instrument Development II	Not Elsewhere Classified
71	54								T, G
72	55		T, G						
73	56		T, G						
74	57								S, G
75	58								S, G
76	59								S, G
79	60								S, G
81	61								J, G
82	62								J, G
83	63								J, G
84	64				J, G				
85	65								D, G
86	66								D, G
87	67								D, G
88	68								D, G
92	69								S, G
94	70								S, G
95	71								S, G
96	72				T, G				
97	73								T, G
98	74								T, G
99	75					D, G			
100	76								D, G
101	77								D, G
102	78								D, G
103	79								J, G
105	80								J, G
106	81					J, G			
107	82								J, G
5.1	83								S, G
5.2	84								S, G
5.3	85								S, G
20.1	86								S, G
20.2	87								S, G
32.1	88		T, G						
32.2	89		T, G						
32.3	90								T, G
37.1	91							T, G	
37.2	92							T, G	
59.1	93								D, G
59.2	94					D, G			
61.1	95			D, G					
61.2	96								D, G
62.1	97								D, G
62.2	98								D, G
77.1	99								J, G
77.2	100								J, G
77.3	101								J, G
78.1	102								J, G
78.2	103								J, G
104.1	104								J, G
104.2	105								J, G

Box H.10: Stage 2 Reconciled Codes

Article Log Number	Study Number	Cross-sectional Descriptive	Cross-sectional Predictive	Cross-sectional Explanatory	Longitudinal Descriptive	Longitudinal Predictive	Longitudinal Explanatory	Retrospective Descriptive	Retrospective Predictive	Retrospective Explanatory
1	1			G, T, S, J						
2	2			G, T, S, J						
3	3			G, T, S, J						
4	4			G, T, S, J						
6	5			G, T, S, J						
7	6				G, T, S, J					
8	7	reconciled								
11	8			G, T, S, J						
13	10			reconciled						
14	11	reconciled								
16	12			G, T, S, J						
17	13			G, T, S, J						
18	14			G, T, S, J						
21	16			G, D, J, S						
22	17	G, D, T, J								
23	18	G, D, T, J								
24	19			G, D, S, J						
25	20						T, S, G, J			
26	21									S, G, T, J
30	22	S, G, T, J								
31	23					D, G, J, T				
33	24	D, T, G, J								
34	25			reconciled						
36	27			reconciled						
38	28	S, J, G								
39	29	S, J, G								
40	30						S, J, G			
41	31							S,J,G		
42	32	T, J, G								
43	33	T, J, G								
44	34	reconciled								
46	35	T, J, G								
48	37	S, T, G								
49	38		S, T, G							
50	39			T, S, G						
51	40				reconciled					
52	41				T, D, G					
53	42				S, D, G					
55	43	J, D, G								
56	44		J, D, G							
57	45					J, D, G				
58	46	S, T, G								
60	47			S, T, G						
64	49			S, T, G						
65	50	T, D, G								

Box H.10: Stage 2 Reconciled Codes continued

Article Log Number	Study Number	Cross-sectional Descriptive	Cross-sectional Predictive	Cross-sectional Explanatory	Longitudinal Descriptive	Longitudinal Predictive	Longitudinal Explanatory	Retrospective Descriptive	Retrospective Predictive	Retrospective Explanatory
66	51	T, D, G								
67	52	reconciled								
68	53				reconciled					
71	54			T, G						
74	57	S, G								
75	58	S, G								
76	59			S, G						
79	60	S, G								
81	61	J, G								
82	62			J, G						
83	63			J, G						
85	65				D, G					
86	66	D, G								
87	67	D, G								
88	68	D, G								
92	69	S, G								
94	70	S, G								
95	71	S, G								
97	73	T, G								
98	74				T, G					
100	76	D, G								
101	77	D, G								
102	78				D, G					
103	79			J, G						
105	80			J, G						
107	82			J, G						
5.1	83	S, G								
5.2	84	S, G								
5.3	85	S, G								
20.1	86	S, G								
20.2	87	S, G								
32.3	90			T, G						
59.1	93	D, G								
61.2	96				D, G					
62.1	97	D, G								
62.2	98	D, G								
77.1	99			J, G						
77.2	100			J, G						
77.3	101			J, G						
78.1	102			J, G						
78.2	103			J, G						
104.1	104			J, G						
104.2	105			J, G						

APPENDIX I

PROCEDURES USED TO DEVELOP AND VALIDATE THE POPULATION VALIDITY CLASSIFICATION SYSTEM

Elaborated below are findings from the procedures used to validate the coding of the population validity assessment strategy. The population validity assessment strategy was designed to classify each of 105 studies in the research synthesis population as having high population validity, moderate population validity, or low population validity.

These procedures are briefly discussed in the third component of the theoretical framework located in Chapter II. The procedures used in achieving the three objectives of developing, implementing, and validating the population validity assessment are described in the five subsequent phases. Phase one explains procedures used in developing the population validity assessment strategy. Phase two describes implementation procedures. Phases three, four, and five are devoted to findings for each of the substantive concerns of hearing, speech, and language.

Phase One

Phase one describes the process used to develop the population validity assessment strategy. The phase was developed in two stages.

Stage One. Five researchers were used to develop the population validity assessment. These five individuals were the same advanced doctoral-level students who participated in the third component (research design classification systems) of this theoretical framework.

The participation of these five individuals in the population validity coding provides additional continuity in this inquiry. These five researchers participated in weekly focus groups over a ten month period where dinner was provided. The weekly focus groups targeted topics of population validity and conclusion validity. Additionally, these five researchers possess a unique familiarity with the article formats, substantive topics, and the 16 types of research designs used in the research synthesis population of 105 studies.

Stage Two. The objective of stage two was to develop an inventory of questions that would allow an informed judgment to be made about the quality of population validity in each of the 105 studies in the research synthesis population.

Using knowledge gained from having (a) read this large corpus of studies, (b) developed the research design code, and (c) participated in weekly focus groups, these five researchers contributed questions that focused on making a determination of population validity within studies.

The author of this dissertation and a researcher with over 30 years experience compiled the contributed questions. Trends within the contributed questions were identified that when taken collectively provided a basis for making informed decisions about population validity.

Eight essential questions emerged. These questions are elaborated in Table 2.10 and discussed below.

Question one addresses how well the sample used in the study was linked to the purpose of the study. This information involves the authors providing information as to what characteristics the group of participants possessed that enabled them to be related to the purpose of the study.

Question two evaluates to what extent the target population was mentioned or defined.

Question three looks at how transparent the connection is between the sample and the target population. It was noted that often the sample was referred to as participants. The use of the term participants implies that those individuals included in the study are not representative of the population of interest, but rather a sample of convenience.

Questions four and five evaluate a priori decisions about sample size(s) and proposed data analysis strategies.

Questions six and seven look at how the authors, in light of the information previously discussed, temper the generalizability of study findings or address factors that might compromise the study findings.

Question eight asks researchers to make an informed decision about how well sampling procedures may be replicated.

Phase Two

Phase two describes the process used to implement the population validity assessment strategy. This phase involved focus group meetings and was developed in two stages.

Stage One. In stage one, the five coders were asked to read population validity assessment in Chapter II of the theoretical framework. Group discussions targeted the concept of population validity assessment in light of the theoretical framework and the eight essential questions.

Stage Two. The five coders were provided copies of the population validity questions and eight studies from the research synthesis population. Included in these eight studies were four studies that exemplified high population validity and four studies that exemplified low population validity.

The coders independently reviewed the studies and made notes justifying the ratings. At a follow-up meeting, the five coders independently reported their findings and agreed or disagreed with the proposed ratings of the sample studies. The independent reviewer provided a rationale for each final decision in a debriefing session attended by all independent coders. The final result of the stage two validation coding task was 100% agreement. No additional questions were needed to make the determination of population validity.

Phase Three

The purpose of phase three is to validate the population validity assessment for the substantive concern of hearing. This validation was achieved in two stages.

Stage One. Using the first 10 of 23 studies that addressed the substantive concern of hearing, four coders independently addressed the eight questions used to guide the decision about population validity. At the debriefing meeting, coders shared their findings. After all coders had the opportunity to share, each was asked to rate studies as having high population validity, moderate population validity, or low population validity. Stage one coding results are provided in Box I.1.

Results of the first coding yielded high agreement (90%) between the four coders. Only one study (#28) needed to be reconciled. This study was reconciled using an independent reviewer. If needed, this same procedure will be used to reconcile future coding differences. In this instance, reconciliation resulted in the study being classified as having high population validity.

Stage Two. Given the high rate of agreement between the four coders, the final 13 studies addressing the substantive topic of hearing were evaluated using three coders. Coding procedures outlined in stage one were used. Findings are presented in Box I.2

Results of the stage two coding task yielded high agreement (92%) between the three coders. Only one study (#43) needed to be reconciled. This resulted in the study being classified as having low population validity.

Box I.1: Code One Hearing

Article Log Number	Study Number	Type	High Population Validity	Moderate Population Validity	Low Population Validity
1	1	Hearing			T, S, J, G
2	2	Hearing			D, J, S, G
20	86	Hearing			T, J, D, G
20	87	Hearing			T, J, D, G
21	16	Hearing			T, S, D, G
22	17	Hearing			T, S, J, G
38	28	Hearing	J, S	G, D	
39	29	Hearing			T, D, J, G
40	30	Hearing			D, T, S, G
41	31	Hearing			T, S, J, G

Phase Four

The purpose of phase four is to validate the population validity assessment for the substantive concern of speech. This validation was achieved in two stages.

Box I.2: Code Two Hearing

Article Log Number	Study Number	Type	High Population Validity	Moderate Population Validity	Low Population Validity
55	43	Hearing		S	J, G
56	44	Hearing			T, G, J
57	45	Hearing		D, T, G	
71	54	Hearing			D, S, G
72	55	Hearing			S, J, G
73	56	Hearing			G, T, J
74	57	Hearing			T, D, G
75	58	Hearing			T, J, G
76	59	Hearing			S, J, G
77	99	Hearing			S, D, G
77	100	Hearing			S, D, G
77	101	Hearing			S, D, G
92	69	Hearing			D, T, G

Stage One. Given the training and high rate of agreement between the researchers, three coders were used in this stage. Using the first 10 of 23 studies that addressed the substantive concern of speech, three coders independently addressed the eight essential questions used to guide the decision about population validity. Stage one coding results for the substantive topic of speech are provided in Box I.3. Results of the first coding yielded high agreement (100%) between the three coders. No studies needed to be reconciled in the stage one speech coding task.

Box I.3: Speech

Article Log Number	Study Number	Type	High Population Validity	Moderate Population Validity	Low Population Validity
3	3	Speech			example
4	4	Speech			G, J, S
5	83	Speech			D, T, G
5	84	Speech			D, T, G
5	85	Speech			D, T, G
6	5	Speech			S, J, G
7	6	Speech			D, T,G
8	7	Speech			D, J, G
23	18	Speech		example	
24	19	Speech			S, T, G

Stage Two. Given the high rate of agreement between coders, the final 13 studies addressing the substantive concern of speech were evaluated using two coders. Stage two coding results for the substantive topic of speech are provided in Box I.4.

Results of the stage two coding task again yielded high agreement (100%) between the two coders. No studies needed to be reconciled in the stage two speech coding task.

Phase Five

The purpose of phase five is to validate the population validity for the substantive concern of language. This validation process was achieved in two stages.

Box I.4: Speech Code Two

Article Log Number	Study Number	Type	High Population Validity	Moderate Population Validity	Low Population Validity
25	20	Speech			D, G
26	21	Speech			T, G
42	32	Speech		S, G	
43	33	Speech			J, G
44	34	Speech			D, G
58	46	Speech			T, G
59	93	Speech			example
59	94	Speech			example
78	102	Speech			J, G
78	103	Speech			J, G
79	60	Speech			D, G
94	70	Speech			T, G
95	71	Speech			example

Stage One. Given the training and high rate of agreement between the coders, three coders were used in this stage. Using the first 20 of 59 studies that addressed the substantive concern of language, three coders independently addressed the eight essential questions used to guide the decision about population validity. Stage one coding results for the substantive topic of language are provided in Box I.5. Results of the first coding yielded high agreement (100%) between the three coders. No studies needed to be reconciled in the stage one language coding task.

Stage Two. Given the high rate of agreement between coders, the final 39 studies addressing the substantive concern of language were evaluated using two coders.

Stage two coding results for the substantive topic of language are provided in Box I.6. Results of the stage two coding task again yielded high

Box I.5: Language

Article Log Number	Study Number	Type	High Population Validity	Moderate Population Validity	Low Population Validity
11	8	Language		S, J, G	
12	9	Language			T, D, G
13	10	Language	S, J, G		
14	11	Language			T, D, G
16	12	Language			S, J, G
17	13	Language		D, T, G	
18	14	Language			J, S, G,
19	15	Language			T, D, G,
30	22	Language			J, S, G
31	23	Language		D, T, G	
32	88	Language			S, J, G
32	89	Language			S, J, G
32	90	Language			S, J, G
33	24	Language			T, D, G
34	25	Language			S, J, G
35	26	Language			T, D, G
36	27	Language	example		
37	91	Language			T, D, G
37	92	Language			T, D, G
46	35	Language			S, J, G

agreement (100%) between coders. No studies needed to be reconciled in the stage two language coding task.

Summary

This phase provides a summative report of both the stage one and stage two coding tasks across the substantive concerns of hearing, speech, and language.

In the population validity assessment coding task, coders independently classified the 105 studies in the research synthesis study population. A high level of agreement (100%) was achieved on 103 of the total 105 studies. This yielded an interrater reliability of 98%.

More in-depth discussions about population validity assessment findings is in Chapter IV which focuses on hearing, Chapter V which looks at speech, Chapter VI which targets language, and Chapter VII which evaluates trends across the substantive concerns of hearing, speech, and language.

Box I.6: Language Two

Article Log Number	Study Number	Type	High Population Validity	Moderate Population Validity	Low Population Validity
47	36	Language			D, G
48	37	Language			D, G
49	38	Language		D, G	
50	39	Language		D, G	
51	40	Language			D, G
52	41	Language			D, G
53	42	Language			D, G
60	47	Language			D, G
61	95	Language		D, G	
61	96	Language			D, G
62	97	Language			J, G
62	98	Language			J, G
63	48	Language			J, G
64	49	Language			J, G
65	50	Language			J, G
66	51	Language			J, G
67	52	Language			J, G
68	53	Language			J, G
81	61	Language			J, G
82	62	Language			J, G
83	63	Language			S, G
84	64	Language			S, G
85	65	Language			S, G
86	66	Language			S, G
87	67	Language			S, G
88	68	Language		S, G	
96	72	Language			S, G
97	73	Language			S, G
98	74	Language	example		
99	75	Language			T, G
100	76	Language			T, G
101	77	Language			T, G
102	78	Language	T, G		
103	79	Language			T, G
104	104	Language			T, G
104	105	Language			T, G
105	80	Language	example		
106	81	Language			T, G
107	82	Language			S, G

Box I.7: Reconciled Codes

Article Log Number	Study Number	Type	High Population Validity	Moderate Population Validity	Low Population Validity
1	1	Hearing			T, S, J, G
2	2	Hearing			D, J, S, G
20	86	Hearing			T, J, D, G
20	87	Hearing			T, J, D, G
21	16	Hearing			T, S, D, G
22	17	Hearing			T, S, J, G
38	28	Hearing		reconciled	
39	29	Hearing			T, D, J, G
40	30	Hearing			D, T, S, G
41	31	Hearing			T, S, J, G
55	43	Hearing			reconciled
56	44	Hearing			T, G, J
57	45	Hearing		D, T, G	
71	54	Hearing			D, S, G
72	55	Hearing			S, J, G
73	56	Hearing			G, T, J
74	57	Hearing			T, D, G
75	58	Hearing			T, J, G
76	59	Hearing			S, J, G
77	99	Hearing			S, D, G
77	100	Hearing			S, D, G
77	101	Hearing			S, D, G
92	69	Hearing			D, T, G
3	3	Speech			example
4	4	Speech			G, J, S
5	83	Speech			D, T, G
5	84	Speech			D, T, G
5	85	Speech			D, T, G
6	5	Speech			S, J, G
7	6	Speech			D, T,G
8	7	Speech			D, J, G
23	18	Speech		example	
24	19	Speech			S, T, G

Box I.7: Reconciled Codes (Continued)

25	20	Speech			D, G
26	21	Speech			T, G
42	32	Speech		S, G	
43	33	Speech			J, G
44	34	Speech			D, G
58	46	Speech			T, G
59	93	Speech			example
59	94	Speech			example
78	102	Speech			J, G
78	103	Speech			J, G
79	60	Speech			D, G
94	70	Speech			T, G
95	71	Speech			example
11	8	Language		S, J, G	
12	9	Language			T, D, G
13	10	Language	S, J, G		
14	11	Language			T, D, G
16	12	Language			S, J, G
17	13	Language		D, T, G	
18	14	Language			J, S, G,
19	15	Language			T, D, G,
30	22	Language			J, S, G
31	23	Language		D, T, G	
32	88	Language			S, J, G
32	89	Language			S, J, G
32	90	Language			S, J, G
33	24	Language			T, D, G
34	25	Language			S, J, G
35	26	Language			T, D, G
36	27	Language	example		
37	91	Language			T, D, G
37	92	Language			T, D, G
46	35	Language			S, J, G
47	36	Language			D, G
48	37	Language			D, G
49	38	Language		D, G	
50	39	Language		D, G	
51	40	Language			D, G
52	41	Language			D, G
53	42	Language			D, G
60	47	Language			D, G
61	95	Language		D, G	
61	96	Language			D, G
62	97	Language			J, G
62	98	Language			J, G
63	48	Language			J, G
64	49	Language			J, G
65	50	Language			J, G
66	51	Language			J, G

Box I.7: Reconciled Codes (Continued)

67	52	Language			J, G
68	53	Language			J, G
81	61	Language			J, G
82	62	Language			J, G
83	63	Language			S, G
84	64	Language			S, G
85	65	Language			S, G
86	66	Language			S, G
87	67	Language			S, G
88	68	Language		S, G	
96	72	Language			S, G
97	73	Language			S, G
98	74	Language	example		
99	75	Language			T, G
100	76	Language			T, G
101	77	Language			T, G
102	78	Language	T, G		
103	79	Language			T, G
104	104	Language			T, G
104	105	Language			T, G
105	80	Language	example		
106	81	Language			T, G
107	82	Language			S, G

APPENDIX J

PROCEDURES USED TO DEVELOP AND VALIDATE THE CONCLUSION VALIDITY CLASSIFICATION SYSTEM

Elaborated below are findings from the procedures used to develop and validate the coding of the conclusion validity assessment strategy. The conclusion validity assessment strategy was designed to classify each of 105 studies in the research synthesis population as having high conclusion validity, moderate conclusion validity, or low conclusion validity.

These procedures are briefly discussed in the fourth component of the theoretical framework located in Chapter II. The procedures used in achieving the three objectives of developing, implementing, and validating the conclusion validity assessment are described in more depth in the five subsequent phases. Phase one explains procedures used in developing the conclusion validity assessment strategy. Phase two describes implementation procedures. Phases three, four, and five are devoted to findings for each of the substantive concerns of hearing, speech, and language.

Phase One

Phase one describes the process used to develop the conclusion validity assessment strategy. The phase was developed in two stages.

Stage One. Five researchers were used to develop the conclusion validity assessment. These five individuals were the same advanced doctoral-level students who participated in the third component (research design classification systems) and fourth component (population validity assessment task) of this theoretical framework.

The participation of these five individuals in the conclusion validity coding provides additional continuity in this inquiry. These five researchers participated in weekly focus groups where dinner was provided. The weekly focus groups were over a ten month period and directly targeted the topics of population validity and conclusion validity. Additionally, these five researchers possess a unique familiarity with the article formats, the substantive topics, and the 16 types of research designs used in the research synthesis population of studies.

Stage Two. The objective of stage two was to develop an inventory of questions that would allow an informed judgment to be made about the quality of conclusion validity in each of the 105 studies in the research synthesis population.

Using the knowledge gained from having (a) read this large corpus of studies, (b) developed the research design code, (c) participated in the population assessment task, and (d) participated in weekly focus groups, these five researchers contributed questions that focused on making a determination of conclusion validity within studies.

Four evaluation components when taken collectively would provide a basis for making informed decisions about conclusion validity. These decisions would allow each of the 105 studies in the research synthesis population to be defined as having high conclusion validity, moderate conclusion validity, or low conclusion validity. These four evaluation components are presented in Table 2.11 and elaborated below.

Phase Two

Phase two describes the process used to implement the conclusion validity assessment strategy. This phase involved focus group meetings and was developed in two stages.

Stage One. In stage one, the five coders were asked to read conclusion validity assessment in Chapter II of the theoretical framework. Group discussions targeted the concept of conclusion validity assessment in light of the theoretical framework and the four evaluation components.

Stage Two. The five coders were provided copies of the conclusion validity assessment and ten studies from the research synthesis population that addressed the substantive topics of hearing. The coders independently read the studies and made notes justifying the ratings. At a follow-up meeting, the five coders independently reported their findings and using the discriminate analysis confirmed or denied the ratings of the sample studies. The independent reviewer provided a rationale for each final decision in a debriefing session attended by all independent coders. The results of the stage two coding task are presented in Box J.1. The final result of the stage two validation coding task was 100% agreement. No additional questions were needed to make the determination of conclusion validity.

Box J.1: Phase Two Hearing Code

Article Log Number	Study Log Number	Type	High Conclusion Validity	Moderate Conclusion Validity	Low Conclusion Validity
1	1	Hearing			T, S, J, G
2	2	Hearing			D, J, S, G
20	86	Hearing			D, J, T, G
20	87	Hearing			D, J, T, G
21	16	Hearing			D, T, S, G
22	17	Hearing			J, T, S, G
38	28	Hearing			D, J, S, G
39	29	Hearing		J, T, D, G	
40	30	Hearing			D, T, S, G
41	31	Hearing			J, T, S, G

Phase Three

The purpose of phase three is to validate the conclusion validity assessment for the substantive concern of hearing. Given the high interrater agreement in the phase two results, this process was achieved in one stage.

Stage One. Using the next 13 of 23 studies that addressed the substantive concern of hearing, three coders independently addressed the four questions used to guide the decision about conclusion validity. At the debriefing meeting, the coders shared their findings. After all coders had the opportunity to share, each was asked to rate the study as having high conclusion validity, moderate conclusion validity, or low conclusion validity. The stage one coding results are provided in Box J.2.

Box J.2: Phase Three Hearing Code

Article Log Number	Study Log Number	Type	High Conclusion Validity	Moderate Conclusion Validity	Low Conclusion Validity
55	43	Hearing			J, S, G
56	44	Hearing			J, T, G
57	45	Hearing		D, T, G	
71	54	Hearing			D, S, G
72	55	Hearing			J, S, G
73	56	Hearing			J, T, G
74	57	Hearing			D, T, G
75	58	Hearing			J, T, G
76	59	Hearing			J, S, G
77	99	Hearing			D, S, G
77	100	Hearing			D, S, G
77	101	Hearing			D, S, G
92	69	Hearing			D, T, G

Results of the phase three coding yielded high agreement (100%) between the four coders. As such, no reconciliation was needed to reach agreement between coders.

Phase Four

The purpose of phase four is to validate the conclusion validity assessment for the 23 studies addressing the substantive concern of speech. This validation process was achieved in two stages.

Stage One. Given the training and high rate of agreement between the researchers, three coders were used in this stage. Using the first 10 of 23 studies that addressed the substantive concern of speech, three coders

independently addressed the four questions used to guide the decision about conclusion validity. The stage one coding results for the substantive topic of speech is provided in Box J.3. Results of the first coding yielded high agreement (100%) between the three coders.

Box J.3: Stage One Speech Coding Results

Article Log Number	Study Log Number	Type	High Conclusion Validity	Moderate Conclusion Validity	Low Conclusion Validity
3	3	Speech			J, T, G
4	4	Speech			J, S, G
5	83	Speech		T, D, G	
5	84	Speech		T, D, G	
5	85	Speech		T, D, G	
6	5	Speech			S, G, J
7	6	Speech			T, D, G
8	7	Speech			D, J, G
23	18	Speech		J, D, G	
24	19	Speech			S, T, G

Stage Two. Given the high rate of agreement between the coders, two coders the final 13 studies that addressed the substantive concern on speech were evaluated using two coders. Results of the stage two speech coding task are provided in Box J.4.

Results of the stage two coding task again yielded high agreement (100%) between the two coders. No studies needed to be reconciled in the stage two coding task.

Box J.4: Stage Two Speech Coding Results

Article Log Number	Study Log Number	Type	High Conclusion Validity	Moderate Conclusion Validity	Low Conclusion Validity
25	20	Speech			D, G
26	21	Speech			T, G
42	32	Speech		S, G	
43	33	Speech			J, G
44	34	Speech			D, G
58	46	Speech		T, G	
59	93	Speech		J, G	
59	94	Speech		J, G	
78	102	Speech			J, G
78	103	Speech			J, G
79	60	Speech			D, G
94	70	Speech			T, G
95	71	Speech			J, G

Phase Five

The purpose of phase five is to validate the conclusion validity assessment in the 59 studies addressing the substantive concern of language. This validation process was achieved in two stages.

Stage One. Given the training and high rate of agreement between the coders, three coders were used in this stage. Using the first 20 of 59 studies that addressed the substantive concern of language, three coders independently addressed the four questions used to guide the decision about conclusion validity. Stage one coding results for the substantive topic of language are provided in Box J.5. Results of the first coding yielded high agreement (100%)

Box J.5: Stage One Language Coding Results

Article Log Number	Study Log Number	Type	High Conclusion Validity	Moderate Conclusion Validity	Low Conclusion Validity
11	8	Language			S, J, G
12	9	Language		T, D, G	
13	10	Language	S, J, G		
14	11	Language			T, D, G
16	12	Language			S, J, G
17	13	Language		D, T, G	
18	14	Language			S, J, G
19	15	Language		T, D, G	
30	22	Language		S, J, G	
31	23	Language			D, T, G
32	88	Language			S, J, G
32	89	Language			S, J, G
32	90	Language			S, J, G
33	24	Language			D, T, G
34	25	Language			S, J, G
35	26	Language			D, T, G
36	27	Language	T, J, G		
37	91	Language			T, D, G
37	92	Language			T, D, G
46	35	Language		S, J, G	

between the three coders. No studies needed to be reconciled in the stage one language coding task.

Stage Two. Given the high rate of agreement between the researchers, two coders were used in the stage two coding task of the final 39 of 59 studies that addressed the substantive concern of language. Results of the stage two coding task are presented in Box J.6.

Box J.6: Stage Two Language Coding Results

Article Log Number	Study Log Number	Type	High Conclusion Validity	Moderate Conclusion Validity	Low Conclusion Validity
47	36	Language			D, G
48	37	Language			D, G
49	38	Language			D, G
50	39	Language			D, G
51	40	Language			D, G
52	41	Language			D, G
53	42	Language			D, G
60	47	Language			D, G
61	95	Language		D, G	
61	96	Language			D, G
62	97	Language			J, G
62	98	Language			J, G
63	48	Language	J, G		
64	49	Language			J, G
65	50	Language			J, G
66	51	Language			J, G
67	52	Language		J, G	
68	53	Language			J, G
81	61	Language			J, G
82	62	Language			J, G
83	63	Language			S, G
84	64	Language			S, G
85	65	Language			S, G
86	66	Language			S, G
87	67	Language			S, G
88	68	Language			S, G
96	72	Language			S, G
97	73	Language			D, G
98	74	Language	J, G		
99	75	Language		T, G	
100	76	Language		T, G	
101	77	Language		T, G	
102	78	Language		T, G	
103	79	Language			T, G
104	104	Language			T, G
104	105	Language			T, G
105	80	Language		T, G	
106	81	Language			T, G
107	82	Language			J, G

Results of the stage two coding task again yielded high agreement (100%) between three coders. No studies needed to be reconciled in the stage two language coding task.

Summary

This phase provides a summative report of both the stage one and stage two coding tasks across the substantive concerns of hearing, speech, and language.

In the conclusion validity assessment coding task, the coders independently classified the 105 studies in the research synthesis study population. A high level of agreement (100%) was met on 105 of the total 105 studies. This yielded an interrater reliability of 100%.

Four studies (four of 105 studies or 3.8%) were rated as having high conclusion validity. Moderate conclusion validity were rated in 21 studies (22 of 105 studies or 20.9%). The vast majority (79 of 105 studies or 75.3%) of the research synthesis study population were rated as having low conclusion validity.

More in-depth discussions about conclusion validity assessment findings will be in Chapter IV which focuses on hearing, Chapter V which looks at speech, Chapter VI which targets language, and Chapter VII which evaluates trends across the three topics.

Box J.7: Reconciled Language Codes

Article Log Number	Study Log Number	Type	High Conclusion Validity	Moderate Conclusion Validity	Low Conclusion Validity
1	1	Hearing			T, S, J, G
2	2	Hearing			D, J, S, G
20	86	Hearing			D, J, T, G
20	87	Hearing			D, J, T, G
21	16	Hearing			D, T, S, G
22	17	Hearing			J, T, S, G
38	28	Hearing			D, J, S, G
39	29	Hearing		J, T, D, G	
40	30	Hearing			D, T, S, G
41	31	Hearing			J, T, S, G
55	43	Hearing			J, S, G
56	44	Hearing			J, T, G
57	45	Hearing		D, T, G	
71	54	Hearing			D, S, G
72	55	Hearing			J, S, G
73	56	Hearing			J, T, G
74	57	Hearing			D, T, G
75	58	Hearing			J, T, G
76	59	Hearing			J, S, G
77	99	Hearing			D, S, G
77	100	Hearing			D, S, G
77	101	Hearing			D, S, G
92	69	Hearing			D, T, G
3	3	Speech			J, T, G
4	4	Speech			J, S, G
5	83	Speech		T, D, G	
5	84	Speech		T, D, G	
5	85	Speech		T, D, G	
6	5	Speech			S, G, J
7	6	Speech			T, D, G
8	7	Speech			D, J, G
23	18	Speech		J, D, G	
24	19	Speech			S, T, G
25	20	Speech			D, G
26	21	Speech			T, G
42	32	Speech		S, G	
43	33	Speech			J, G
44	34	Speech			D, G
58	46	Speech		T, G	
59	93	Speech		J, G	

Box J.7: Reconciled Language Codes (Continued)

Article Log Number	Study Log Number	Type	High Conclusion Validity	Moderate Conclusion Validity	Low Conclusion Validity
59	94	Speech		J, G	
78	102	Speech			J, G
78	103	Speech			J, G
79	60	Speech			D, G
94	70	Speech			T, G
95	71	Speech			J, G
11	8	Language			S, J, G
12	9	Language		T, D, G	
13	10	Language	S, J, G		
14	11	Language			T, D, G
16	12	Language			S, J, G
17	13	Language		D, T, G	
18	14	Language			S, J, G
19	15	Language		T, D, G	
30	22	Language		S, J, G	
31	23	Language			D, T, G
32	88	Language			S, J, G
32	89	Language			S, J, G
32	90	Language			S, J, G
33	24	Language			D, T, G
34	25	Language			S, J, G
35	26	Language			D, T, G
36	27	Language	T, J, G		
37	91	Language			T, D, G
37	92	Language			T, D, G
46	35	Language		S, J, G	
47	36	Language			D, G
48	37	Language			D, G
49	38	Language			D, G
50	39	Language			D, G
51	40	Language			D, G
52	41	Language			D, G
53	42	Language			D, G
60	47	Language			D, G
61	95	Language		D, G	
61	96	Language			D, G
62	97	Language			J, G
62	98	Language			J, G
63	48	Language	J, G		
64	49	Language			J, G

Box J.7: Reconciled Language Codes (Continued)

Article Log Number	Study Log Number	Type	High Conclusion Validity	Moderate Conclusion Validity	Low Conclusion Validity
65	50	Language			J, G
66	51	Language			J, G
67	52	Language		J, G	
68	53	Language			J, G
81	61	Language			J, G
82	62	Language			J, G
83	63	Language			S, G
84	64	Language			S, G
85	65	Language			S, G
86	66	Language			S, G
87	67	Language			S, G
88	68	Language			S, G
96	72	Language			S, G
97	73	Language			D, G
98	74	Language	J, G		
99	75	Language		T, G	
100	76	Language		T, G	
101	77	Language		T, G	
102	78	Language		T, G	
103	79	Language			T, G
104	104	Language			T, G
104	105	Language			T, G
105	80	Language		T, G	
106	81	Language			T, G
107	82	Language			J, G

APPENDIX K

PROCEDURES FOR IDENTIFYING SUBSTANTIVE TOPICS IN STUDIES

Elaborated below are the procedures used to identify the substantive topics addressed in studies in the research synthesis population. As multiple concepts can be included in a single study, the classification system of substantive topics was not designed to be mutually exclusive or exhaustive. Rather, the substantive topic classification system was designed to capture the intent of the study. As such, the classification system was developed in three steps using the *Journal of Speech, Language, and Hearing Research* (*JSLHR*), key words identified in the abstract, and the training and experience of this author.

The initial step to develop the classification system of substantive topics involved using the information provided by the *JSLHR*. This information is elaborated in Phase Two and broadly set forth in Table 3.3. Using the three broad categories in the classification used by the *JSLHR*, the 105 studies in the research synthesis population were grouped under one of the three substantive topics of hearing, speech, or language. These three broad categories captured all of the 105 studies.

The second step in developing the classification accessed the key words as designated by the author in the article abstract. Using the article narrative to provide clarity the key concepts for substantive concerns for the studies were identified.

The final step in developing the substantive concerns of studies hinged on the author's expertise. With over 30 years of experience in the field of

speech-language pathology, the author made the determination of what substantive concern would capture each study. The substantive concerns identified for each of the three broad categories were adequate to capture all of the 105 studies in the research synthesis population.

The substantive concerns for hearing are identified in Table 4.1.

The substantive concerns for speech are identified in Table 5.1.

The substantive concerns for language are identified in Table 6.1.

APPENDIX L

COUNTRY WHERE STUDY WAS CONDUCTED

Detailed below are the procedures used to gather information about the country of origin of the research synthesis article population. Two coders independently accomplished this coding in three distinct steps. Box L.1 provides broad results of the coding process.

The first step centered on using the description and information contained in the Methods section of each article to identify the country of origin. Based on the information contained in the Methods section, each article was coded in one of two ways. If the country where the study was conducted could be identified, the article was coded as containing adequate information to make a determination. The country of origin was identified in half (46 out of 92) of the articles. The remaining 46 articles in which the country of origin could not be established were coded as not containing adequate information and retained for use in the second step in the coding process.

The second step in determining the country of origin used the 46 retained articles from step one. Because adequate information was not provided in the text of the articles, coders relied upon author information to make the step two determination. The *Journal of Speech, Language, and Hearing Research* (*JSLHR*) provides information about all authors' employment and/or research affiliation and the location of each facility. Using the author information, the coders were able to establish the country of origin on the remaining 46 articles.

Steps one and two were adequate to establish a country of origin for all (100%) of the 92 articles contained in the research synthesis population. All 92 articles were used in the third step of this coding procedure.

The third step involved documenting the country decision from steps one and two. Each coder documented the country of origin of each of the articles. Results show that the two coders were in 100 percent agreement on the country of origin.

Box L.1: Country Where Study was Conducted

	Determined by			
Article Number	Description provided in the article		Author information	Country
	Yes	No	Yes	
1		*	*	United States
2		*	*	United States
3		*	*	United States
4	*			United States
5	*			United States
6	*			United States
7		*	*	United States
8		*	*	United States
11		*	*	United States
12		*	*	United States
13	*			United Kingdom
14		*	*	Finland
16	*			Belgium
17	*			United Kingdom
18	*			United States
19	*			United Kingdom
20		*	*	United States
21	*			United States
22		*	*	United States
23	*			United States
24	*			United States
25	*			United States
26		*	*	United States
30		*	*	South Africa
31	*			United Kingdom
32	*			United States
33	*			United States
34		*	*	United States
35	*			Australia
36	*			United States
37	*			United States
38		*	*	United States
39	*			United States
40		*	*	United States
41	*			United Kingdom
42	*			United States

Box L.1: Country Where Study was Conducted (Continued)

	Determined by			
Article Number	Description provided in the article		Author information	Country
	Yes	No	Yes	
43	*			United States
44		*	*	United States
46	*			United States
47		*	*	United States
48		*	*	United States
49		*	*	United States
50		*	*	United States
51	*			United States
52		*	*	United States
53		*	*	Israel
55	*			United Kingdom
56	*			Australia
57	*			Australia
58		*	*	United States
59		*	*	United States
60		*	*	United States
61		*	*	United States
62	*			Israel
63		*	*	United States
64	*			United States
65		*	*	United States
66		*	*	United States
67	*			United States
68	*			United Kingdom
71		*	*	United States
72	*			Israel
73		*	*	Israel
74	*			United States
75		*	*	United States
76		*	*	United Kingdom
77	*			The Netherlands
78	*			United States
79		*	*	France
81		*	*	United States
82	*			United States
83		*	*	United States
84		*	*	United States
85	*			United States
86		*	*	Hong Kong, SAR
87	*			United Kingdom
88		*	*	United States
92	*			United States
94		*	*	United States
95	*			United States
96	*			United States
97	*			United States
98		*	*	United States
99		*	*	United Kingdom
100		*	*	United Kingdom
101	*			United Kingdom
102		*	*	United States
103		*	*	United States
104	*			Hong Kong, SAR
105	*			The Netherlands
106	*			United States
107		*	*	Hong Kong, SAR
Total	46	46	46	

APPENDIX M

FORMAT OF ARTICLES

The *JSLHR* publishes a variety of papers including data-based research reports, reviews, and tutorials and each submission is required to contain an abstract. The Information for Authors published in each *JSLHR* provides guidelines for the subsections of the abstract. The journal requires that the abstract contain four subsections: Purpose, Method, Results, and Conclusions. Information about what is to be included in each of the four subsections is provided. However, the *JSLHR* does not provide any explicit recommendations for submitted manuscripts.

Because specific guidelines for the organization of manuscripts are not provided by the journal, the general assumption is that the manuscripts would follow the same organizational framework designated for the abstract. A review of the organization of the articles found that 91 of the 92 retained articles (98.9%) included an Introduction section. The articles began with an unlabeled introductory part and included subsections that provided background rationale for the study, the literature review, definitions of key vocabulary and concepts, and the purpose of the study including research questions or hypotheses. The exception (#31) began with a section heading.

For articles involving a single study subsequent major sections are delineated by a bold line the length of the section title and the title of the section. The format of the ten articles that included more than one study varied. If a section pertained to the entire article, the bold line extended across the entire column. The shorter line and titles were used to organize the individual studies.

In those ten articles, the individual studies contained their own Methods, Results, and Discussion sections. The authors then engaged in a General Discussion or Conclusion section involving the results of all the studies in the article. These ten articles are not included in the remaining descriptions of the articles of single studies.

The Methods section included in each of the 82 single study articles included information about participants, procedures, research design, materials, interrater reliability, and data analysis. Only one article (#97) referred to “subjects” as opposed to “participants”.

The Results section reported and discussed the findings, variability, between-group effects, within-group effects, gender differences, and relationships. The Discussion section provided concluding remarks, interpretation of findings, implications of the findings, and limitations of the study. One article (#74) combined the Results and Discussion sections under a joint heading and then provided a separate Summary and Conclusion section.

Following the typical Discussion section other major sections were included in 18 articles. These sections included were designated as Summary, Conclusion, General Discussion, Clinical Implications, and Conclusions and Further Implications.

Box M.1: Article Format

Article Number	Study Number	Unlabeled Introduction	Method	Results	Discussion	Summary	Conclusion	General Discussion	Clinical Implication	Conclusion and Further Implications	Results and Discussions	Summary and Conclusions	Overview	Vowel Productivity During Vocal Tract Growth
1	1	✓	✓	✓	✓	✓								
2	2	✓	✓	✓	✓		✓							
3	3	✓	✓	✓	✓									
4	4	✓	✓	✓	✓		✓							
5		✓												
	83		✓	✓										
	84		✓	✓										
	85		✓	✓										
6	5	✓	✓	✓	✓			✓						
7	6	✓	✓	✓	✓									
8	7	✓	✓	✓	✓									
11	8	✓	✓	✓	✓									
12	9	✓	✓	✓	✓									
13	10	✓	✓	✓	✓									
14	11	✓	✓	✓	✓									
16	12	✓	✓	✓	✓									
17	13	✓	✓	✓	✓									
18	14	✓	✓	✓	✓				✓					
19	15	✓	✓	✓	✓					✓				
20		✓			✓									
	86		✓	✓										
	87		✓	✓										
21	16	✓	✓	✓	✓	✓								
22	17	✓	✓	✓	✓		✓							
23	18	✓	✓	✓	✓									
24	19	✓	✓	✓	✓									
25	20	✓	✓	✓	✓									
26	21	✓	✓	✓	✓									
30	22	✓	✓	✓	✓									
31	23		✓	✓	✓									
32		✓						✓						
	88		✓											
	89		✓											
	90		✓											
33	24	✓	✓	✓	✓									
34	25	✓	✓	✓	✓									
35	26	✓	✓	✓	✓									
36	27	✓	✓	✓	✓									
37		✓					✓	✓						
	91		✓											
	92		✓											
38	28	✓	✓	✓	✓									
39	29	✓	✓	✓	✓									
40	30	✓	✓	✓	✓									
41	31	✓	✓	✓	✓									
42	32	✓	✓	✓	✓									

Box M.1: Article Format (Continued)

Article Number	Study Number	Unlabeled Introduction	Method	Results	Discussion	Summary	Conclusion	General Discussion	Clinical Implication	Conclusion and Further Implications	Results and Discussions	Summary and Conclusions	Overview	Vowel Productivity During Vocal Tract Growth
43	33	✓	✓	✓	✓		✓							
44	34	✓	✓	✓	✓	✓								
46	35	✓	✓	✓	✓									
47	36	✓	✓	✓	✓		✓							
48	37	✓	✓	✓				✓						
49	38	✓	✓	✓	✓									
50	39	✓	✓	✓	✓									
51	40	✓	✓	✓	✓									
52	41	✓	✓	✓	✓									
53	42	✓	✓	✓	✓									
55	43	✓	✓	✓	✓									
56	44	✓	✓	✓	✓		✓							
57	45	✓	✓	✓	✓			✓						
58	46	✓	✓	✓	✓									
59		✓									✓			
	93		✓	✓	✓									
	94		✓											
60	47	✓	✓	✓	✓									
61		✓												
	95		✓								✓			
	96		✓								✓			
62		✓						✓						
	97		✓	✓	✓									
	98		✓	✓	✓									
63	48	✓	✓	✓	✓									
64	49	✓	✓	✓	✓									
65	50	✓	✓	✓	✓									
66	51	✓	✓	✓	✓	✓								
67	52	✓	✓	✓	✓									
68	53	✓	✓	✓	✓									
71	54	✓	✓	✓	✓	✓								
72	55	✓	✓	✓	✓									
73	56	✓	✓	✓	✓									
74	57	✓	✓								✓	✓		
75	58	✓	✓	✓	✓	✓								
76	59	✓	✓	✓	✓		✓							
77		✓						✓				✓		
	99	✓	✓	✓									✓	
	100	✓	✓	✓									✓	
	101	✓	✓	✓									✓	
78		✓						✓						
	102	✓	✓	✓										
	103	✓	✓	✓										
79	60	✓	✓	✓	✓		✓							✓
81	61	✓	✓	✓	✓									
82	62	✓	✓	✓	✓									

Box M.1: Article Format (Continued)

Article Number	Study Number	Unlabeled Introduction	Method	Results	Discussion	Summary	Conclusion	General Discussion	Clinical Implication	Conclusion and Further Implications	Results and Discussions	Summary and Conclusions	Overview	Vowel Productivity During Vocal Tract Growth
83	63	✓	✓	✓	✓									
84	64	✓	✓	✓	✓									
85	65	✓	✓	✓	✓									
86	66	✓	✓	✓	✓									
87	67	✓	✓	✓	✓									
88	68	✓	✓	✓	✓		✓							
92	69	✓	✓	✓	✓									
94	70	✓	✓	✓	✓									
95	71	✓	✓	✓	✓									
96	72	✓	✓	✓	✓									
97	73	✓	✓	✓	✓									
98	74	✓	✓	✓	✓									
99	75	✓	✓	✓	✓									
100	76	✓	✓	✓	✓									
101	77	✓	✓	✓	✓									
102	78	✓	✓	✓	✓									
103	79	✓	✓	✓	✓									
104		✓					✓							
	104		✓	✓	✓									
	105		✓	✓	✓									
105	80	✓	✓	✓	✓									
106	81	✓	✓	✓	✓									
107	82	✓	✓	✓	✓									

APPENDIX N

INACCURACIES AND AMBIGUITIES

Elaborated below are numerous items related to inaccuracies and ambiguities in articles published in Volume 47 of the *Journal of Speech, Language, and Hearing Research* (*JSLHR*) published in 2004. These items were noted while reading the research synthesis population of articles. In themselves, these inaccuracies and ambiguities provide supporting evidence about low population validity and conclusion validity ratings that were achieved.

Box N.1: Inaccuracies and Ambiguities

Article	Page	Findings
1		Table 1 provided information about individuals with cochlear implants, but no comparison information was provided for the individuals with normal hearing.
3	23	Used a five-point ordinal scale but scale was not defined.
	24	In Table 1 the types and numbers of injections (bilateral versus unilateral) do not agree with what is provided in the article.
	24	There is no variability between participants' pre-injection/post-injection severity. Pre-injection data has eight ties and post-injection data has five ties.
	24	No data provided for comparison group.
	24	Only ten participants covered an age span of 36 years (38-74 years of age).
	26	Provides means and standard deviations but no information provided about whether the data are normally distributed.
	28	Participant selection was blatantly biased. "A second criterion for speaker selection was a documented pre-injection-to-post-injection perceptual rating change indicating that each speaker responded favorably to Botox."
4	33	Abstract describes healthy young men included in the study as being age 29 ± 3 years. Methods section describes healthy young men included in the study as being between 25-35 years of age.

Box N.1: Inaccuracies and Ambiguities (Continued)

Article	Page	Findings
	33	Abstract describes healthy older men included in the study as being 69 ± 7 years. Methods section describes healthy older men included in the study as being between 60-83 years of age.
	35	Table 1 presents demographic and neuropsychological assessment data on the two groups in this study. All participants were right handed. Of interest is that the Group: Young reported a handedness mean of 21.45, standard deviation of 2.74, and a range of 17-24. Group: Old reported a handedness mean of 22.72, standard deviation of 1.93, and a range of 18-24. No additional explanations were provided in the text.
	40	Data in Table 6 is incorrect. Columns total more than 100%. N values are missing.
5	54	The articles states that “statistically significant differences in the effects of the different training tasks were found in left HG and in the auditory cortical areas considered as a whole.” However, no data are provided.
8	87	The participants in the study were 15 adult male speakers. A mean age of 24.7 years and a standard deviation of 12.5 years were given. However, one standard deviation below the mean provides an age of 12.2 years which does not qualify as an adult.
	91	Refers to “differences in the number of observations” but does not discuss why there were different numbers of observations.
	92-93	The order of the five dependent variables in Table 2 differs from the order of the graphic description of the data. This is confusing for the reader.
8	93	Table 2 reports 85 *f* tests (17 across five variables)
	95	The scales on the y-axis differ
	97	Table 4 reports Pearson correlations for 30 task pairs across four variables (120 *r* values) for an N of 15.
13	148	The only criteria given for the selection of participant was the enrollment in a language unit. However, there was no discussion about the criteria for enrollment in the unit.

Box N.1: Inaccuracies and Ambiguities (Continued)

Article	Page	Findings
14	164	"The control group purposely represented a wider range of ages to account for the wide age range seen in the three generations of the SLI family." The ages of participants in the SLI group (n=5) were between 11 years and 70 years with a range of 56 years. The ages of participants in the comparison group (n=6) were between 15 years and 61 years with a range of 46 years.
	164	No discussion of where the comparison group came from.
	165	The grandmother in the SLI groups was eliminated from the study, but this does not seem to be reflected in the data analysis.
21	260	Table 1 provides pure-tone threashold information for older listeners and averages for older listeners and younger listeners. The written report states that threasholds for younger listeners are also provided.
	260	The text states, "For slightly more than half of the older listeners…". How many is this?
22	271	Researchers attempted to exclude left-handed individuals "as much as possible". However, *one* left-handed person was included in the study.
24	299-300	Ns in Table 2 and Table 3 differ across all three groups
34	428	Outliers were mentioned but unclear how they were handled in the data analysis
35	440	Sample sizes were n_{speech}=12 and $n_{users\ of\ AAC}$=28
	443	Researchers used n=34 but does not explain where the number came from.
37	466	"Participants with lesser amounts of otitis media were selected pseudorandomly for a study of the association between otitis media and developmental outcome." Pseudorandomly is not defined.
41	529	Article discusses the following sample sizes: Original sample: n_{SLI} = 11 $n_{control}$ = 13 Added "to boost power" 5 3 Total sample size = 16 = 16

Box N.1: Inaccuracies and Ambiguities (Continued)

Article	Page	Findings
	530	Participants were administered tests that were not standardized for their age. Scores for 17-21 year olds on the Standard Progressive Matrices (Raven, Raven & Court, 1998) were calculated on the norms for 16 year-olds.
	530	Participants were administered tests that were not standardized for their age. Scores for 18-21 year olds on the British Picture Vocabulary Scale (Dunn, Dunn, Whetton, & Pintilie, 1982) were calculated on the norms for 17 year-olds.
	530	Participants were administered tests that were not standardized for their age. Scores for 17-21 year olds on the Clinical Evaluation of Language Fundamentals-Revised (Semel, Wiig, & Secord, 1987) were calculated on the norms for 16 year-olds.
	530	Participants were administered tests that were not standardized for their age. Scores for 19-21 year olds on Recreating Sentences subtest of the Test of Language Competence-Expanded Edition (Wiig & Secord, 1989) were calculated on the norms for 18 year-olds.
	530	Participants were administered tests that were not standardized for their age. Scores for 19-21 year olds on Figurative Language subtest of the Test of Language Competence-Expanded Edition (Wiig & Secord, 1989) were calculated on the norms for 18 year-olds.
	531	Participants were administered tests that were not standardized for their age. Scores for 17-21 year olds on the Martin and Pratt Nonword Reading Test (Martin & Pratt, 1999) were calculated on the norms for 16;0 to 16;11 year-olds.
	531	Participants were administered tests that were not standardized for their age. Scores for 15-21 year olds on the Naming Speed Test for Pictures were calculated on the norms for 14;11 year-olds.
47	617	Reported 39 t-tests
48	624	Authors claim that this is a retrospective study but do not provide any support for this claim.
49	642	Table 1 does not provide n for groups
53	697-698	Table 1 and Table 2 do not provide n for either group.

Box N.1: Inaccuracies and Ambiguities (Continued)

Article	Page	Findings
55	727	Authors used a scale of word familiarity of 1 to 7 but do not define the high end and low end of the scale.
58	768	Authors report that participants were part of a previous study but it is not clear if any of the data from the previous study were used in this study.
74	1003	Clearer definitions of terms are needed. For example, the authors attempted to define individuals who wear hearing aids as *full-time users* and *part-time users*. Background information and an example were provided. However, confusion abounds with this definition-- "Thus, the categories were defined as follows: Full-time users wore their hearing aids whenever they needed them. Part-time users wore their hearing aids only occasionally."
76	1023	Auditory processing efficiency was the topic of interest with this study. It is interesting though that the only inclusion criterion was that the participants/guardians of participants *reported* no previous hearing difficulties.
88	1198	Table 1 presents test performance and production probe accuracy. Means and standard deviations are given for variables of interest for each of the three groups involved in the study. Gender was a variable. For each group, the number of the boys were listed as the mean and the number of the girls were listed as the standard deviation.
105	1415	The study includes a normative population of 500 children. However, there is no mention in the article where this population came from, who this population is, when the data were collected, or what kind of data were collected.

VITA

G L E N D A E L K I N S B Y R N S

4225 TAMU · Texas A&M University · College Station, TX 77843

EDUCATION

Ph. D.: Texas A&M University
(Educational Psychology, May 2007)

M.Ed.: Southwest Texas State University
(Speech Pathology, May 1976)

B.S.: Southwest Texas State University
(Education, 1975)

CERTIFICATIONS

American Speech-Language-Hearing Association: Certificate of Clinical Competency

Texas State Teaching Certificate

SELECTED PROFESSIONAL EXPERIENCE

Lecturer	Texas A&M University	College Station, TX	2005-Present
Research Assistant	Texas A&M University	College Station, TX	2005-Present
Speech Pathologist	1976-2005		

PUBLICATIONS

Parker, R., Byrns, G., and Cryer, J. (in press). Controlling baseline trend in single case research. *School Psychology Quarterly.*

Co-authored curriculum, *Comprehensive Curriculum and Assessment for Special Students*, published by G.G. Consulting

Co-authored eligibility criteria for speech and language. TSHA Task Force on Eligibility