A Reliability Generalization of the Parental Authority Questionnaire

Lynn M. Dean

Follow this and additional works at: http://digscholarship.unco.edu/dissertations

Recommended Citation
Dean, Lynn M., "A Reliability Generalization of the Parental Authority Questionnaire" (2016). Dissertations. 396.
http://digscholarship.unco.edu/dissertations/396

This Text is brought to you for free and open access by the Student Research at Scholarship & Creative Works @ Digital UNC. It has been accepted for inclusion in Dissertations by an authorized administrator of Scholarship & Creative Works @ Digital UNC. For more information, please contact Jane.Monson@unco.edu.
A RELIABILITY GENERALIZATION OF THE PARENTAL AUTHORITY QUESTIONNAIRE

A Dissertation Submitted in Partial Fulfillment of the Requirements of the Degree of Doctor of Philosophy

Lynn M. Dean

College of Education and Behavioral Sciences
School of Psychological Sciences
Educational Psychology

December 2016
This Dissertation by: Lynn M. Dean

Entitled: *A Reliability Generalization of the Parental Authority Questionnaire*

has been approved as meeting the requirement for the Degree of Doctor of Philosophy in College of Education and Behavioral Sciences in School of Psychological Sciences, Program of Educational Psychology

Accepted by the Doctoral Committee

______________________________
Steven Pulos, Ph.D., Research Advisor

______________________________
Teresa McDevitt, Ph.D., Committee Member

______________________________
Nancy Karlin, Ph.D., Committee Member

______________________________
Susan Thompson, Ph.D., Faculty Representative

Date of Dissertation Defense ________________________________

Accepted by the Graduate School

______________________________
Linda L. Black, Ed.D.
Associate Provost and Dean
Graduate School and International Admissions
ABSTRACT

Dean, Lynn M. A Reliability Generalization of the Parental Authority Questionnaire. Published Doctor of Philosophy dissertation, University of Northern Colorado, 2016.

How parents interact with their children impacts many crucial facets of children’s lives. Over the last 4 decades, researchers have identified four different parenting styles: authoritative, authoritarian, permissive, and disengaged. Hundreds of studies conducted all over the world, have identified correlations between parenting style and many different child outcomes (e.g., academic achievement, coping, self-esteem, and decision-making abilities). The interpretability of these results has been limited by the unknown psychometric properties of the instruments used.

The building block of any psychometrically sound measure is instrument reliability, which is the degree of non-error variance in the instrument’s results. Sound estimates of reliability may be obtained from the synthesis of existing research. To this end, Reliability generalization (RG) focuses on the estimate of reliability reported by primary studies to produce a measure of central tendency of the reliability of the data collected by the instrument across different studies and the impact of identified moderator variables on the magnitude and distribution of reliability estimates. This reliability generalization dissected the most commonly used measure to identify parenting style--the Parental Authority Questionnaire (PAQ; Buri, 1991). Two different statistical methods
for determining the reliability of the PAQ were compared: the random-effect model and the varying coefficient model.

In this investigation, all six scales of the PAQ were found to be reliable at a level generally acceptable for research purposes, but not necessarily reliable to the degree necessary when making decisions that affect individuals. Moreover, each of the six scales was not equally reliable. Seven moderator variables were identified that significantly affected scale reliability; (a) scale mean, (b) scale variance, (c) percent of male participants, (d) developmental level of participants, (e) the language the PAQ was administered in, (f) the type of publication, and (g) the year of publication.

A comparison of the two different RG meta-analytic methods revealed that both models yielded similar results with regard to scale reliabilities. However, the varying-coefficient model provided smaller confidence intervals and was more sensitive at detecting moderator variables. This dissertation has implications for both those who intend to use the PAQ as well as those who conduct reliability generalizations. First, before researchers plan interventions based on the results of studies that seem to correlate parental authority with various other psychological, sociological, or educational constructs, care must be taken to ensure that sound methodological practices are in place throughout the research process. These practices should include calculating reliability each time an instrument such as the Parental Authority Questionnaire is administered. Second, this research yielded valuable information for meta-analytic researchers by demonstrating, with actual study data, how methods of analysis in reliability generalization could differ in not only identifying reliability estimates but variables that moderate reliability as well.
ACKNOWLEDGEMENTS

I would like to take this opportunity to thank each member of my committee, especially my two mentors who guided me, not only in the dissertation process, but also throughout my entire doctoral program: Steven Pulos and Theresa McDevitt. I would also like to thank my husband and three grown children who have been extremely supportive during these past few years, each of whom has taken up the slack in familial responsibilities at one time or another. I would especially like to thank Marie for allowing me to attend her chosen University and for all those “payback” lunches she delivered to me. It was she, after all, who inadvertently got me into this in the first place. That said, it was not my plan to complete this degree, but His. What He has in store for me next is anyone’s guess. Whatever it is, life certainly won’t be boring.
TABLE OF CONTENTS

CHAPTER
I. INTRODUCTION ........................................................................................................... 1
   Purpose ........................................................................................................................... 5
   Rationale ......................................................................................................................... 6
   Research Questions ....................................................................................................... 6
   Delimitations .................................................................................................................. 7
   Definitions ....................................................................................................................... 7

II. LITERATURE REVIEW ............................................................................................... 9
   Baumrind’s Two-factor Model ...................................................................................... 9
   Subsequent Research by Others .................................................................................. 16
      Objective versus Subjective Perspective ................................................................. 18
      Parent versus Child Perceptions .............................................................................. 19
      Gender Differences ................................................................................................. 21
      Other Characteristics of the Respondent .................................................................. 22
      Culture and Social Milieu ....................................................................................... 23
      Direct Observation and Its Foibles ......................................................................... 25
      Construct Validity ..................................................................................................... 26
      Instrument Reliability ............................................................................................... 27
   A Micro Look at the Parental Authority Questionnaire (PAQ) ................................... 28
   Reliability and the Parental Authority Questionnaire ............................................... 29
   Reliability Generalization ......................................................................................... 31

III. METHODS .................................................................................................................. 33
   Literature Search ......................................................................................................... 33
   Selection of Studies ...................................................................................................... 34
   Institutional Review Board Approval ......................................................................... 34
   Data Collected ............................................................................................................. 34
CHAPTER III. continued

Data Analysis ........................................................................................................35

Meta-analytic Model ............................................................................................35

     Fixed-effects model .......................................................................................35
     Random-effects (RE) model ............................................................................36
     Varying coefficient (VC) model .....................................................................37

To Transform or Not to Transform ....................................................................37
To Weight or Not to Weight ................................................................................38
Steps in the Reliability Generalization Meta-analytic for the
     Random-effects Model ................................................................................39
Steps in the Varying Coefficients Alpha Model ..............................................41
Comparing the Results of the Two Different Models .....................................42

IV. RESULTS .......................................................................................................43

Data Extraction ..................................................................................................43
Random-effect Model Results ...........................................................................45
Moderator Analysis ..............................................................................................46
Test of Scale Mean as Moderator Variable ......................................................47
Test of Scale Variance as a Moderator Variable ..............................................48
Test of Percent Male as a Moderator Variable ................................................49
Test of Language of Instrument as Moderator Variable ...............................50
Test of Developmental Level as a Moderator Variable .................................53
Test of Publication Type as a Moderator Variable .........................................55
Test of Publication Year as a Moderator Variable .........................................57
The Varying Coefficient Model Results ............................................................57
Analysis of Moderator Variables .......................................................................59

     Continuous Variables .....................................................................................59
     Test of Scale Mean as a Moderator Variable .................................................60
     Test of Scale Variance as a Moderator Variable ..........................................61
     Test of Percent Male as a Moderator Variable .............................................62

     Categorical Moderator Variables .................................................................63

     Test of Instrument Language as a Moderator Variable ...............................63
     Test of Developmental Level as a Moderator Variable ...............................64
     Test of Type of Publication as a Moderator Variable ..................................67
     Test of Year of Publication as a Moderator Variable ...................................67
CHAPTER V. CONCLUSIONS.................................................................................................70

Methodological Mistakes...............................................................................................71

Issues with Reliability....................................................................................................72
Reliability Induction .........................................................................................................72
Other Problems with Reliability.......................................................................................73
Issues with Inconsistent Scale Construction and Language ........................................74
Other Methodological Issues ..........................................................................................76

Methodological and Reliability Issues Inherent in the Studies
Included in the Reliability Generalization .......................................................................77

Reliability of Scales and Moderator Effects ..................................................................77
Comparison of the Random-effect and Varying Coefficient Model ...............................78

Moderator Effects ...........................................................................................................80

Moderating Effect of Scale Mean on Reliability ............................................................80
Moderating Effect of Scale Variance on Reliability .......................................................81
Moderating Effects of Percent Male on Scale Reliabilities ...........................................82
Moderating Effects of the Language the Instrument .....................................................83
Moderating Effect of Developmental Level ...................................................................85
Moderating Effect of Publication Type .........................................................................86
Moderating Effect of Year of Publication .....................................................................87

Limitations .....................................................................................................................88
Implications of Results of this Meta-Analysis ...............................................................89
Areas of Future Research ..............................................................................................90

REFERENCES ..................................................................................................................91

APPENDIX

A. Institutional Review Board .......................................................................................108

B. Meta-analysis Reference List ...................................................................................110
LIST OF TABLES

Table
1. Random-effect Model Reliability Estimates by Scale .........................46
2. Number of Studies Reporting Each of the Continuous Variables
   Considered ............................................................................................................47
3. Random-effect Model Test of Scale Mean as Moderator Variable ........48
4. Random-effect Model Test of Scale Variance as Moderator Variable.....49
5. Random-effect Model Test of Percent Male as Moderator Variable.......50
6. Random-effect Model Test of Instrument Language as a Moderator
   Variable .................................................................................................................52
7. Random-effect Model Test of Developmental Level as a Moderator
   Variable .................................................................................................................54
8. Random-effect Model Test of Publication Type as a Moderator
   Variable .................................................................................................................56
9. Random-effect Model Test of Year Published as a Moderator
   Variable .................................................................................................................58
10. Varying Coefficient Model Reliability Estimates by Scale ..................59
11. Varying Coefficient Model with Scale Mean as a Moderator
    Variable ...............................................................................................................61
12. Varying Coefficient Model with Scale Variance as a Moderator
    Variable ...............................................................................................................62
13. Varying Coefficient Model with Percent Male as a Moderator
    Variable ...............................................................................................................63
14. Varying Coefficient Model with Instrument Language as a Moderator
    Variable ...............................................................................................................65
<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>15.</td>
<td>Varying Coefficient Model with Developmental Level as a Moderator Variable</td>
<td>66</td>
</tr>
<tr>
<td>16.</td>
<td>Varying Coefficient Model with Publication Type as a Moderator Variable</td>
<td>68</td>
</tr>
<tr>
<td>17.</td>
<td>Varying Coefficient Model with Publication Year as a Moderator Variable</td>
<td>69</td>
</tr>
<tr>
<td>18.</td>
<td>Comparison of Reliability Estimates by Model</td>
<td>79</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Figure
1. Baumrind’s Theoretical Parental Authority Circumplex Model ............11
2. A Theoretical Circumplex illustrating Baumrind’s revised Parental Authority Prototypes..............................................................16
CHAPTER I

INTRODUCTION

Over the last few decades, mothers and fathers have become confused about the best ways to parent their children. According to Criss and Larzelere (2013), “Part of this confusion can be attributed to the wide range of parenting advice in popular books, which often have reflected transient fads more than scientific evidence” (p. 1).

Parents are their children’s first teachers imparting by word and deed, knowledge, cultural values, and societal expectations. How parents go about these tasks is important. Eleanor Maccoby (Luskin, 2001) agreed. In an interview with Stanford University’s Bing Times Online, she said:

Of course parents matter in the way children live their daily lives and in what happens to them. You may ask why anyone would doubt it. . . . Socialization researchers start out with the preconception, the idea, that there a good child-rearing styles and bad child-rearing styles, and that parents who use good ones will have better children than those who use bad ones. Just as we all know the rules for a healthy lifestyle, we know all the rules for good child-rearing. Give children plenty of love and approval, set limits, enforce them firmly but fairly, don’t use physical punishment or make belittling remarks, be consistent and so on. (para 1, 4)

Many parents strive to be good parents. Although Maccoby’s formula sounds simple, it often has been difficult for parents to implement it. Moreover, parenting in the 21st Century is challenging. This was the conclusion Gardner and her research team at the University of Oxford came to after they analyzed longitudinal data from the British
Household Panel Survey (BHPS) as well as data generated by a Nuffield project (Gardner, Maughan, Scott, & Collishaw, 2009). According to Gardner, the advent of such new technology as the Internet and cell phones has made it more difficult for parents to monitor who their children are spending time with, what their children are watching, and what other online activities they are engaged in. Moreover, young adults are responding to the tight job market and higher cost of living by continuing to live with their parents well into their 20s. Thus, today’s parents must develop new, more complex skills than their earlier counterparts.

Although Gardner conducted her research in the United Kingdom, these trends have been mirrored in the developed countries around the world. Thus, according to Gardner, today’s parents must develop new, more complex skills than their earlier counterparts. But what are these new skills, and where do parents learn them? In short, what constitutes good parenting today?

Previous generations of parents have asked the same question. And previous researchers have strived to answer it. Although parenting roles and styles have changed over the centuries, the advent of empirical methods of investigation in the early part of the last century focused researchers’ attention on the issue (Stearns, 2003). In the 1930s and 1940s, Baldwin, Kalhorn and Breese (1945) conducted a longitudinal study in which he observed parent-child interactions over the years and identified two different parenting styles (democratic and autocratic). Nearly simultaneously, Radke (1946) authored The Relation of Parental Authority to Children’s Behavior and Attitudes. Even though her research was hailed as a good first attempt at careful experimental research, reviewers
noted the lack of “clear-cut” results due to the difficulty in studying parent-child interaction in the home environment (Bernhardt, 1945).

In the late 1960s, Baumrind (1966, 1967, 1968a, 1968b; Baumrind & Black, 1967) undertook such a project and conducted a series of studies examining the interactions between parents and their children. The focus of the early research was on children rather than parents. Specifically, Baumrind investigated “whether preschool children who are assertive, self-reliant, self-controlled, buoyant, and affiliative” are parented differently than “children who are discontented, withdrawn, and distrustful,” or “children who have little self-control or self-reliance and tend to retreat from novel experiences” (Baumrind & Black, 1967, pp. 291-293). Her parenting styles grew out of the empirical description of these parents. Subsequent research has focused on investigating the resulting parenting styles.

In all of Baumrind’s research, her focus was always on optimal parenting and never on pathogenic or pathological parenting. She wanted to know what kind of parenting strategies parents of competent children practiced. Much of the other research at the time focused on parenting dimensions of problem children. In her research endeavors, Baumrind explicitly focused on parenting of normal children.

Baumrind (1966) began her research on parent-child interactions by studying 140 Berkeley families with children enrolled in University-run nursery schools. Researchers first observed the children for 3-5 months in the school setting and while they were engaged in taking the Stanford-Binet. According to Baumrind (1968a), “Observation over a 3-5 month period permits the observer to distinguish enduring response tendencies from maturational factors” (p. 2). Then, each child was rated with a 72-item Q sort which was
designed to rate the child on eight different constructs (with 9 Q-sort items each): (a) stress tolerance, (b) self-confidence versus fearfulness, (c) achievement orientation, (d) approach-oriented versus withdrawn, (e) autonomous versus suggestible, (f) rebellious versus dependable, (g) destructive versus constructive, and (h) alienated versus trusting.

Observers collected parent data during extended home visits during the 3 hours before bedtime. Prior to the home visit, parental attitudes were obtained through self-report data. Each parent was interviewed and rated in a Q-sort process similar to one described above. Constructs included (a) articulated child-rearing philosophy versus lack of philosophy, (b) directive versus non-directive behavior, (c) firm versus lax enforcement, (d) confidence versus lack of confidence as a parent, and (e) discourages versus encourages infantile behavior. While these continued to be her primary research methods, subsequent research was augmented by questionnaires when the children were capable of reading. Unlike research by others, self-report instruments played only a small part, if at all, in her research. Even though in her initial research Baumrind (1968a) also used likert-type self-report measures, she abandoned them after issues with such measures and parents unwillingness to choose a response because they took issue with the wording of the item or felt that a response of the item implied either a positive or negative evaluation.

After this experience, I much prefer an interview to a self-report inquiry as a method for assessing parental attitudes and values. My assumption that parents would prefer to take a self-report inquiry in the comfort of their own home than to be interviewed was incorrect. The parents we studied preferred for the most part to discuss those views which mattered to them, in depth, in an interview. (Baumrind, 1968a, p. 6)

In her seminal monograph, Current Patterns of Parental Authority, Baumrind (1971a) described the three parenting styles she identified in this and previous studies
(authoritative, authoritarian, and permissive). Baumrind’s work became the foundation for future research on parent-child interactions.

But Baumrind’s methods of observing parenting style were time and labor intensive. It was simply not practical for researchers with limited means or those working in less privileged environments. While several subsequent studies used self-report measures (e.g., Dornbusch, Ritter, Ledierman, Roberts, & Fraleigh, 1987), it was nearly two decades later that Buri (1991) published the first widely used instrument to assess parenting styles as set forth by Baumrind (1971a). Although the Parental Authority Questionnaire (PAQ) was designed to assess parenting style according to Baumrind’s three prototypes (authoritative, authoritarian and permissive), Buri’s methods did not require as huge an investment of research time and money as Baumrind’s methods did. The PAQ has been widely used since that time to study parenting styles and how they impact various aspects of child development. Yet in spite of its wide usage, little has been known about its psychometric properties and how they varied across populations and samples.

**Purpose**

The purpose of this study was to analyze the reliability of the Parental Authority Questionnaire (PAQ) using reliability generalization, a meta-analytical technique, which analyzes the central-tendency, variability, and systematic variation of reliability coefficients. This study also compared two different meta-analytic models for evaluating instrument reliability across studies: the random-effect model and the varying coefficient model.
Rationale

Since its development in the late 1980s, and subsequent publication in 1991, the Parental Authority Questionnaire (PAQ; Buri, 1991) has been widely used to measure parental authority and to categorize parents into three distinct prototypes (permissive parents, authoritarian parents, and authoritative parents). Using this measure, hundreds of studies conducted all over the world, in dozens of disciplines, have identified correlations not only between parenting style and academic achievement (Hickman, Bartholomae, & McKenry, 2000) but also the ability to transition successfully into college or the military (Wintre & Yaffe, 2000), adolescent decision-making, identity style and identity commitment (Berzonsky, 2004), self-esteem (Cardinali & D’Allura, 2001), spiritual maturity (Bryant, 2003), body attitudes (Corgiat & Templer, 2003), marital quality (Gersho, 1999), and even insomnia (Brassington, 1994). Indeed, over 500 studies, in a variety of disciplines, have cited the Parental Authority Questionnaire. With so much riding on this one measure of parental authority, one must ask: Just how reliable is this instrument? Is it valid? Without answers to these questions, the magnitude of the relationship between parenting style and its effect on the children could not be known. The practical and theoretical significance of parenting styles--and other independent variables--rests upon the magnitude of the effect and not just statistical significance. Statistical significance can occur with a tiny and trivial magnitude (Ellis & Steyn, 2003).

Research Questions

Q1 Does the Parental Authority Questionnaire produce reliable information across groups when used by adolescents, young-adult or adult children to evaluate their parents’ parenting styles?

Q2 What moderator variables (e.g., country of origin, culture, age of respondent) can be identified that might affect reliability?


**Delimitations**

This meta-analysis only included those studies that used the PAQ to assess the parenting styles as reported by offspring, those studies that were available to the researcher, and those that reported reliability on at least one of the six scales. Only studies that used the original PAQ as developed by Buri (1991) were included.

**Definitions**

Since Buri (1991) constructed the PAQ specifically to measure Baumrind’s (1971a) parental authority styles, I used their definitions of the constructs.

*Authoritarian parents.* According to Buri (1991), authoritarian parents were highly directive and expected their mandates to be obeyed without discussion or question. He asserted that these parents used punitive measures to control their children’s behaviors, and he characterized these parents as more detached and less warm than other parents.

*Authoritative parents.* According to Buri (1991), authoritative parents provided clear direction for their children but couched such direction with “warmth, flexibility, and verbal give-and-take” (p. 111).

*Parental authority.* According to Buri (1991), parental authority referred to the manner or style with which parents guided their children’s lives and behaviors. Parental Authority, and thus parenting style, was determined by both a parent’s level of responsiveness and demandingness.

*Permissive parents.* According to Buri (1991), permissive parents were relatively non-controlling, made few demands on their children, allowed them to make their own decisions, and regulate their own behavior.
Responsiveness. According to Baumrind (2005), responsiveness referred to “the extent to which parents foster individuality and self-assertion by being attuned, supportive and acquiescent to children’s requests. It includes warmth, autonomy support, and reasoned communication” (pp. 61-62).
CHAPTER II

LITERATURE REVIEW

What is good parenting? Who makes the best parents? What kind of parenting style is most beneficial? Most experts have agreed that when it comes to the cognitive development of children, an authoritative parenting style, as defined by Diana Baumrind in her more than a half century of research, is best (Gauvain, Perez, & Beebe, 2013). Indeed, a meta-analysis of 34 primary studies done by Rosenzweig (2001) concluded that positive parenting practices (including authoritative parenting) have accounted for nearly one quarter of the variance in student achievement outcomes. Moreover, an additional one-third of the variance in student achievement could be accounted for by negative parenting practices and that these practices have led to academic failure. The researcher found that negative parenting practices have accounted for nearly one-third of the variance in student achievement and have led to academic failure. As educators, the ultimate goal should be to foster the cognitive development of children. Therefore, it is important that educators harness all of the tools at their disposal to help students reach academic goals, including partnering with those who are a child’s first teachers—his or her parents.

Baumrind’s Two-factor Model

Baumrind’s theory grew out of her empirical research on children and their families. Unlike other studies of the time, she wanted to know what kind of parents produced optimally competent children, who she defined as being assertive, self-reliant,
self-controlled, buoyant, and affiliative (Baumrind & Black, 1967). Interestingly, educational attainment was never a primary focus of her research. In addition to competent children, Baumrind identified two other groups of children: those that were discontent and withdrawn, and those who lacked self-control and self-reliance. According to Baumrind and Black (1967).

Findings from the first study can be summarized briefly as follows: parents of the most assertive, self-reliant, and self-controlled children were controlling, demanding, communicative, and loving; parents of the unhappy and disaffiliated group were relatively controlling and detached; and parents of the least self-reliant and self-controlled group of children were noncontrolling, nondemanding, and relatively warm. (p. 292)

Her subsequent decades of work focused on these parenting styles she discovered from the empirical description of these parents.

According to Baumrind’s two-factor typology, the ideal parent, who she labeled as authoritative, exhibited both high levels of responsiveness (often referred to as nurturance) and demandingness. In her model, each of these traits was represented on its own orthogonal scale and they intersected to form what could be characterized as a theoretical circumplex model with level of responsiveness on one axis and level of demandingness on the other (see Figure 1; It is important to note that Baumrind [1991] added the disengaged parent in later research). In a circumplex, which is a type of factor analysis, each axis of the bi-factor model forms two radii from the central point (intersection) which cordon off each of four quadrants (Guttman, 1954). What distinguishes a circumplex model from a typical factor analysis is that all points will fall along a diameter of a circle drawn with the intersection at its core, thus, neither orthogonal nor oblique rotations will improve the fit. Instead, the quadrants are determined on a theoretical basis.
After her initial empirical research, Baumrind’s theoretical model emerged and was informed by Schaefer’s (1955) earlier Circumplex Model for Maternal Behavior. For his model, Schaefer defined the two axes as Autonomy versus Control and Love versus Hostility. According to Schaefer (1955), “the advantage of the circumplex order over the previous organizations is that it directs attention to the sequential order of variables and does not divide the matrix into discrete clusters or factors” (p. 232). Schaefer fit the data presented in three previous studies (Baldwin et al., 1945; Sanford, Atkins, Miller, &
Cobb, 1943; Schaefer, Bell, & Bayley, 1959) to his circumplex model. Schaefer (1955) concluded that,

By organizing several empirical studies of maternal behavior using Guttman’s (1954) circumplex model, it was demonstrated that many of the existing concepts of maternal behavior can be ordered within a two-dimensional space” and that those two dimensions were love versus hostility and autonomy versus control. (p. 234)

Although it is true that Baumrind’s two dimensions have not been entirely congruent with Schaefer’s, hers (responsiveness and demandingness) were an outgrowth of her own research, which was also informed by Baldwin et al.’s (1945) research and the patterns of parent behavior. Baldwin et al. identified three paradigms that influenced parent behavior (democracy in the home, acceptance of the child, and indulgence). In terms of democracy in the home, researchers considered how parents justified their policies, how democratic those policies were, whether their suggestions were non-coercive, how readily they explained their decisions, how they directed criticism and approval, how they clarified their policies, how well they understood their child, and how non-restrictive their rules were. In terms of acceptance of the child, researchers considered the rapport the parents had with the child, how affectionate they were, how they directed criticism or approval, how effective their policies were, how child-centered the home environment was, and the degree of non-disciplinary friction. In terms of indulgence, researchers considered how protective the parents were, whether they “babied” the child, how child-centered the home was, how they accepted the child, how they showed concern for their child’s welfare, the length of the contact mothers had with their children, and the intensity of that contact. Baldwin et al. acknowledged, however, that “only its grossest aspects [of parent behavior] have been delineated [by their
research, but the basic structure seems substantial enough to serve as a foundation for further research” (p. 71).

Baumrind took up the challenge. But Baumrind started with the children--normal children. She first identified “competent” students and then worked “backward” to study how their mothers and fathers parented them. Baumrind (1967) defined competent children as those who were self-reliant, self-controlled, explorative, and content, as opposed to children who were discontent, withdrawn, and distrustful, or those who retreated from novel experiences and had little self-control or self-reliance. In each of three separate studies (Baumrind, 1967, 1971b; Baumrind & Black, 1967). Baumrind spent three months observing children in a preschool setting and then in structured situations with their parents. She then conducted two 3-hour home visits at each child’s home in which she observed parent-child interactions. Finally, she conducted structured interviews with each child’s mother and father. The result was Baumrind’s (1971a) seminal monograph, *Current Patterns of Parental Authority*, in which she described the three parenting styles she identified in this and previous studies (authoritative, authoritarian, and permissive). Subsequent research followed the children and their parents into adolescence.

Baumrind discovered that the competent children had parents that she later called authoritative parents. According to Baumrind (2005, 2013), authoritative parents exhibited both high responsiveness and high demandingness, authoritarian parents exhibited low responsiveness but high demandingness, and permissive parents exhibited high responsiveness but low demandingness. She characterized responsiveness as the “extent to which parents foster individuality and self-assertion by being attuned,
supportive, and acquiescent to children’s requests; it includes warmth, autonomy support, and reasoned communication” (Baumrind, 2005, p. 61). On the other hand, according to Baumrind (2005), demandingness “refers to the claims parents make on children to become integrated into society by behavior regulation, direct confrontation, and maturity demands (behavioral control) and supervision of children’s activities (monitoring)”(p. 62). In subsequent research, Baumrind pointed out that responsiveness and demandingness were not on opposite ends of the same scale, but rather two separate orthogonal scales (Baumrind, 2013).

Baumrind’s parental prototypes were born out of the way in which parents synergistically express, or fail to express, responsiveness and demandingness. Each parental prototype reflected a complex continuum of behavior. One facet of distinguishing behavior between the parental prototypes was the way in which parents wielded their authority. Authoritative parents used confrontive control rather than coercive control. Confrontive control was goal-oriented, behavior-directed, tailored to a child’s temperament and stage of development, subject to rationalization, and open to negotiation. Coercive control, on the other hand, was first and foremost about maintaining control and demanding obedience, rather than about managing behavior. It was undermining autonomy and was manipulative, punitive, intrusive, and restrictive (Baumrind, 2013). According to Baumrind (2013), this was one of the crucial differences between the authoritative and authoritarian parenting prototypes. While authoritarian parents would, on occasion, exercise confrontive control, authoritative parents refrained from exercising coercive control. Permissive parents, on the other hand, viewed any type of control, confrontive or coercive, as an “undesirable, counteractive force to children’s
growth-enhancing motive to act autonomously” (Baumrind, 2013, p. 18). Early in her research Baumrind (1966) hypothesized that:

Authoritarian control and permissive noncontrol may both shield the child from the opportunity to engage in vigorous interaction with people. Demands which cannot be met or no demands, suppression of conflict or sidestepping of conflict, refusal to help or too much help, unrealistically high or low standards, all may curb or understimulate the child so that he fails to achieve the knowledge and experience which could realistically reduce his dependence upon the outside world. The authoritarian and permissive parent may both create, in different ways, a climate in which the child is not desensitized to the anxiety associated with nonconformity. Both models minimize dissent, the former by suppression, and the latter by diversion on indulgence. To learn how to dissent, the child may need a strongly held position from which to diverge and then be allowed under some circumstances to pay the price for nonconformity by being punished. Spirited give and take which the home, if accompanied by respect and warmth, may teach the child how to express aggression in self-serving and prosocial causes and to accept partially unpleasant circumstances of such action. (p. 904)

After obtaining a second wave of data when the participants in her longitudinal study were nine, Baumrind (1991) identified a fourth type of parent, the disengaged one. She characterized these parents as neither responsive nor demanding.

Later, when these same children were 15-years-old, Baumrind conducted a follow-up study and further differentiated her parenting styles by looking at the degree of imbalance in their demandingness-responsiveness ratio, as well as level of intrusiveness (Baumrind, 2005; see Figure 2). Authoritative parents were highly demanding, highly responsive, and lowly intrusive. Directive parents (those who were highly demanding) were divided into Authoritarian-Directive (highly demanding, highly intrusive, and lowly responsive) and Nonauthoritarian-Directive (highly demanding, moderately responsive, and moderately or lowly intrusive). Lenient parents (lowly demanding) were divided into Permissive (lowly demanding and highly responsive) and Democratic (moderately demanding and highly responsive). Disengaged parents (those who were lowly
demanding and lowly responsive) were divided into Rejecting (hostile and intrusive) and Neglecting (indifferent). Good-enough parents were identified as those that were moderately demanding and moderately responsive (Baumrind, 2005).

![Diagram of Baumrind's revised Parental Authority Prototypes]

**Figure 2.** A Theoretical Circumplex illustrating Baumrind’s revised Parental Authority Prototypes.

**Subsequent Research by Others**

Other researchers accepted and built on Baumrind’s model. In their chapter in volume four of the *Handbook of Child Psychology* (“Socialization in the Context of the Family: Parent-Child Interaction”), Maccoby and Martin (1983) synthesized the research to date, linking authoritative parenting to children’s competence, including social assertiveness, social responsibility, and, cognitive competence. They also examined research regarding the negative impacts of authoritarian parenting on children’s moral
development, self-concepts, and aggression. As for permissive parenting, their review of the research concluded that,

The issue of permissiveness hardly comes up with respect to behaviors that are socially approved or desired. Thus permissiveness is a dimension of parental variation that is brought to the fore when the child engages in some form of behavior that disturbs or threatens others, or otherwise contravenes some social norm. (Maccoby & Martin, 1983, p. 44)

Although Baumrind had not yet identified the disengaged parental prototype, Maccoby and Martin (1983) concluded that the degree of parental involvement, including both positive and negative interactions, was an important factor to consider in the study of parental prototypes:

A pattern of family functioning in which children are required to be responsive to parental demands, and parents accept a reciprocal responsibility to be as responsive as possible to their children’s reasonable demands and points of view has been labeled “authoritative” by Baumrind. We use this term and add the label ”reciprocal.” (p. 46)

They went on to say: “We have seen that the authoritative-reciprocal pattern of parenting is associated with children’s being independent, ’agentic‘ in both the cognitive and social spheres, socially responsible, able to control aggression, self-confident, and high in self-esteem” (p. 48). Other researchers agreed. Darling and Steinberg (1993) felt that Baumrind’s parenting styles,

Produced a remarkably consistent picture of the type of parenting conducive to the successful socialization of children into the dominant culture of the United States. (p. 487)

They lauded Maccoby and Martin’s (1983) synthesis stating that their two-dimensional approach helped “facilitated investigations of the generalizability of Baumrind’s model to populations quite different from the one in which the typology arose” (p. 491). Later, Gray and Steinberg (1999) further tried to “unpacked”
authoritative parenting by examining the three basic tenets of the construct, as they
defined it, with regard to adolescents (acceptance-involvement, strictness/supervision,
and psychological autonomy granting). They studied a sample of 8,700 adolescents and
found that each of the three tenets were associated with academic competence. Gray and
Steinberg (1999) concluded that their findings,

Both support and extend previous research on the relationship between
authoritative parenting and adolescent adjustment. (p. 583)

Our investigation provides evidence that three of its [authoritative parenting]
central features—parent involvement, behavioral control, and autonomy granting—
contribute in unique ways to psychosocial development, academic competence,
behavior problems, and internal distress. (p. 586)

Yet it has been unclear if Gray and Steinberg truly did “unpack” authoritative
parenting because they made a radical change in methodology, departing from
Baumrind’s original methods and employing self-report measures to collect data rather
than extensive observation that Baumrind favored. This change is discussed in greater
detail later in this dissertation.

Objective versus Subjective
Perspective

As noted earlier, Gray and Steinberg (1999) and his colleagues, among others,
departed radically from Baumrind’s methodology. They focused on self-report measures,
usually coming from the children, while Baumrind’s research methodology consisted
primarily of months of observation in multiple settings, structured interviews, and
observation at school, home, and in laboratory settings. Self-report measures played only
a small role in Baumrind’s studies and were only used in upper grades (Steven Pulos,
personal correspondence, October 23, 2013).
Unlike Baumrind’s research, which was done in the tradition of the Berkeley Growth studies, Gray and Steinberg (1999) used questionnaires administered to adolescent subjects to gather their data. They asked the teens how they would characterize their parents in response to given situations. While Baumrind’s research involved trained researchers spending hundreds of hours observing participants in both school and home environments, and hours more coding and triangulating the qualitative data, instruments such as those used by Gray and Steinberg were efficient to administer and easier to code (Cundy, 1979). Even so, these measures relied on the subjective perceptions of the respondents. Moreover, critics, including Baumrind (2005) herself, argued that self-report measures could not accurately capture parent-child relationships: “conclusions relating parenting style to family conflict based on parenting measures that assess conceptions of parental authority may differ from those that assess how parental authority is exercised in practice” (p. 64).

**Parent versus Child Perceptions**

Indeed, a number of researchers have identified a variety of issues that may confound children’s, and even parents’, perceptions of parental authority. In 1969, Goldin conducted a comprehensive review of studies that included children’s reports of parental behavior. He organized his review around three orthogonal factors synthesized from both Schaefer’s (1955) research, as well as Siegelman’s (1965): Loving (Acceptance-Rejection), Demanding (Psychological Autonomy-Psychological Control), and Punishment (Firm-Control-Lax Control). It was important to recall that Baumrind’s work was not yet published, thus, it was not included in this analysis. Goldin (1969) reviewed more than 60 studies conducted between 1940 and 1965 and concluded that children’s
reports of their parents’ behaviors differed from parents’ reports and said differences were thought to be related to sex, social class, and the children’s behavior.

More recently, other researchers have found similar differences. In her 1995 study looking at parenting styles and adolescents’ conceptions of parental authority using the Parental Authority Questionnaire, Smetana (1995) found that adolescents perceived their parents as permissive and more authoritarian than their parents viewed themselves. Parents tended to view themselves as more authoritative with regard to their parenting style. Correlations between parent and child responses indicated significant differences. With regard to parent and adolescent responses, they were correlated (or not correlated) at .21 ($p < .05$) for permissive, .42 ($p < .001$) for authoritarian, and .22 ($p < .05$) for the authoritative parenting prototype.

In 2001, Feinberg, Neiderhiser, Howe, and Hetherington conducted a study that investigated adolescent, parent, and observer perceptions of parental warmth and negativity and the underlying biases of parents and children. According to the study’s authors, “the literature on family relations suggests that shared reality is surprisingly limited” (Feinberg et al., 2001, p. 1226). Results of their study, which showed very low correlations between parent-child reports ($r = .28-.31$), parent-observer reports ($r = .12-.25$), and child-observer ($r = .09-.22$), supported this assumption. Yet Feinberg and his colleagues did not find children’s reports to be useless. They found that, although children’s reports were different, they added predictive power to more objective measures of parenting.

After surveying 302 mothers, fathers, and their children, Villar, Luengo, Gómez-Fraguela, and Romero (2006) used a multi-trait model and conducted a confirmatory
factor analysis (CFA) to determine the validity and convergence of five separate constructs (family conflict, family communication, permissive parenting, authoritative parenting, and authoritarian parenting). They concluded that, while mothers and fathers tended to agree with regard to parenting style, their children viewed their styles differently.

Gender Differences

As Goldin’s (1969) research suggested researchers have found that children perceived parental behaviors differently depending, not only upon their gender, but also upon their parent’s gender. Using the Parent Perception Inventory (PPI), Hazzard, Christensen, and Margolin (1983) surveyed 75 children, ages 5 to 14, and their parents and also discovered differences between the perceptions of children and parents regarding parents’ behaviors. Although they concluded there were no age related differences, the researchers did find significant gender effects. Boys viewed their parents’ behaviors more positively, especially with regard to their fathers, and adolescents in general viewed their mothers’ behavior as being more punitive.

Russell et al. (1998) looked at the flip side of this issue: how mothers and fathers parent differently depending upon the gender of their children. Researchers used the Parenting Practices Questionnaire (PPQ; Robinson, Mandleco, Olsen, & Hart, 1995) to survey 305 parents (108 mother-father couples, 5 fathers, and 84 mothers) of 4-year-old preschool children (102 girls and 95 boys). They found “prominent” sex-based differences. The study found that mothers were more authoritative in nature as compared to fathers who were more likely to be either authoritarian or permissive parents. Further,
as a whole, parents used the authoritarian style more when dealing with boys and the authoritative style with girls.

McKinney and Renk (2008) found similar results. Researchers surveyed 151 male and 324 female college students between the ages of 18 and 22 by administering the PAQ, the Rosenberg Self-Esteem Inventory (Rosenberg, 1965), the Beck Depression Inventory (Beck, Steer, & Brown, 1996), and the Manifest Anxiety Scale (Taylor, 1953) to look at differences between maternal and paternal parenting styles of adolescents. Based on the results of their research, McKinney and Renk (2008) concluded “that fathers and mothers may parent their sons and daughters differently” (p. 821).

Other Characteristics of the Respondent

Another issue that could confound perception of parental authority was the ages of the respondents and which “parents” they were evaluating: (a) their parents as they were when the respondents were children; (b) their parents as they were when they were adolescents; or (c) they were when they were adults. For example, parenting perceived as “too strict” or authoritarian as adolescents, in hindsight, might be viewed as quite appropriate when the respondents were adults. As the relationship between parents and their children changed over time, so might children’s perception of parenting style change over time (Nucci, Hasebe, & Lins-Dyer, 2005). According to De Los Reyes and Kazdin (2005), “Evidence suggests that the perspective or stance people have when they recall events from memory determines their memory recall” (p. 491). Hardt and Rutter (2004) agreed, “The recall experiences that rely heavily on judgement and interpretation have not been found to have satisfactory validity” (p. 270). Since many studies that
considered parenting style used undergraduate psychology students as their respondent populations, we must consider this issue.

To further cloud children’s perception of the parenting styles of their fathers and mothers was the phenomena that parents act differently at different times in their children’s lives. Moreover, mothers and fathers may treat each of their children differently (Maccoby, 1992; Nucci et al., 2005). According to Baumrind (2013), “Authoritative parents tailor the balance between asserting control and allowing freedom to the resilience, unique temperament and developmental stage of their child” (p. 19). Feinberg, Neiderhiser, Simmens, Reiss, and Hetherington (2000) looked at this issue of “differential parenting” of siblings and concluded that: “Development, at least for the 80% of American children who have at least one sibling, cannot be understood simply in terms of separate dyadic parent-child relationships and dyadic sibling relationships” (p. 1626).

Smetana and Campione (2003) agreed that, depending upon the circumstances, parents change how they treat their children: “Research has demonstrated that parenting practices are affected by situational factors” (p. 1210). These factors included the child’s behavior and the type of transgression, as well as whether the wrong was merely a lack of consideration for others or one that might involve harming others.

**Culture and Social Milieu**

Children’s cultural and/or socio-economic background also affected how they viewed their parents with regard to parenting style. It was important to remember that Baumrind studied normal, well-adjusted European-American children in middle-class
families. It was still not clear whether her model could be applied to families in other cultures or social milieus (Darling & Steinberg, 1993).

McBride-Chang and Chang (1998) used the PAQ to survey 906 adolescents and 1,091 parents. They concluded that “categorizing parenting styles as authoritative, authoritarian, or permissive may not be a culturally relevant dimension of socialization in Hong Kong adolescents” (p. 430). Chao (2001) used the PAQ to study the differences in parenting styles between Chinese immigrant and European American parents. Her results indicated that Chinese mothers favored authoritarian and permissive parenting styles. Kim and Gim Chung (2003) used the PAQ to survey 144 Korean American college students and their findings only partially supported Baumrind’s model as applied to Korean American families and concluded that their findings “demonstrate the importance of considering acculturation issues in parenting studies” (p. 481).

Similarly, Varela et al. (2004) used the PAQ to examine differences in parenting style between families of Mexican descent and Caucasian non-Hispanic families. While researchers found no parenting style differences between Mexican and Caucasian parents, their results did suggest that ethnic minority status, rather than affiliation with Mexican culture, played a role in greater use of authoritarian parenting style.

In studying how spouses and offspring reported on the political positions of other family members, Cundy (1979) concluded: “Short of direct observation, there is no dependable way to produce valid assessments of self-reports and external observer perceptions” (p. 499).
Direct Observation and Its Foibles

But even direct observation has had its issues. In 1987, Achenbach, McConaughy, and Howell conducted a meta-analysis of 119 studies to determine how consistently different observers (parents, teachers, mental health workers, outside observers, peers, and the subjects themselves) rated the emotional/behavioral problems of children and adolescents. The researchers found that ratings by observers who had similar relationships with the children, mothers and fathers, for example, had correlations that hovered around .60, while ratings by observers who had different types of relationships with the children, teachers and parents, for example, had correlations of only .28. Moreover, the correlations between subject ratings and observer ratings were only .22.

More recently, De Los Reyes and Kazdin (2005) delved into this issue of “informant discrepancies” and explained the inconsistencies between parent, child, teacher, and other observers by proposing that it was role of the context--the situation in which the observation or behavior occurred--which was key. They explained, for example, that parents were more likely to blame their child’s behavior on that child’s disposition, while children were more likely to attribute said behavior to outside forces. According to De Los Reyes and Kazdin (2005), “discrepancies are, in part, a function of the different attributions informants have of the causes of the child’s behavior and the different perspectives that informants have” (p. 502). After later consideration of the issue, De Los Reyes concluded that discrepancies such as these should not necessarily be seen as a failure to corroborate other data, but rather these discrepancies were an indicator of underlying issues within the relationships being studied. De Los Reyes (2011) explained:
Informant reports of children’s behavior should reveal important information on how children’s behavioral expressions vary across situations and time. It was this focus on qualitative distinctions between parent and child reports that proved to provide important information on the prediction of an outcome commonly linked to parental monitoring. (pp. 2, 4)

Even so, he acknowledged that these discrepancies might also be interpreted by other researchers as “a disconnection among measures in how they represent the behaviors being assessed” (pp. 2-3).

Most researchers have agreed that there was value in collecting and triangulating multiple sources and kinds of data and studying informant discrepancies as well as instrument discrepancies (Creswell & Clark, 2007). Even so, one must ask if these disparate results were truly due to solely situational or contextual differences, or whether they were due to a lack of instrument reliability or validity. This question has become even more germane when only one source, such as a questionnaire, was used to gather information about relationships between two or more people. While both forms of research could be useful, they are indeed different and both contribute to the variance in predicting behavior. Therefore, we must ask how reliable our methods are and what are the true effects of any given treatment or intervention.

**Construct Validity**

All of these issues come into play when addressing the validity of a particular instrument. Construct validity refers to how well an instrument measures the psychological construct. Although there were no current studies addressing the overall validity of quantitative instruments that measure parental authority, Holden and Edwards (1989) conducted a meta-analytical review of more than 80 instruments measuring
parental attitudes about child rearing. The researchers examined three types of construct validity (convergent, discriminant, and population validity).

Convergent validity refers to how well the instrument produces results that are consistent with the results of others measures of the same construct. Discriminant validity (divergent validity) refers to how well an instrument does not measure what it is not supposed to measure (Holden & Edwards, 1989). What the researchers uncovered was a surprising lack of validity across the measures. Most studies reviewed did not provide evidence of the convergent validity of their subscales, rarely examined how their measures differed from other “distinct constructs,” and provided “no evidence” indicating that their results could be generalized to other populations (Holden & Edwards, 1989, pp. 41-42).

**Instrument Reliability**

Holden and Edwards (1989) also considered the reliability of the instruments in his meta-analytical review, including test-retest and internal consistency reliability. As defined by Holden and Edwards (1989), test-retest reliability looks at the consistency of a subject’s responses on the same items over a short period of time. As Cronbach (1951) noted more than 50 years ago, while both internal consistency and test-retest reliability were termed reliability, they were quite different statistics and should rightly be given different names, (e.g., test-retest should be called a coefficient of stability). It could be possible for one to be high and the other low. Holden and Edwards (1989) noted that only 21 of the 83 studies reviewed provided test-retest data, less than half reported a measure of internal consistency, and only 11 reported on the factor structure of the instrument.
A Micro Look at the Parental Authority Questionnaire (PAQ)

In light of the results of Holden and Edward’s (1989) research and the questions that have been raised regarding the use of quantitative measures to study the construct of parental authority, it seemed prudent to take a closer look at one of the most widely used instruments to assess parental authority--the PAQ (Parental Authority Questionnaire) developed by Buri (1991). With more than 500 citations in the fields of psychology, education, business, the law, and others, Buri’s (1991) Parental Authority Questionnaire has been used in studies across age groups, gender, ethnic groups and nationalities. Since its development in the late 1980s, and subsequent publication in 1991, the Parental Authority Questionnaire (Buri, 1991) has been widely used to measure parental authority and to categorize parents into Baumrind’s three distinct prototypes (permissive parents, authoritarian parents, and authoritative parents). Using this measure, hundreds of studies conducted all over the world have identified correlations not only between parenting style and academic achievement (Hickman et al., 2000) but also the ability to transition successfully into college or the military (Wintre & Yaffe, 2000), adolescent decision-making, identity style and identity commitment (Berzonsky, 2004), self-esteem (Cardinali & D’Allura, 2001), spiritual maturity (Bryant, 2003), body attitudes (Corgiat & Templer, 2003), marital quality (Gersho, 1999), and even insomnia (Brassington, 1994).

Buri (1991) developed the instrument as an efficient way of determining parenting style so as to study its effects on different areas of development. He felt that the research approach taken by Baumrind (extensive observation and follow-up interviews) was time-consuming, expensive, and required extensive training of researchers. He also
criticized previous attempts to develop an effective measure because researchers either took into account the academic achievement of the parents or because they did not allow for differentiation between mothers’ and fathers’ possible disparate parenting styles.

Buri used a systematic approach in developing his questionnaire, vetting his questions with panels of experts and then administering it to different groups of college students to determine test-retest reliability and internal consistency reliability. Although Buri’s (1991) instrument captured adolescent and adult children’s current attitudes regarding their parents’ parenting practices, he contended that the PAQ would capture parenting style as “phenomenologically appraised” by their offspring (p. 111).

Although Buri’s (1991) PAQ focused on parental demandingness, he conceded that the dimensions of warmth and nurturance were not adequately captured by his instrument. Thus, the experts who vetted the survey questions were not able to consider nurturance when deciding whether a particular item should be included in the survey. In order to address this issue, Buri administered both the PAQ and the Parental Nurturance Scale (Buri & Mueller, 1988) to a group of college students. His reasoning was that authoritativness should be positively correlated with parental nurturance and authoritarianism should be negatively correlated with nurturance. The results supported his hypothesis.

Reliability and the Parental Authority Questionnaire

Yet with so much riding on this one measure of parental authority, as researchers, we must ask ourselves: Just how reliable are the results of this instrument across different groups and populations? Similar to Holden and Edward’s (1989) review, a cursory glance at the research using the PAQ showed that many studies did not include reliability
estimates (Brassington, 1994; Chao, 1996;), and others cited the reliability of the instrument as reported by Buri (reliability inductions; Dreyer, 1995; Lai & McBride-Chang, 2001; Wintre & Yaffe, 2000). Yet, reliability does not “travel” with the instrument (Thompson & Vacha-Haase, 2000). Instead, it must be calculated each time the instrument is used. Indeed, with respect to the PAQ’s reliability across studies, there were several problems: (a) most studies did not report or examine the reliability of the instrument with respect to their study; (b) even when reliability was reported, it was frequently on small sample sizes which with large confidence intervals and did not give an exact reliability estimate for the population; and (c) even when reliability was reported on a sufficient sample size, it was not clear if the reliability was affected by task variables and demographic variables.

It was important to consider the reliability of the instrument for each separate population it was administered to because unreliability contributed noise to the data analysis such that the greater the noise, the greater the distortion of the data. This distortion could occur in two ways: (a) the unreliability distorts the accuracy with which we could measure the effect size such that low reliability may obscure the true effect size that exists among the constructs being investigated, and (b) low reliability reduces the statistical power such that a greater sample size would be needed to maintain the power constant and enable researchers to make a correct decision regarding accepting or rejecting hypotheses. Moreover, having an accurate estimation of the power would be essential in determining sample size. If it was assumed that the reliability was greater than it was, the researcher would likely overestimate the power and select too few participants (Vacha-Haase & Thompson, 2011).
Furthermore, unless one knows the true reliability of an instrument, the true effect size of the relationship of interest could not be determined because the relationship among constructs would be distorted. Thus, the researcher would underestimate the extent of the relationship, or we may even commit Type I error and say there was no relationship when there actually may be one (Creswell & Clark, 2007).

In summary, without accurate reliability estimates, it would be difficult to understand the extent of the effect, which would hamper the ability to properly plan future studies. Thus, in order to understand the results yielded by the PAQ among different groups and populations, it was necessary to determine if the studies were reporting true effect sizes or if the significance found was due to distorted reliability. Unless the reliability of the PAQ was investigated, outside factors that could moderate the magnitude of the reliability, such as age of respondent or language of the instrument, could not be determined.

**Reliability Generalization**

Reliability generalization is a way to determine, across studies, a given instrument’s: (a) typical reliability of scores, (b) variability of reliability coefficients, and (c) sources of variability of reliability coefficients (Vacha-Haase, 1998). An outgrowth of validity generalization (Hunter & Schmidt, 1990; Schmidt & Hunter, 1977), reliability generalization (RG) is a meta-analytic technique developed by Vacha-Haase (1998) to evaluate the sources of measurement error variance in scores across studies. Using RG, the mean measurement error across studies which have used the PAQ could be characterized, as well as the moderator variables that contributed to the variability of variance. Unlike a typical meta-analysis, which investigates validity by aggregating effect
sizes, RG focuses on the estimate of reliability reported by the primary studies to produce a measure of central tendency of the reliability of the data collected by the instrument across different studies, as well as the impact of the moderator variables identified on the magnitude and distribution of reliability estimates.
CHAPTER III

METHODS

Literature Search

Since all studies that would have used the Parental Authority Questionnaire would have cited Buri, I first conducted a cited reference search in the Web of Science (also known as Web of Knowledge) database as well as on Google Scholar. Then, using the search terms Buri, PAQ, and Parental Authority Questionnaire, I searched the following academic databases: Academic Search Premier, Annual Reviews, CINAHL, ERIC, EBSCO, Education Source, Family and Society StudiesWorldwide, Fuente Academica, HAPI (Hispanic American Periodicals Index) Online, Health and Psychosocial Instruments, JSTOR, Nursing and Allied Health Source, PAIS International, Proquest Dissertations and Theses, PsycEXTRA, PsycINFO, Science Direct, Sociological Abstracts, Scirus, NDLTD, Digital Archives of the British Library, TROVE, and PubMed. I limited my search to articles published after 1989, the year in which the PAQ was created, through 2015. The search was conducted in two phases. The initial phase was conducted in October 2013 and included studies published up to that point. A follow-up search was conducted in 2016 in order to capture any studies published after the initial search. The second search did not include Scirus, as that database was eliminated shortly after the initial search phase.
Selection of Studies

The following criteria were used to select the studies included in this reliability generalization: (a) the study must have the reliability of at least one of the six scales; (b) the PAQ must have been completed by offspring evaluating their parents’ authority styles, as this was Buri’s intention when he created the questionnaire, and all the items and item language remained intact; (c) the study must have been published in either English or Spanish (the two languages I could read and understand). That said, in the study review I tracked and identified those studies that were excluded from the RG for one or more of the following reasons: (a) reported no reliability, or no reliability other than Buri’s, or that of a previous study conducted by another researcher; (b) reported only a range of reliabilities; (c) included scales that had been altered in any way other than translation (e.g., parents self-reporting on their own parenting, or scales that have been altered to make language easier for younger offspring to understand; and (d) reported on a "composite" parent, the most dominant parent (unless that parent was identified as either mother or father), or other than a biological parent.

Institutional Review Board Approval

It should be noted that this study did not require the approval of the Institutional Review Board (IRB) because I was using data already published. A copy of the email confirming this from UNC IRB chair Maria Lehman is included as Appendix A.

Data Collected

For each qualifying study, I collected (if available) the reliabilities, means, and standard deviations for each of the scales (authoritarian, authoritative, and permissive) for both mothers and fathers separately. I also considered variables that might moderate the
effect and collected data on the percentage of male and female respondents, age of participants, ethnicity, country, language, and socio-economic status.

Data Analysis

Reliability generalization (RG) focuses on the estimate of reliability reported by the primary studies to produce a measure of central tendency of the reliability of the data collected by the instrument across different studies, as well as the impact of the moderator variables identified on the magnitude and distribution of reliability estimates. Since the Vacha-Haase’s (1998) development of reliability generalization (RG), the methodology has grown in both sophistication and complexity. Recently, a synthesis of the different recommended approaches to RG was published (Sánchez-Meca, López-López, & López-Pina, 2013). In this article, the authors recommended that “the main criterion for the model choice should be the extent to which the meta-analyst intends to generalize the results” (p. 402).

Meta-analytic Model

There were three possible models to consider: the fixed-effects model, the random-effects model, and the varying coefficients model.

**Fixed-effects model.** This model “assumes a priori that exactly the same $p$ (or $\delta$) [the population reliability] underlies all studies in the meta-analysis (i.e., $SDp = 0$)” (Hunter & Schmidt, 2004, p. 201) and that differences among studies were due only to sampling error. As discussed in Chapter II, there were a number of population characteristics (gender of respondent, language of respondent, and age of respondent) that would violate the underlying assumption ($SDp = 0$) that there would be no variation within the population reliability as a whole. In short, our sample populations were too
heterogeneous. There were just too many differences within the sample population for this assumption to be true.

**Random-effects (RE) model.** The random-effects model builds on the fixed-effects model but allows “for the possibility that the population parameters ($p$ or $\delta$ values) vary from study to study. A major purpose of random-effects models is to estimate this variance”(Hunter & Schmidt, 2004, p. 201). According to Sánchez-Meca et al. (2013), “the RE model assumes there is a hypothetical population of parametric coefficients alpha with a mean $\mu_\alpha$ and variance $\tau^2$ focusing on obtaining estimates of $\mu_\alpha$ and $\tau^2$” (p. 409). The purpose of this model is to be able to generalize the results to a broader superpopulation of hypothetical studies. That said, there were three main issues with this model that must be considered. First, the studies in the meta-analyses were not randomly selected from the population at large and, therefore, at least in theory, it was inappropriate to estimate the average population coefficient of the “superpopulation” of potential studies (Bonett, 2010; Sánchez-Meca et al., 2013). Second, when dealing with a small number of studies, estimates of the between studies variance were often inaccurate. This, in turn, affects the accuracy of the statistical analyses. Finally, since the normality assumption could not be met, then the assumption of a superpopulation of hypothetical reliability coefficients also could not be met (Sánchez-Meca et al., 2013). According to Sánchez-Meca et al., (2013),

The meta-analyst is unlikely to have sufficient information to assess the plausibility of many of these assumptions and the appropriateness of inferences made by primary study authors on the basis of coefficient alpha. (p. 308).

It should be noted that, even though Rodriguez and Maeda (2006) made this point eloquently, they advocated a random-effects model in their 2006 article regarding
meta-analyses of coefficient alpha, because at that time, Bonett (2010) had not yet published his ground-breaking article on the varying coefficients model.

**Varying coefficient (VC) model.** Like the fixed-effect model, this recently developed model advocated by Bonett (2010) attempts to generalize the results only to like populations within the studies. According to Sánchez-Meca et al. (2013), “The VC model is a kind of FE model where it is assumed that the coefficient obtained in each study is estimating different population reliability coefficients” (p. 408). The advantage of the VC model is that it does not rest on the somewhat unrealistic assumptions of the other models, including RE, (e.g., studies randomly selected from a normal hypothetical superpopulation of studies). Furthermore, using this model, linear regression can be used to analyze the effects of both categorical and continuous moderator variables. The disadvantage of this model is that results can only be generalized to like populations within the studies and not the superpopulation at large.

After careful consideration, it was decided to analyze the data using both a random-effects model and a varying coefficient model. While each model offered advantages and disadvantages individually, the advantage of using these models was that both a new, cutting-edge, more conservative approach to analyzing the data (VC) as well as a less conservative, more traditional approach (RE) could be employed. Moreover, this comparison offered a real-world contribution to the methodological approaches to reliability generalization.

**To Transform or Not to Transform**

One of the first questions a meta-analyst must answer when planning an RG study is whether or not to transform coefficients alpha. The answer, of course, depends upon
the model one chooses. Such transformations could usually be done in order to normalize
the distribution and stabilize variances. This issue would provide one of the first sources
of variability between different RG methods. Most often the Fisher $z$ transformation has
been employed to transform reliability coefficients computed as Pearson correlations.
However, for transforming coefficient alpha both Rodriguez and Maeda (2006) and
Hakstian and Whalen (1976) recommended a transformation procedure dubbed the
Hakstian-Whalen transformation which uses the cubed root of the variance:

$$ T_i = 3\sqrt{1 - \hat{\alpha}_i}. $$

This transformation is based on the $F$ statistic. Although Rodriguez and
Maeda (2006) pointed out that, like Fisher $z$, this transformation was also biased, “the
bias in estimates of coefficient alpha are less than that of Fisher $z$” (p. 309).

Thus, the Hakstian-Whalen transformation was employed in the RE model, as
advocated by Rodriguez and Maeda but not in the VC model because Bonett favored a
transformation that normalized the distribution of coefficients alpha while stabilizing
their variances. According to Bonett (2010),

> The use of the Hakstian-Whalen transformation in meta-analysis of alpha
> reliabilities is analogous to the use of Fisher-transformation correlations in the
> meta-analysis of Pearson correlations. . . . The weighted average of the Hakstian-
> Whalen-transformed reliabilities is then reverse transformed to give and estimator
> of $p$. (p. 369)

This reverse transformation added to the bias. Instead, for the VC model, the
formula advocated by Bonett (2010) was employed to transform the coefficients alpha:

$$ L_i = \ln (q - |\hat{\alpha}_i|). $$

**To Weight or Not to Weight**

In any given meta-analysis, there were studies with varying sample sizes.
Researchers must decide how to balance the results of studies with small sample sizes
against those with larger sample sizes, as treating the results of the smaller studies in the same manner as the larger studies may skew the results (Hunter & Schmidt, 2004). There were four weighting methods that could be applied in RG studies: (a) unweighted and applying an ordinary least squares technique; (b) weighting by the inverse variance in a fixed-effect model; (c) weighting by the inverse variance in a random-effects model; and, (d) weighting by sample size (Sánchez-Meca et al., 2013).

In the RE model, the weighting was designed to take into account both within study variance as well as between-study variance (Sánchez-Meca et al., 2013). For this investigation the inverse variance method: \( W_{RE}^i = \frac{1}{\sqrt{V_i + z^2}} \) for the RE model was employed (Sánchez-Meca et al., 2013). No weighting method was used for VC model because, according to Bonett (2010):

It can be shown that the weighted average is a biased estimator of \( p \) when the \( m \) weights (which are functions of sample sizes) are unequal and the constant coefficient assumption…does not hold. In typical meta-analytic studies, the sample sizes are unequal across studies and the constant coefficient assumption is unrealistic. (p. 369)

**Steps in the Reliability Generalization**

**Meta-analytic for the Random-effects Model**

The following meta-analytic procedures were drawn from those recommended by Sánchez-Meca et al. (2013)

Step 1. Transform the reliability coefficients:

\[ T_i = \sqrt{i} \sqrt{1 - \hat{\alpha}_i} \]

where \( \hat{\alpha}_i \) is the coefficient alpha for the \( i^{th} \) study.

Step 2. Conduct a test of homogeneity. This test takes into account within-study differences by comparing the squared differences between the observed and weighted
average (Campbell, Pulos, Hogan, & Murry, 2005): $Q = \sum_i W_i^{FE} (y_i - \bar{y}_{FE})^2$

where $V(y_i)$ is the sampling variance for the $i^{th}$ study and $W_i^{FE} = \frac{1}{V(y_i)}$. (Recall that the RE model builds on the FE model, therefore, the RE models uses the weights that would have been assigned in the fixed-effect model to determine heterogeneity and between studies variance.

Step 3. Estimate between-studies variance: $\tau^2 = \frac{Q - (k - 1)}{c}$ where $k$ is the number of studies and

$$c = \sum_i W_i^{FE} - \frac{\Sigma_i (0W_i^{FE})^2}{\Sigma_i W_i^{FE}}$$

Step 4. Weight the results by the inverse variance:

$$W_i^{RE} = \frac{1}{V(y_i + \tau^2)}.$$

Step 5. Compute the average reliability estimate ($\bar{Y}$) using the transformed coefficients:

$$\bar{Y}_{RE} = \frac{\Sigma_i W_i^{RE} Y_i}{\Sigma_i W_i^{RE}}.$$

Step 6. Compute the sampling variance of $\bar{Y}$:

$$V(\bar{Y}_{RE}) = \frac{1}{\Sigma_i W_i^{RE}}$$

Step 7. Compute the vector of the regression coefficients with $p - 1$ predictors as follows: $b_{WLS} = (X'WX)^{-1} X'WY$.

Step 8. Create the variance-covariance matrix of regression coefficients:

$$\sum b_{WLS} = (X'WX)^{-1}.$$

$$z_j = \frac{b_j}{\sqrt{V(b_j)}}.$$

Step 9. Perform the test of statistical significance:
Step 10. Compute $R^2$ (the proportion of variance in the reliability coefficients explained by the moderators):

$$R^2 = \frac{\hat{\tau}^2 - \hat{\tau}_{res}^2}{\hat{\tau}^2}$$

where $\hat{\tau}_{res}^2 = \frac{y'Py - (k - p - 1) \cdot tr(P)}{tr(P)}$ and \( P = W - WX(XWX)'X'W \) and where \( W \) is the diagonal \( k \times k \) matrix of the weights of all the studies: \( \{W_i^{1FE}\} \).

Steps in the Varying Coefficients Alpha Model

Step 1. Transform coefficient alpha: \( L_i = \ln(\xi - 1\bar{\alpha}_i) \).

Step 2. Compute the average reliability estimate:

$$\bar{\alpha}_\mu = \frac{\Sigma_i \bar{\alpha}_i}{k}$$

where \( k \) is the number of studies.

Step 3. Compute the sampling variance:

$$V(\bar{\alpha}_\mu) = \frac{\Sigma_i V(\bar{\alpha}_i)}{k^2}$$

Step 4. Create the vector of regression coefficients with \( p - 1 \) predictors:

$$b_{OLS} = (X'X)^{-1}X'y.$$ 

Step 5. Created the variance-covariance matrix of the regression coefficients:

$$\sum b_{VC} = (X'X)^{-1}X'VX(X'X)^{-1}.$$ 

$$z_j = \frac{b_j}{\sqrt{V(b_j)}}.$$ 

Step 6. Perform the test of statistical significance:

$$R^2_{adj} = 1 - \left(1 - \tau_\lambda^2\right)\frac{k-1}{k-2}.$$

It should be noted that, due to the number of statistical tests performed simultaneously on this small data set, a Bonferroni correction was necessary to avoid Type I Error (finding a significant difference where none truly exists; Glass & Hopkins, 1996).
Comparing the Results of the Two Different Models

After reviewing the results of the two different models, I then compared the results of each model to determine if there were noteworthy differences in the results. Little has been known about how the methods compare with actual data. An examination of all studies that cited the seminal paper on VC in the meta-analysis of alpha (Bonett, 2010) failed to find any comparison of methods with actual data. Even though the methods have been quite different in methods and assumptions, it could be quite possible that the differences have been trivial from a substantive rather than a methodology perspective. Specifically, I compared the confidence intervals and the ability to detect moderator variables.
CHAPTER IV

RESULTS

In this chapter, I have discussed the results of the study review process and outlined the myriad of methodological issues which led to study exclusion, as well as the methodological issues inherent in many of the studies included in the meta-analysis. The results of the random-effect model have been highlighted, discussed, and included the estimated scale reliability results and the identification of significant moderating effects. The results of the vary-coefficient model have also been presented and have also included estimated scale reliabilities and the identification of the significant moderating effects.

Data Extraction

A search of the databases identified in Chapter III yielded a list of 1,577 journal articles, dissertations, theses, unpublished papers, and government reports. Of those, 622 were duplicate entries. In total, I reviewed 995 separate studies. Only 548 studies actually used the Parental Authority Questionnaire (PAQ). Each study was evaluated to determine if it met the inclusion criteria as identified in Chapter III: (a) the study must have the reliability of at least one of the six scales; (b) the PAQ must have been completed by offspring evaluating their parents’ authority styles and all the items and item language remained intact; and (c) the study must have been published in either English or Spanish. It should be noted that I could not locate or access 33 of the studies on the list, and 43 studies were published in a language other than English or Spanish.
Of the remaining 548 studies, 162 were excluded due to issues regarding the reported reliabilities or lack thereof. Researchers in 90 studies only reported Buri’s (1991) reliabilities (reliability induction), 2 reported reliabilities cited by other researchers (reliability induction), 11 reported a range of reliabilities for the scales used instead of reliabilities for each of the 6 scales, 19 reported the reliability of the overall instrument, 12 reported reliabilities for the combined mother and then father scales, and 28 reported no reliabilities at all.

Authors of 160 studies changed the items or language of the scales in one or more ways. In 74 studies, items were adapted so that parents could report on their own parental authority while 8 other studies adapted the study to report on the parental unit (e.g., mothers and fathers combined, grandparents, or other caregivers). In 60 studies, the wording of items was modified and 18 studies used a shortened version of the PAQ.

Methodological issues and/or missing data led to the exclusion of 57 studies. Of those, authors of 45 studies reported no data or results, 6 reported data that was unclear or confusing, and at least 6 studies reported reliability data already published in a previous study by the same researcher.

Of the remaining 128 studies that were to be included initially, upon closer inspection, 59 were excluded for one or more of the reasons cited above. In the end, 69 studies were included in the reliability generalization. It should be noted that 2 studies (Kawamura, Frost, & Harmatz, 2002; Newman et al., 2015) included more than 1 independent data set, thus yielding 72 discrete sets of data for analyzation.
**Random-effect Model Results**

I first conducted the reliability generalization using the random-effect model in order to answer the two research questions of this investigation: (a) Does the Parental Authority Questionnaire (PAQ) produce reliable information across groups when used by adolescent, young-adult or adult children to evaluate their parents’ parenting styles and (b) What moderator variables can be identified that might affect reliability.

As expected, in each analysis, the tests for heterogeneity were significant at the .0001 level, indicating that the variability in the data was due to more than sampling error. It should be noted that, due to the number of statistical tests performed simultaneously on this small data set, a Bonferroni correction was necessary to avoid Type I Error (finding a significant difference where none truly exists; Glass & Hopkins, 1996).

The reliability estimates for each of the 6 scales are presented in Table 1: mother authoritarian (.8318), mother authoritative (.8381), mother permissive (.7477), father authoritarian (.8667), father authoritative (.8629), father permissive (.7595). Not all studies used all six scales. The analysis also showed that there were differences in reliabilities depending upon the scale with the permissive scale exhibiting lower reliability than the authoritarian and authoritative scales. Likewise, reliabilities were lower when participants were rating the parental authority of their mothers versus their fathers. Table 1 also provides the ranges of the confidence intervals and the minimum and maximum reliability reported per scale. For the mother authoritarian scale, a confidence interval of .8158 to .8470 indicated that researchers could predict with 95% confidence that the actual reliability of this scale (for the studies similar to those in this
investigation) lied somewhere between those two numbers. The confidence intervals on the remaining scales were: mother authoritative (.8197-.8567), mother permissive (.7273-.7670), father authoritarian (.8482-.8835), father authoritative (.8460-.8785), and father permissive (.7424-.7758).

Table 1

Random-effect Model Reliability Estimates by Scale

<table>
<thead>
<tr>
<th>Scale</th>
<th># of Data Sets</th>
<th>Alpha</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mother Authoritarian</td>
<td>67</td>
<td>0.8318</td>
<td>0.8158</td>
<td>0.847</td>
<td>0.57</td>
<td>0.91</td>
</tr>
<tr>
<td>Mother Authoritative</td>
<td>62</td>
<td>0.8389</td>
<td>0.8197</td>
<td>0.8567</td>
<td>0.54</td>
<td>0.99</td>
</tr>
<tr>
<td>Mother Permissive</td>
<td>61</td>
<td>0.7477</td>
<td>0.7273</td>
<td>0.7670</td>
<td>0.48</td>
<td>0.86</td>
</tr>
<tr>
<td>Father Authoritarian</td>
<td>58</td>
<td>0.8667</td>
<td>0.8482</td>
<td>0.8835</td>
<td>0.6</td>
<td>0.94</td>
</tr>
<tr>
<td>Father Authoritative</td>
<td>57</td>
<td>0.8629</td>
<td>0.846</td>
<td>0.8785</td>
<td>0.63</td>
<td>0.95</td>
</tr>
<tr>
<td>Father Permissive</td>
<td>54</td>
<td>0.7595</td>
<td>0.7424</td>
<td>0.7758</td>
<td>0.53</td>
<td>0.86</td>
</tr>
</tbody>
</table>

Moderator Analysis

Although data were collected on a number of possible moderators, due to limitations of the data available, I was only able to perform analyses which investigated the effect of seven different variables (scale mean, scale variance, percent males, language of the instrument/culture, developmental level of the participant, publication type, and year published). It should be noted that, for the continuous variables reported by the study authors, not every study reported each of the variables. Table 2 summarizes the number of studies reporting these variables by scale.
Of the four categorical variables, two (publication type and year of study) were inherently included in every study, language the questionnaire was administered in was reported by all studies, and developmental level was imputed by me based on either mean age or, if that statistic was unavailable, then grade level. Due to limitations in the data (low frequencies in many of the categories), each of the four categorical variables was coded as a dichotomous variable yielding: (a) In English versus Not in English; (b) Under Age 20 versus 20 and Older; (c) Dissertations and Theses versus Journal Articles; and (d) Before 2000 versus 2000 and later.

**Test of Scale Mean as Moderator Variable**

With the exception of the mother authoritarian scale, little more than half of the studies reporting scale reliability also reported scale mean (please refer to Table 2 in the previous section). The analysis of those studies that did include scale mean (see Table 3) showed no significant moderating effect of the mean on scale reliability with \( p \)-values on
the Omnibus Test of Coefficients (QM) ranging from .0026 to 4.0055. In this investigation, no significant effect of scale mean on scale reliability could be detected.

Table 3

*Random-effect Model Test of Scale Mean as Moderator Variable*

<table>
<thead>
<tr>
<th>Scale</th>
<th># of Data Sets</th>
<th>p-value*</th>
<th>Estimate</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mother Authoritarian</td>
<td>40</td>
<td>0.0454</td>
<td>0.0217</td>
<td>0.0006</td>
<td>0.0426</td>
</tr>
<tr>
<td>Mother Authoritative</td>
<td>36</td>
<td>0.1968</td>
<td>-0.0188</td>
<td>-0.0467</td>
<td>0.0099</td>
</tr>
<tr>
<td>Mother Permissive</td>
<td>35</td>
<td>0.1845</td>
<td>-0.0149</td>
<td>-0.0364</td>
<td>0.0072</td>
</tr>
<tr>
<td>Father Authoritarian</td>
<td>33</td>
<td>0.5168</td>
<td>0.0057</td>
<td>-0.0016</td>
<td>0.0232</td>
</tr>
<tr>
<td>Father Authoritative</td>
<td>32</td>
<td>0.2413</td>
<td>-0.0176</td>
<td>-0.0466</td>
<td>0.0119</td>
</tr>
<tr>
<td>Father Permissive</td>
<td>29</td>
<td>0.9597</td>
<td>0.0018</td>
<td>-0.0021</td>
<td>0.0054</td>
</tr>
</tbody>
</table>

* Test of significance were performed on transformed variable

**Test of Scale Variance as a Moderator Variable**

Little more than half the studies reporting scale reliability also reported scale variance. The analysis of those studies that did include scale variance showed a significant moderating effect of scale variance on reliability for both the mother authoritative and father authoritative scales with p-values on the Omnibus Test of Coefficients (QM) of less than .0001 (see Table 4). For the mother authoritative scale, the regression analysis indicated that for every .0108 increase in slope reliability rose by 1%. Likewise, for the Father Authoritative scale, every .0063 increase in slope, reliability rose by 1%. For the other four scales, there was no significant moderating effect of variance on scale reliability with QM p-values ranging from .0769 to .4829. Because scale
reliability is a measure of consistency of responses, and variability can be interpreted as a measure of inconsistency of responses—the greater the variance, the more inconsistent the responses were—then it should be expected that scale variance would impact reliability.

The fact that scale variance was a significant moderating variable indicated that there was a higher degree of variability in scale reliabilities from one study to another,

Table 4

*Random-effect Model Test of Scale Variance as Moderator Variable*

<table>
<thead>
<tr>
<th>Scale</th>
<th># of Data Sets</th>
<th>p-value*</th>
<th>Estimate</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mother Authoritarian</td>
<td>37</td>
<td>0.3787</td>
<td>0.0018</td>
<td>-0.0021</td>
<td>0.0054</td>
</tr>
<tr>
<td>Mother Authoritative</td>
<td>34</td>
<td>**</td>
<td>0.0108</td>
<td>0.0078</td>
<td>0.0137</td>
</tr>
<tr>
<td>Mother Permissive</td>
<td>33</td>
<td>0.2375</td>
<td>-0.0003</td>
<td>-0.0009</td>
<td>0.0003</td>
</tr>
<tr>
<td>Father Authoritarian</td>
<td>32</td>
<td>0.0769</td>
<td>0.0027</td>
<td>-0.0003</td>
<td>0.006</td>
</tr>
<tr>
<td>Father Authoritative</td>
<td>31</td>
<td>**</td>
<td>0.0063</td>
<td>0.0039</td>
<td>0.009</td>
</tr>
<tr>
<td>Father Permissive</td>
<td>28</td>
<td>0.4829</td>
<td>0</td>
<td>-0.0003</td>
<td>0.0006</td>
</tr>
</tbody>
</table>

*p* Test of significance were performed on transformed variable; **Significant at .0001

*Test of Percent Male as a Moderator Variable*

Most of the studies reported demographic information that indicated the percentage of participants who were male. The percentage of male participants varied across studies with a mean of 35% and a median of 38%. Four studies included no male participants and one study included all males.

The analysis of those studies (see Table 5) showed no significant moderating effect of gender of participants on scale reliability with *p*-values on the Omnibus Test of
Coefficients (QM) ranging from .0251 to .4.9529. In this investigation, no significant effect of percent male on scale reliability could be detected.

Table 5

*Random-effect Model Test of Percent Male as Moderator Variable*

<table>
<thead>
<tr>
<th>Scale</th>
<th># of Data Sets</th>
<th>p-value*</th>
<th>Estimate</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mother Authoritarian</td>
<td>61</td>
<td>0.3968</td>
<td>0.0015</td>
<td>-0.8085</td>
<td>0.0018</td>
</tr>
<tr>
<td>Mother Authoritative</td>
<td>56</td>
<td>0.0235</td>
<td>-0.0024</td>
<td>0.0063</td>
<td>0.0015</td>
</tr>
<tr>
<td>Mother Permissive</td>
<td>55</td>
<td>0.0260</td>
<td>-0.0036</td>
<td>-0.0066</td>
<td>0.0036</td>
</tr>
<tr>
<td>Father Authoritarian</td>
<td>55</td>
<td>0.8740</td>
<td>-0.0003</td>
<td>-0.0045</td>
<td>0.0036</td>
</tr>
<tr>
<td>Father Authoritative</td>
<td>54</td>
<td>0.5419</td>
<td>-0.0011</td>
<td>-0.0004</td>
<td>0.0024</td>
</tr>
<tr>
<td>Father Permissive</td>
<td>51</td>
<td>0.7593</td>
<td>-0.0001</td>
<td>-0.003</td>
<td>0.0024</td>
</tr>
</tbody>
</table>

* Test of significance were performed on transformed variable

Test of Language of Instrument as Moderator Variable

The majority of studies (57) included in this reliability generalization administered the Parental Authority Questionnaire (PAQ) in English. In the 15 remaining studies, the PAQ was administered in 10 different languages: Arabic (3), Bahasa (1), Chinese (3), Gujarati (1), Hebrew (1), Indonesian (1), Malay (1), Pakistani (1), Turkish (2), and Urdu (1). Due to the low frequencies, a dichotomous variable was created: English and Not in English. The analysis of this variable showed that the language in which the PAQ was administered had a significant moderating effect on the reliability of the instrument for each of the six scales. Although the results of the random-effect model showed that each of the scales was reliable when administered in English or another
language, the Omnibus Test of Coefficients indicated that there were significant differences, with $p$-values < .0001, in the reliability of each scale depending upon the language of the instrument. These differences are summarized in Table 6.

Note that the confidence intervals were larger for the reliabilities of studies not administered in English. The confidence intervals indicated the range of reliabilities in which one could predict with 95% certainty that the true reliabilities of the scales (for studies similar to those in this investigation) lay. Thus, for example, while the estimated alpha for the mother authoritarian scale administered in a language other than English was .7345, one could predict with 95% certainty that the actual reliability of the mother authoritarian scale in this investigation lay somewhere between .7148 and .8094. Likewise, while the estimated alpha for the mother authoritarian scale administered in English was .8510, one could predict with 95% confidence that the true reliability (what the alpha would have been without error) of the mother authoritarian scale in this investigation lied somewhere between .8371 and .8722. The larger the range of the confidence interval, the less precisely the true reliability could be predicted.
Table 6

**Random-effect Model Test of Instrument Language as a Moderator Variable**

<table>
<thead>
<tr>
<th>Scale</th>
<th># of Data Sets</th>
<th>Language</th>
<th>Alpha</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mother Authoritarian</td>
<td>67</td>
<td>Not in English</td>
<td>0.7345</td>
<td>0.7148</td>
<td>0.8094</td>
</tr>
<tr>
<td></td>
<td></td>
<td>English</td>
<td>0.8510</td>
<td>0.8371</td>
<td>0.8722</td>
</tr>
<tr>
<td>Mother Authoritative</td>
<td>62</td>
<td>Not in English</td>
<td>0.7653</td>
<td>0.7148</td>
<td>0.8095</td>
</tr>
<tr>
<td></td>
<td></td>
<td>English</td>
<td>0.8554</td>
<td>0.8371</td>
<td>0.8722</td>
</tr>
<tr>
<td>Mother Permissive</td>
<td>61</td>
<td>Not in English</td>
<td>0.6334</td>
<td>0.582</td>
<td>0.7518</td>
</tr>
<tr>
<td></td>
<td></td>
<td>English</td>
<td>0.7691</td>
<td>0.6803</td>
<td>0.7855</td>
</tr>
<tr>
<td>Father Authoritarian</td>
<td>58</td>
<td>Not in English</td>
<td>0.7394</td>
<td>0.6911</td>
<td>0.7824</td>
</tr>
<tr>
<td></td>
<td></td>
<td>English</td>
<td>0.8890</td>
<td>0.8760</td>
<td>0.9008</td>
</tr>
<tr>
<td>Father Authoritative</td>
<td>57</td>
<td>Not in English</td>
<td>0.7728</td>
<td>0.7274</td>
<td>0.8128</td>
</tr>
<tr>
<td></td>
<td></td>
<td>English</td>
<td>0.8810</td>
<td>0.8660</td>
<td>0.8932</td>
</tr>
<tr>
<td>Father Permissive</td>
<td>54</td>
<td>Not in English</td>
<td>0.6702</td>
<td>0.6267</td>
<td>0.7100</td>
</tr>
<tr>
<td></td>
<td></td>
<td>English</td>
<td>0.7757</td>
<td>0.7607</td>
<td>0.789</td>
</tr>
</tbody>
</table>

* Significant at .0001 level
Test of Developmental Level as a Moderator Variable

In the group of studies that was included in this investigation, the PAQ was administered to adolescents, young-adults, and adults. The average age of participants ranged from 12.28 to 49.81. There were 8 studies reporting the majority of participants under the age of 16, 26 studies reporting the majority of participants 16 and older but under 20 years old, 23 studies reporting the majority of participants 20 years old and older but under 35, and 8 studies reporting the majority of participants over the age of 35. Due to the low frequencies, the 4 categories were collapsed into 2 (participants under age 20 and participants age 20 and older).

Of the 36 studies in which the PAQ was administered to participants under 20 years old, 16 were published prior to 2000, and in 11 instances the survey was not administered in English. Of the other half of the studies in which the PAQ was administered to participants age 20 years old or older, 15 were conducted prior to 2000 and 20 were conducted in 2000 or later. For those studies with participants 20 years old or older, the survey was administered in a language other than English in 25 of the 36 studies.

The analysis of this dichotomous variable showed that developmental level had a significant moderating effect on the reliability of the instrument for each of the six scales. Although the results of the analysis showed that each of the scales was reliable when administered to participants in both age groups, the Omnibus Test of Coefficients indicated that there were significant differences, with \( p \)-values < .0001, in the reliability of each scale depending upon whether the participants were younger than 20 years old or 20 and older. These differences are summarized in Table 7.
### Table 7

**Random-effect Model Test of Developmental Level as a Moderator Variable**

<table>
<thead>
<tr>
<th>Scale</th>
<th># of Data Sets</th>
<th>p-value*</th>
<th>Age</th>
<th>Alpha</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mother Authoritarian</td>
<td>67</td>
<td>*</td>
<td>Under 20</td>
<td>0.8149</td>
<td>0.7900</td>
<td>0.8378</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>20 and up</td>
<td>0.8459</td>
<td>0.8253</td>
<td>0.8648</td>
</tr>
<tr>
<td>Mother Authoritative</td>
<td>62</td>
<td>*</td>
<td>Under 20</td>
<td>0.8169</td>
<td>0.7875</td>
<td>0.8435</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>20 and up</td>
<td>0.8581</td>
<td>0.8339</td>
<td>0.8798</td>
</tr>
<tr>
<td>Mother Permissive</td>
<td>61</td>
<td>*</td>
<td>Under 20</td>
<td>0.721</td>
<td>0.6889</td>
<td>0.751</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>20 and up</td>
<td>0.7668</td>
<td>0.7419</td>
<td>0.7901</td>
</tr>
<tr>
<td>Father Authoritarian</td>
<td>58</td>
<td>*</td>
<td>Under 20</td>
<td>0.8416</td>
<td>0.8143</td>
<td>0.866</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>20 and up</td>
<td>0.8906</td>
<td>0.8683</td>
<td>0.9102</td>
</tr>
<tr>
<td>Father Authoritative</td>
<td>57</td>
<td>*</td>
<td>Under 20</td>
<td>0.8409</td>
<td>0.817</td>
<td>0.8626</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>20 and up</td>
<td>0.8865</td>
<td>0.8655</td>
<td>0.9052</td>
</tr>
<tr>
<td>Father Permissive</td>
<td>54</td>
<td>*</td>
<td>Under 20</td>
<td>0.7442</td>
<td>0.7198</td>
<td>0.7673</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>20 and up</td>
<td>0.7747</td>
<td>0.7516</td>
<td>0.7963</td>
</tr>
</tbody>
</table>

* Significant at .0001 level
Note that, while the confidence intervals were larger for the reliabilities of the mother authoritarian, mother permissive and father authoritarian scale for participants under age 20, the ranges for the other scales were not disparate. Larger confidence intervals indicate less precise estimates than more narrow ones.

**Test of Publication Type as a Moderator Variable**

Although the search process identified all types of published and unpublished studies (journal articles, dissertations, theses, unpublished papers, and government reports), only 26 dissertations and theses and 46 journal articles contained adequate data to be included in this investigation. The PAQ was not administered in English in 3 of the dissertations/theses studies and 12 of the studies published in journal articles. Thirteen of the dissertations/theses studies were conducted prior to 2000 and 24 of the studies published in journal articles were conducted prior to the year 2000.

The analysis of this variable showed that publication type had a significant moderating effect on scale reliabilities with $p$-values < .0001 on the Omnibus Test of Coefficients. For each parental authority scale, reliabilities were higher for dissertation or thesis research than for research published in journal articles. Although the range of the confidence intervals were greater for the father authoritative and father authoritarian scales for the studies published in journal articles, the range of confidence intervals on the other scales were not disparate. The results of this analysis are summarized in Table 8.
Table 8

Random-effect Model Test of Publication Type as a Moderator Variable

<table>
<thead>
<tr>
<th>Scale</th>
<th>Total K</th>
<th>p-value*</th>
<th>Publication Type</th>
<th>Alpha</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mother Authoritarian</td>
<td>67</td>
<td>**</td>
<td>Dissertation/Thesis</td>
<td>0.8477</td>
<td>0.8233</td>
<td>0.8697</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Journal Article</td>
<td>0.8213</td>
<td>0.7998</td>
<td>0.8410</td>
</tr>
<tr>
<td>Mother Authoritative</td>
<td>62</td>
<td>**</td>
<td>Dissertation/Thesis</td>
<td>0.8609</td>
<td>0.8330</td>
<td>0.8856</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Journal Article</td>
<td>0.8238</td>
<td>0.7981</td>
<td>0.8472</td>
</tr>
<tr>
<td>Mother Permissive</td>
<td>61</td>
<td>**</td>
<td>Dissertation/Thesis</td>
<td>0.7683</td>
<td>0.7290</td>
<td>0.7954</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Journal Article</td>
<td>0.7315</td>
<td>0.7036</td>
<td>0.7575</td>
</tr>
<tr>
<td>Father Authoritarian</td>
<td>58</td>
<td>**</td>
<td>Dissertation/Thesis</td>
<td>0.8896</td>
<td>0.864</td>
<td>0.9118</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Journal Article</td>
<td>0.8498</td>
<td>0.8248</td>
<td>0.8724</td>
</tr>
<tr>
<td>Father Authoritative</td>
<td>57</td>
<td>**</td>
<td>Dissertation/Thesis</td>
<td>0.8853</td>
<td>0.8615</td>
<td>0.9061</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Journal Article</td>
<td>0.8474</td>
<td>0.8248</td>
<td>0.8678</td>
</tr>
<tr>
<td>Father Permissive</td>
<td>54</td>
<td>**</td>
<td>Dissertation/Thesis</td>
<td>0.7717</td>
<td>0.7460</td>
<td>0.7954</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Journal Article</td>
<td>0.7502</td>
<td>0.7270</td>
<td>0.7720</td>
</tr>
</tbody>
</table>

* Test of significance were performed on transformed variable; ** Significant at .0001 level
Test of Publication Year as a Moderator Variable

The Parental Authority Questionnaire (PAQ) has been used continuously to assess parental authority since its development by Buri (1991). Of the studies included in this investigation, 36 were published prior to 2000 and 35 were published in 2000 or later. In 3 of the 36 studies published prior to 2000, the PAQ was administered in a language other than English. In 12 of the 35 studies published 2000 or later, the PAQ was administered in a language other than English.

The analysis of this variable showed that year of publication had a significant moderating effect on scale reliabilities with $p$-values $< .0001$ on the Omnibus Test of Coefficients. Studies published prior to 2000 had higher scale reliabilities than those published after 2000. With the exception of the mother authoritative scale, the confidence intervals were larger on the other five scales for studies published after 2000. A summary of the analysis is presented in Table 9.

The Varying Coefficient Model Results

The results of the varying coefficient model indicated that all six of the scales of the Parental Authority Questionnaire (PAQ) were reliable: mother authoritarian (.8225), mother authoritative (.8278), mother permissive (.7389), father authoritarian (.8536), father authoritative (.8528), and father permissive (.7538). A summary of the results is presented in Table 10. It was important to note that each of the studies did not use all six scales. The result of the analysis also showed that there were differences in reliabilities depending upon the scale with the permissive scale exhibiting lower reliability than the authoritarian and authoritative scales. Likewise, reliabilities were lower when participants were rating the parental authority of their mothers versus their fathers.
Table 9

*Random-effect Model Test of Year Published as a Moderator Variable*

<table>
<thead>
<tr>
<th>Scale</th>
<th>Total K</th>
<th>$p$-value*</th>
<th>Year Published</th>
<th>Alpha</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mother Authoritarian</td>
<td>67</td>
<td>**</td>
<td>Before 2000</td>
<td>0.8433</td>
<td>0.8219</td>
<td>0.8713</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2000 and Later</td>
<td>0.8189</td>
<td>0.7943</td>
<td>0.8538</td>
</tr>
<tr>
<td>Mother Authoritative</td>
<td>62</td>
<td>**</td>
<td>Before 2000</td>
<td>0.8487</td>
<td>0.8234</td>
<td>0.8714</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2000 and Later</td>
<td>0.8265</td>
<td>0.7961</td>
<td>0.8538</td>
</tr>
<tr>
<td>Mother Permissive</td>
<td>61</td>
<td>**</td>
<td>Before 2000</td>
<td>0.7593</td>
<td>0.7313</td>
<td>0.7353</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2000 and Later</td>
<td>0.7353</td>
<td>0.7050</td>
<td>0.7634</td>
</tr>
<tr>
<td>Father Authoritarian</td>
<td>58</td>
<td>**</td>
<td>Before 2000</td>
<td>0.8807</td>
<td>0.8582</td>
<td>0.9007</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2000 and Later</td>
<td>0.8467</td>
<td>0.8160</td>
<td>0.8737</td>
</tr>
<tr>
<td>Father Authoritative</td>
<td>57</td>
<td>**</td>
<td>Before 2000</td>
<td>0.8720</td>
<td>0.8505</td>
<td>0.8218</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2000 and Later</td>
<td>0.8498</td>
<td>0.7115</td>
<td>0.8747</td>
</tr>
<tr>
<td>Father Permissive</td>
<td>54</td>
<td>**</td>
<td>Before 2000</td>
<td>0.7745</td>
<td>0.7533</td>
<td>0.7945</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2000 and Later</td>
<td>0.7385</td>
<td>0.7115</td>
<td>0.7638</td>
</tr>
</tbody>
</table>

* Test of significance were performed on transformed variable; ** Significant at .0001 level
Table 10

_Varying Coefficient Model Reliability Estimates by Scale_

<table>
<thead>
<tr>
<th>Scale</th>
<th>Alpha</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mother Authoritarian</td>
<td>0.8225</td>
<td>0.8185</td>
<td>0.8282</td>
</tr>
<tr>
<td>Mother Authoritative</td>
<td>0.8278</td>
<td>0.8239</td>
<td>0.8333</td>
</tr>
<tr>
<td>Mother Permissive</td>
<td>0.7389</td>
<td>0.7329</td>
<td>0.7475</td>
</tr>
<tr>
<td>Father Authoritarian</td>
<td>0.8536</td>
<td>0.8498</td>
<td>0.8586</td>
</tr>
<tr>
<td>Father Authoritative</td>
<td>0.8528</td>
<td>0.8493</td>
<td>0.8576</td>
</tr>
<tr>
<td>Father Permissive</td>
<td>0.7538</td>
<td>0.7479</td>
<td>0.7619</td>
</tr>
</tbody>
</table>

The confidence intervals for each of the scales are also presented in Table 10. For the mother authoritarian scale, the confidence interval was .8185 to 8282 which indicated that one could predict with 95% confidence that the actual reliability of the mother authoritarian scale for this group of studies fell within that range. The confidence intervals for the other scales were equally as narrow: mother authoritative (.8239-.8333), mother permissive (.7329-.7475), father authoritarian (.8498-.8586), father authoritative (.8493-.8576), and father permissive (.7479-.7619).

**Analysis of Moderator Variables**

**Continuous Variables**

The regressions employed in the Varying Coefficient Model to look at moderator variables were different from those employed in the random-effect model. Bonett (2008, 2010), who developed the method, criticized the mixed-effect regression model as being
biased and requiring assumptions that were difficult to support in the real world. He proposed an alternative regression model based on ordinary least squares which employed log-complement transformation to stabilize the variance and approximate normal distribution with when samples were small. It was important to note that this procedure differed from the regressions performed in the random-effect model in that: (a) the dependent variable was nonreliability, as opposed to reliability, and (b) the slope of the regression was multiplicative not additive. Thus, the slope referred to the multiplicative change in nonreliability for every increase in the predictor variable.

**Test of Scale Mean as a Moderator Variable**

Results of the logistic regression analysis of scale mean showed that the mean had a significant moderating effect on scale reliabilities with the exception of the father permissive scale. A summary of the results is presented in Table 11. Negative estimate results indicated that, as the mean goes up, nonreliability goes down, and positive estimate results indicated that, as the mean goes up so does nonreliability. Thus, for example, as the mean of the authoritarian scale increased by one point, the nonreliability decreased (thus, reliability increased) by a multiplicative factor of .0374. Alternatively, for the mother authoritative scale, as the mean increased by 1 point, the nonreliability of the scale increased by a multiplicative factor of .0541. With the exception of the mother authoritarian and father authoritarian scales (and the father permissive scale as discussed above), an increase in the mean resulted in a decrease in reliability. The confidence intervals in this logistic regression referred to the range of slopes within which the actual slope of the effect of variance could be predicted to fall with 95% accuracy.
Table 11

*Varying Coefficient Model with Scale Mean as a Moderator Variable*

<table>
<thead>
<tr>
<th>Scale</th>
<th>Estimate</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mother Authoritarian</td>
<td>-0.0347</td>
<td>-0.046</td>
<td>-0.0234</td>
<td>*</td>
</tr>
<tr>
<td>Mother Authoritative</td>
<td>0.0541</td>
<td>0.0456</td>
<td>0.0625</td>
<td>*</td>
</tr>
<tr>
<td>Mother Permissive</td>
<td>0.0248</td>
<td>0.0138</td>
<td>0.0358</td>
<td>*</td>
</tr>
<tr>
<td>Father Authoritarian</td>
<td>-0.0079</td>
<td>-0.0157</td>
<td>-0.0003</td>
<td>*</td>
</tr>
<tr>
<td>Father Authoritative</td>
<td>0.0398</td>
<td>0.0292</td>
<td>0.0485</td>
<td>*</td>
</tr>
<tr>
<td>Father Permissive</td>
<td>0.0038</td>
<td>-0.0075</td>
<td>0.0151</td>
<td>0.5105</td>
</tr>
</tbody>
</table>

* Significant at .0001 level

**Test of Scale Variance as a Moderator Variable**

Results of the logistic regression indicated that scale variance had a significant moderating effect on scale reliabilities for all but the mother authoritarian scale which had a *p*-value of only .009. These results are summarized in Table 12.
For the five scales in which a significant moderating effect of scale variance on scale mean was detected, all but the mother permissive scale (.0005) had negative slopes indicating that, as the variance increased, nonreliability went down. Thus, for every point the variance increased, reliability went up multiplicatively by the estimate.

**Test of Percent Male as a Moderator Variable**

Results of the logistic regression showed that the percentage of male participants had a significant moderating effect on reliabilities for the mother authoritative, mother permissive, and father authoritarian scales. In each case, the results indicated that, as the percentage of males increased, the nonreliability increased, and thus, reliability decreased. These results are summarized in Table 13. So, for example, on the mother authoritative scale, as the number of male participants increased by 1%, the reliability of
the scale decreased multiplicatively by .0045. For the other three scales, there was no detectible significant effect of the percentage of male participants on scale reliabilities.

Table 13

*Varying Coefficient Model with Percent Male as a Moderator Variable*

<table>
<thead>
<tr>
<th>Scale</th>
<th>Estimate</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mother Authoritarian</td>
<td>0.0027</td>
<td>0.0011</td>
<td>0.0062</td>
<td>0.0009</td>
</tr>
<tr>
<td>Mother Authoritative</td>
<td>0.0045</td>
<td>0.0029</td>
<td>0.0076</td>
<td>*</td>
</tr>
<tr>
<td>Mother Permissive</td>
<td>0.0059</td>
<td>0.0042</td>
<td>0.0076</td>
<td>*</td>
</tr>
<tr>
<td>Father Authoritarian</td>
<td>0.0007</td>
<td>0.0008</td>
<td>0.0023</td>
<td>*</td>
</tr>
<tr>
<td>Father Authoritative</td>
<td>0.002</td>
<td>0.005</td>
<td>0.0034</td>
<td>0.0084</td>
</tr>
<tr>
<td>Father Permissive</td>
<td>0.0008</td>
<td>0.0008</td>
<td>0.0023</td>
<td>0.3296</td>
</tr>
</tbody>
</table>

* Significant at .0001 level

**Categorical Moderator Variables**

**Test of Instrument Language as a Moderator Variable**

Results of the moderator analysis using the varying coefficient model showed that the language of the instrument had a significant moderating effort on reliability. For each of the six scales, studies in which the Parental Authority Questionnaire was administered in English had higher reliabilities than those that were administered in a language other than English. For example, the estimated reliability of the mother authoritarian scale was .7218 for studies in which the PAQ was administered in a language other than English, while the reliability was .8442 for studies in which the PAQ was administered in English. The difference in the reliabilities of the mother authoritarian scale was -.1224, which was
significant at the < .0001 level. Results of this analysis are summarized in Table 14. Note, also that, while the confidence intervals were narrow across all scales, the range of the confidence intervals was larger for the studies in which the PAQ was administered in a language other than English.

**Test of Developmental Level as a Moderator Variable**

Results of the moderator analysis using the varying coefficient model showed that the developmental level of the participant had a significant moderating effect on the reliabilities for all six scales with the $p$-values of the differences in alpha at < .0001. Reliabilities on all 6 scales were higher for participants 20 years and older than for those studies in which participants were under the age of 20. The confidences intervals were quite narrow across all scales, with larger ranges for the mother permissive and father permissive scales for both groups of participants. The range for the father authoritarian scale was slightly larger for studies in which participants were under the age of 20. A summary of the results of this analysis is presented in Table 15.
# Table 14

*Varying Coefficient Model with Instrument Language as a Moderator Variable*

<table>
<thead>
<tr>
<th>Scale</th>
<th>Language</th>
<th>Alpha</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mother Authoritarian</td>
<td>Not in English</td>
<td>0.7218</td>
<td>0.7093</td>
<td>0.7358</td>
<td></td>
</tr>
<tr>
<td></td>
<td>English</td>
<td>0.8442</td>
<td>0.8403</td>
<td>0.8497</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Difference</td>
<td>-0.1224</td>
<td>-0.136</td>
<td>0.1078</td>
<td>*</td>
</tr>
<tr>
<td>Mother Authoritative</td>
<td>Not in English</td>
<td>0.7528</td>
<td>0.7407</td>
<td>0.7662</td>
<td></td>
</tr>
<tr>
<td></td>
<td>English</td>
<td>0.8474</td>
<td>0.8434</td>
<td>0.8530</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Difference</td>
<td>-0.0947</td>
<td>-0.1080</td>
<td>-0.0862</td>
<td>*</td>
</tr>
<tr>
<td>Mother Permissive</td>
<td>Not in English</td>
<td>0.6353</td>
<td>0.6176</td>
<td>0.6556</td>
<td></td>
</tr>
<tr>
<td></td>
<td>English</td>
<td>0.7641</td>
<td>0.7581</td>
<td>0.7724</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Difference</td>
<td>-0.1288</td>
<td>-0.1484</td>
<td>-0.1076</td>
<td>*</td>
</tr>
<tr>
<td>Father Authoritarian</td>
<td>Not in English</td>
<td>0.7294</td>
<td>0.7147</td>
<td>0.7462</td>
<td></td>
</tr>
<tr>
<td></td>
<td>English</td>
<td>0.8808</td>
<td>0.8776</td>
<td>0.8852</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Difference</td>
<td>-0.1514</td>
<td>-0.1667</td>
<td>-0.1343</td>
<td>*</td>
</tr>
<tr>
<td>Father Authoritative</td>
<td>Not in English</td>
<td>0.7624</td>
<td>0.7504</td>
<td>0.7761</td>
<td></td>
</tr>
<tr>
<td></td>
<td>English</td>
<td>0.8715</td>
<td>0.8678</td>
<td>0.8873</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Difference</td>
<td>-0.1091</td>
<td>-0.1220</td>
<td>-0.0949</td>
<td>*</td>
</tr>
<tr>
<td>Father Permissive</td>
<td>Not in English</td>
<td>0.6664</td>
<td>0.6475</td>
<td>0.6880</td>
<td></td>
</tr>
<tr>
<td></td>
<td>English</td>
<td>0.7707</td>
<td>0.7651</td>
<td>0.7784</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Difference</td>
<td>-0.1043</td>
<td>-0.1247</td>
<td>-0.0819</td>
<td>*</td>
</tr>
</tbody>
</table>

* Significant at .0001 level
Table 15

*Varying Coefficient Model with Developmental Level as a Moderator Variable*

<table>
<thead>
<tr>
<th>Scale</th>
<th>Age</th>
<th>Reliability Estimate</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mother Authoritarian</td>
<td>Under 20</td>
<td>0.7938</td>
<td>0.7870</td>
<td>0.8025</td>
<td></td>
</tr>
<tr>
<td></td>
<td>20 Years or Older</td>
<td>0.8367</td>
<td>0.8318</td>
<td>0.8431</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Difference</td>
<td>-0.0428</td>
<td>-0.0522</td>
<td>-0.0329</td>
<td>*</td>
</tr>
<tr>
<td>Mother Authoritative</td>
<td>Under 20</td>
<td>0.8028</td>
<td>0.7970</td>
<td>0.8104</td>
<td></td>
</tr>
<tr>
<td></td>
<td>20 Years or Older</td>
<td>0.8447</td>
<td>0.8385</td>
<td>0.8523</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Difference</td>
<td>-0.0419</td>
<td>-0.0515</td>
<td>-0.0321</td>
<td>*</td>
</tr>
<tr>
<td>Mother Permissive</td>
<td>Under 20</td>
<td>0.7127</td>
<td>0.7029</td>
<td>0.7249</td>
<td></td>
</tr>
<tr>
<td></td>
<td>20 Years or Older</td>
<td>0.7559</td>
<td>0.7483</td>
<td>0.7659</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Difference</td>
<td>-0.0431</td>
<td>-0.0571</td>
<td>-0.0288</td>
<td>*</td>
</tr>
<tr>
<td>Father Authoritarian</td>
<td>Under 20</td>
<td>0.8106</td>
<td>0.8036</td>
<td>0.8192</td>
<td></td>
</tr>
<tr>
<td></td>
<td>20 Years or Older</td>
<td>0.8858</td>
<td>0.8817</td>
<td>0.8909</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Difference</td>
<td>-0.0751</td>
<td>-0.0838</td>
<td>-0.0656</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Father Authoritative</td>
<td>Under 20</td>
<td>0.8190</td>
<td>0.8130</td>
<td>0.8266</td>
<td></td>
</tr>
<tr>
<td></td>
<td>20 Years or Older</td>
<td>0.8810</td>
<td>0.8765</td>
<td>0.8866</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Difference</td>
<td>-0.0619</td>
<td>-0.0702</td>
<td>-0.0531</td>
<td>*</td>
</tr>
<tr>
<td>Father Permissive</td>
<td>Under 20</td>
<td>0.7349</td>
<td>-0.7262</td>
<td>0.7460</td>
<td></td>
</tr>
<tr>
<td></td>
<td>20 Years or Older</td>
<td>0.7669</td>
<td>0.7591</td>
<td>0.7766</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Difference</td>
<td>-0.0320</td>
<td>-0.0450</td>
<td>-0.0184</td>
<td>*</td>
</tr>
</tbody>
</table>

* Significant at .0001 level
Test of Type of Publication as a Moderator Variable

Results of the moderator variable analysis using the varying coefficient model showed that the type of publication had a significant moderating effect on reliabilities across all scales with \( p \)-values of \(< .0001 \) for the differences in reliabilities. For each of the six scales, studies undertaken as part of a dissertation or thesis showed higher scale reliabilities than those that were published in a journal article. Confidence intervals were narrow across all scales regardless of publication type, although the ranges of the reliabilities for the mother permissive and the father permissive scales were larger for both types of publications. The results of this analysis are summarized in Table 16.

Test of Year of Publication as a Moderator Variable

Results of the moderator variable analysis using the varying coefficient model showed that the year in which a study was published had a significant moderating effect on scale reliabilities across all scales with \( p \)-values of \(< .0001 \). For all six scales, reliabilities were higher for studies published before the year 2000 than for those published in 2000 or later. The ranges of the confidence intervals were relatively narrow across all the scales for all studies no matter when they were published. That said, the ranges of the confidence intervals were larger for the father authoritarian scale in studies published 2000 and later, as well as the father permissive scale for studies published before 2000. The results of this analysis are summarized in Table 17.
Table 16

*Varying Coefficient Model with Publication Type as a Moderator Variable*

<table>
<thead>
<tr>
<th>Scale</th>
<th>Publication Type</th>
<th>Reliability Estimate</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mother Authoritarian</td>
<td>Dissertation/Thesis</td>
<td>0.8368</td>
<td>0.8316</td>
<td>0.8434</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Journal Article</td>
<td>0.7984</td>
<td>0.7904</td>
<td>0.8065</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Difference</td>
<td>0.0384</td>
<td>0.0288</td>
<td>0.0476</td>
<td>*</td>
</tr>
<tr>
<td>Mother Authoritative</td>
<td>Dissertation/Thesis</td>
<td>0.8341</td>
<td>0.8288</td>
<td>0.8410</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Journal Article</td>
<td>0.8129</td>
<td>0.0612</td>
<td>0.8212</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Difference</td>
<td>0.0213</td>
<td>0.0114</td>
<td>0.0309</td>
<td>*</td>
</tr>
<tr>
<td>Mother Permissive</td>
<td>Dissertation/Thesis</td>
<td>0.7590</td>
<td>0.7507</td>
<td>0.7696</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Journal Article</td>
<td>0.7156</td>
<td>0.7067</td>
<td>0.7268</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Difference</td>
<td>0.0434</td>
<td>0.0295</td>
<td>0.0572</td>
<td>*</td>
</tr>
<tr>
<td>Father Authoritarian</td>
<td>Dissertation/Thesis</td>
<td>0.8699</td>
<td>0.8656</td>
<td>0.8751</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Journal Article</td>
<td>0.8141</td>
<td>0.8062</td>
<td>0.8237</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Difference</td>
<td>0.0558</td>
<td>0.0452</td>
<td>0.0653</td>
<td>*</td>
</tr>
<tr>
<td>Father Authoritative</td>
<td>Dissertation/Thesis</td>
<td>0.8598</td>
<td>0.8549</td>
<td>0.8658</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Journal Article</td>
<td>0.8281</td>
<td>0.8216</td>
<td>0.8362</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Difference</td>
<td>0.0316</td>
<td>0.0222</td>
<td>0.0406</td>
<td>*</td>
</tr>
<tr>
<td>Father Permissive</td>
<td>Dissertation/Thesis</td>
<td>0.7691</td>
<td>0.7626</td>
<td>0.7776</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Journal Article</td>
<td>0.7228</td>
<td>0.7119</td>
<td>0.7362</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Difference</td>
<td>0.0463</td>
<td>0.0314</td>
<td>0.0601</td>
<td>*</td>
</tr>
</tbody>
</table>

* Significant at .0001 level
Table 17

*Varying Coefficient Model with Publication Year as a Moderator Variable*

<table>
<thead>
<tr>
<th>Scale</th>
<th>Publication Year</th>
<th>Reliability Estimate</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mother Authoritarian</td>
<td>Before 2000</td>
<td>0.8392</td>
<td>0.8335</td>
<td>0.8462</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2000 and Later</td>
<td>0.8030</td>
<td>0.7974</td>
<td>0.8104</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Difference</td>
<td>0.0304</td>
<td>0.2680</td>
<td>0.0452</td>
<td>*</td>
</tr>
<tr>
<td>Mother Authoritative</td>
<td>Before 2000</td>
<td>0.8440</td>
<td>0.8382</td>
<td>0.8511</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2000 and Later</td>
<td>0.8109</td>
<td>0.8051</td>
<td>0.8185</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Difference</td>
<td>0.0331</td>
<td>0.0235</td>
<td>0.0423</td>
<td>*</td>
</tr>
<tr>
<td>Mother Permissive</td>
<td>Before 2000</td>
<td>0.7588</td>
<td>0.7507</td>
<td>0.7696</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2000 and Later</td>
<td>0.7210</td>
<td>0.7067</td>
<td>0.7629</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Difference</td>
<td>0.0379</td>
<td>0.0295</td>
<td>0.0572</td>
<td>*</td>
</tr>
<tr>
<td>Father Authoritarian</td>
<td>Before 2000</td>
<td>0.8757</td>
<td>0.8702</td>
<td>0.8819</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2000 and Later</td>
<td>0.8233</td>
<td>0.8171</td>
<td>0.8310</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Difference</td>
<td>0.0521</td>
<td>0.0428</td>
<td>0.0611</td>
<td>*</td>
</tr>
<tr>
<td>Father Authoritative</td>
<td>Before 2000</td>
<td>0.8683</td>
<td>0.8622</td>
<td>0.8754</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2000 and Later</td>
<td>0.8305</td>
<td>0.8253</td>
<td>0.8372</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Difference</td>
<td>0.0378</td>
<td>0.0287</td>
<td>0.0466</td>
<td>*</td>
</tr>
<tr>
<td>Father Permissive</td>
<td>Before 2000</td>
<td>0.7668</td>
<td>0.7583</td>
<td>0.7772</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2000 and Later</td>
<td>0.7380</td>
<td>0.7301</td>
<td>0.7485</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Difference</td>
<td>0.0287</td>
<td>0.0153</td>
<td>0.4188</td>
<td>*</td>
</tr>
</tbody>
</table>

*Significant at .0001 level
CHAPTER V

CONCLUSIONS

The results of this meta-analysis have wide-reaching implications. First, this study highlighted serious methodological issues that seemed to be inherent in studies done across social science disciplines. Second, although all six scales of the Parental Authority Questionnaire (PAQ) were found to be reliable at a level generally accepted for research purposes (DeVellis, 2016), the reliability was not to the degree necessary when making decisions that affect individuals (Suhr & Shay, 2009). Third, seven moderator variables were identified that significantly affected scale reliability; (a) scale mean, (b) scale variance, (c) percent of male participants, (d) developmental level of participants, (e) the language the PAQ was administered in, (f) the type of publication, and (g) the year of publication. Fourth, the degree of reliability varied by scale, especially with regard to the permissive scales, suggesting other as yet unidentified variables may moderate the scale reliabilities.

Finally, a comparison of the two different RG meta-analytic methods revealed that both the random-effect model and the varying coefficient model yielded similar results with regard to instrument and scale reliability. That said, the varying-coefficient model provided smaller confidence intervals and was more sensitive at detecting moderator variables.
Methodological Mistakes

In an era when lack of replicability in psychological studies has been cited as one of the top issues in social science research (Open Science Collaboration, 2015), this reliability generalization has illustrated the problematic research practices that have led to this inability to replicate results and highlight the danger in drawing conclusions and planning interventions when relying on such results. The Parental Authority Questionnaire (PAQ) has been administered by researchers in many fields of social science to study the relationships between parental authority and other psychological, sociological, and educational constructs.

For the past 25 years, the psychologists have used the results of these studies to develop treatment plans that address mental health problems such as eating disorders (Corgiat & Templer, 2003), anxiety (Mason, 2004), depression (Van Huisstede, 2013), perfectionism (Çerkez, 2012; Kawamura et al., 2002), narcissism (Lootens, 2010; Ramsey, Watson, Biderman, & Reeves, 1996), anti-social behavior (Schaffer, Clark, & Jeglic, 2009), suicide ideation (Lai & McBride-Chang, 2001), and addiction (Arms, 2009). Educators have used the results to plan curricula that address issues with academic achievement (Hickman et al., 2000), self-esteem (Cardinali & D’Allura, 2001), self-regulation (Abar, Carter, & Winsler, 2009), and classroom behavior (Gonzalez, Doan Holbein, & Quilter, 2002). Parent educators have used the results of studies to define “good parenting” and recommend parental discipline practices and guide parent-child interactions (Burke, 2005; Heer, 2008; Hill, 1995). But can researchers rely on the results yielded by the PAQ to inform such decisions?
**Issues with Reliability**

Reliability is a measure of the consistency of responses to an instrument within individuals and across individuals. It also reflects the degree of measurement error in any given set of responses. It must be calculated for each group of individuals who respond to the measure each and every time the instrument is administered (Thompson & Vacha-Haase, 2000). Instrument reliability is an essential component of validity because unreliability, or nonreliability as Bonett (2010) calls it, distorts the accuracy of the measure in any given group of individuals and leads to inaccurate conclusions. At least 213 studies that used the PAQ were excluded from this reliability generalization because of issues with reported reliabilities, or lack thereof.

**Reliability Induction**

Reliability induction occurs when researchers assume an instrument is reliable based on the reliability estimates of the measure in previous studies (Thompson & Vacha-Haase, 2000). Even when the population of participants in a study is similar to those of previous studies (e.g., college students), reliability needs to be calculated for every administration because situational and cohort effects could affect the responses of individuals and, thus, the reliability of the measure. In this meta-analysis, 92 of the studies reviewed used reliability induction. Patock-Peckham and Morgan-Lopez (2009) was one such example: “Reitman et al. (2002) found the Parental Authority Questionnaire to have acceptable factor structure, internal consistency, and convergent validity for samples with similar demographics to those in this investigation” (p. 217). Ironically, Reitman, Rhode, Hupp, and Altobello (2002) were validating the Parental Authority Questionnaire-Revised (PAQ-R), a measure that transformed the items on the PAQ from child-report to
parent self-report. In other words, the demographic make-up of the Reitman et al. (2002) study was very different from that of the respondents in Patock-Peckham and Morgan-Lopez (2009). But even if it were not that divergent, reliability induction would have been inappropriate.

In another example, Giunta (2002) acknowledged that,

While the DIT and SFI have been previously used to assess delinquency, those instruments and the PAQ fail to provide normative data about adolescent offenders in their testing manuals. Consequently, this population may need its own reliability measure, item response study, and construct validity assessment. (p. 128)

Even so, the researcher still did not do so. Even Buri used reliability induction, citing the scale reliabilities he reported in the PAQ development process.

**Other Problems with Reliability**

In addition to reliability induction, of the 121 other studies excluded from this meta-analysis due to problems with reliability, 28 reported no reliabilities at all, 11 only reported a range of reliabilities, 19 reported reliabilities of the overall instrument, and 12 reported reliabilities for the combined mother and father responses for each of the scales which violated the assumption of independent observations. While one would like to assume that these mistakes were oversights or misunderstandings by researchers, these reporting errors could raise niggling doubts about the veracity of research practices across the discipline. These practices distorted the published results and increase the likelihood of Type I error in which researchers may have identified a significant relationship between variables that did not truly exist.
Issues with Inconsistent Scale Construction and Language

Many of the studies that used the “Parental Authority Questionnaire” changed the scales so dramatically that it became a different instrument altogether. There were 160 such studies identified in this investigation.

Eighteen studies were excluded from this investigation because researchers eliminated items to make the questionnaire shorter. Heaven and Ciarrochi (2008) explained their rationale: “Because of time and space pressures, we used a shortened version of the PAQ by randomly selecting 15 of the 30 items and modifying each item slightly to suit 12-year-olds” (p. 712). Unfortunately, because, theoretically at least, each item was devised by Buri and vetted by practitioners in the field to cover a certain aspect of each parenting style, shortened versions, especially when items were randomly selected, could not adequately assess parental authority. Moreover, statistically speaking, eliminating items decreases test validity. Thus, any changes made to an instrument without examining reliability would be likely to increase imprecision in the use of the instrument to an unknown degree.

Just as Heaven and Ciarrochi (2008) did, 60 other studies changed the wording or modified the item in some way other than by mere translation. In 74 instances, the items on the PAQ were reworded to reflect the parents’ self-reflection of their own parenting style. Yet the self-report aspect of assessing parenting style might very well lead to different results than a children’s assessment of parenting style. According to Çerkez (2012), “this questionnaire is particularly designed to measure children’s perceptions of their fathers’ or mothers’ authorities” (p. 74).
Even so, it could reasonably be argued that, even if the results of the modified scales were proven to be reliable, the self-report instrument could not be considered the same instrument. In fact, Reitman et al. (2002) developed the Parental Authority Questionnaire--Revised, which was essentially the Parental Authority Questionnaire with language modified to reflect self-report parental attitudes. Eight studies excluded from this investigation reported on the parental unit rather than mothers and fathers separately. The make-up of the parental unit varied. One Saudi study (Dwairy, 2006) assessed the parental authority of the Ahel, the group of adults that interact with children in polygamous households. This type of family group can be so different from what is thought of as a family in Western societies that it would seem to be difficult to assess “parental authority” with an instrument designed with Western families in mind. Other parental units were comprised of grandparents, adoptive parents, or primary caregivers. Many studies just combined the mother and father together (e.g., Boeckel & Castellá Sarriera, 2005; Chan & Chan, 2007) and asked respondents to consider the combined style of their parents. In these instances, phrases such as “my mother” or “my father” were changed to “my parents.”

In studying the effect of assessing the parental unit as a whole, Furnham and Cheng (2000) discovered that, when assessing parenting style for mothers and fathers combined, scale reliabilities were higher (.88, .88, and .82 for authoritarian, authoritative, and permissive) when parenting style was assessed separately for mothers (.87, .83, and .73, respectively) and fathers (.89, .85, and .75, respectively). Furnham and Cheng’s (2000) analysis showed that the gender of the parent being assessed had a moderating effect on reliability with reliabilities higher when offspring were assessing their fathers’
style as opposed to their mothers’ style. Russell et al. (1998) found that mothers were more authoritative in nature as compared to fathers who were more likely to be authoritarian or permissive. McKinney and Renk (2007) found similar results and concluded that fathers and mothers may parent their sons and their daughters differently. If mothers and fathers do indeed parent differently, then using the PAQ to assess the parental unit as a whole will not capture these differences, and the results may not reflect the true magnitude of the relationship between the variables studied.

Other Methodological Issues

How reliability is reported or not reported, as highlighted by this investigation, could be a serious methodological concern that calls into questions the results of the studies across disciplines that conclude there were correlations between parental authority and other psychological, sociological, or educational constructs. But there were other serious methodological issues inherent in the research that might call results into question. For example, 45 of the studies reviewed in this investigation provided no data or results. Six others reported unclear data. Moreover, at least six studies were excluded because the same PAQ data from previous studies was presented. In this last case, it should be pointed out that such recycling of data and/or results was only discovered after the initial review process had been completed and data extraction of the studies to be included began. In each of these instances, the PAQ data was only included once.

Any number of other studies, excluded for other reasons, may have recycled data from previous studies. In the studies reviewed in this meta-analysis, there was no evidence of nefarious intent. Rather, it was more likely that researchers were uninformed about the statistical implications of using the same data to identify correlations between
constructs and that they were unaware that doing so could cause a Type I error. In other words, this practice could have led to the conclusion that correlations existed between constructs when they did not.

**Methodological and Reliability Issues Inherent in the Studies Included in the Reliability Generalization**

Unfortunately, studies that met the inclusion criteria were not immune to methodological issues. Of the 69 studies included in this reliability generalization that used the mother authoritarian scale, only 40 reported scale mean and 37 scale variance. For the other scales, this basic statistical data was sorely lacking with little more than half reporting the mean and variance of the PAQ scales that were administered. Table 2 in Chapter IV summarizes these disparities. Moreover, many of the studies included did not report basic demographic information of participants and, thus, made analyses of the effects of moderating variables more difficult to identify.

Without accurate and thorough reporting of data, it was extremely difficult to reach sound conclusions regarding the relationships between parenting style and other constructs. Nor could we determine the magnitude of such relationships or identify variables that moderate the effect of these relationships.

**Reliability of Scales and Moderator Effects**

In this investigation, all six scales were found to be reliable at a level that was generally acceptable for research purposes (DeVellis, 2016) but not necessarily to the degree necessary when making decisions that affect individuals (Suhr & Shay, 2009). Moreover, if all the reliability data for each of the six scales in all of the studies had been reported, a different outcome regarding scale reliabilities may have been identified. It was
impossible to know whether the reliability data were inadvertently excluded, or if they were excluded because at least one of the scales in the studies demonstrated unacceptable reliability.

Seven moderator variables were identified that significantly affected scale reliabilities (scale mean, scale variance, percent of male participants, developmental level of the participant, the language the scale was administered in, the type of publication, and the year of publication). What was most interesting was that each of the two reliability generalization approaches (the random-effect model and the varying coefficient model) produced differing results.

**Comparison of the Random-effect and Varying Coefficient Model**

Both reliability generalization methods indicated that the six scales of the Parental Authority Questionnaire (PAQ) were reliable, although each to varying degrees. Table 18 summarizes the differences in reliabilities between the two models. Note that, with the exception of the mother authoritative scale, the scale reliabilities were higher when the data was analyzed using the varying coefficient model as opposed to the random-effect model. But the important factor in explaining these differences lay with the confidence intervals identified by each of the models. The varying coefficient model was able to narrow the confidence intervals for each scale. Thus, it provided a more precise estimate of reliability.
<table>
<thead>
<tr>
<th>Scale</th>
<th># of Data Set</th>
<th>Min</th>
<th>Max</th>
<th>Alpha</th>
<th>Random-effect Model</th>
<th>Varying Coefficients Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Confidence Interval</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lower Bound</td>
<td>Upper Bound</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Alpha</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Confidence Interval</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lower Bound</td>
<td>Upper Bound</td>
</tr>
<tr>
<td>Mother Authoritarian</td>
<td>67</td>
<td>0.57</td>
<td>0.91</td>
<td>0.8318</td>
<td>0.8158</td>
<td>0.847</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.8225</td>
<td></td>
</tr>
<tr>
<td>Mother Authoritative</td>
<td>62</td>
<td>0.54</td>
<td>0.99</td>
<td>0.8389</td>
<td>0.8197</td>
<td>0.8567</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.8278</td>
<td></td>
</tr>
<tr>
<td>Mother Permissive</td>
<td>61</td>
<td>0.48</td>
<td>0.86</td>
<td>0.7477</td>
<td>0.7273</td>
<td>0.7670</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.7389</td>
<td></td>
</tr>
<tr>
<td>Father Authoritarian</td>
<td>58</td>
<td>0.6</td>
<td>0.94</td>
<td>0.8667</td>
<td>0.8482</td>
<td>0.8835</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.8536</td>
<td></td>
</tr>
<tr>
<td>Father Authoritative</td>
<td>57</td>
<td>0.63</td>
<td>0.95</td>
<td>0.8629</td>
<td>0.8460</td>
<td>0.8785</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.8528</td>
<td></td>
</tr>
<tr>
<td>Father Permissive</td>
<td>54</td>
<td>0.53</td>
<td>0.86</td>
<td>0.7595</td>
<td>0.7424</td>
<td>0.7758</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.7538</td>
<td></td>
</tr>
</tbody>
</table>
Recall that the confidence intervals are the ranges within which one could predict the actual reliability of the scale with 95% confidence. Therefore, more narrow intervals would indicate more accurate reliability estimates. Indeed, even though the reliability estimate for the mother authoritative scale was lower in the varying coefficient model, it still fell within the 95% confidence interval identified by the Random-effect model.

Moreover, the random-effect model reliability estimate for the mother authoritative scale of .8389 fell outside the confidence intervals for the scale determined by the varying coefficient model. What this meant was that, according to the varying coefficient model, researchers could be 95% confident that the actual reliability of the mother authoritative scale in a population of studies similar to the ones in this investigation was not .8389, because it exceeded the upper bound of the mother authoritative confidence interval which is .8333. Of course, in practical terms, the difference between reliability estimates was miniscule and was highlighted here only to compare model results.

Another more serious concern with regard to scale reliability was the reliability estimates of the permissive scales. In both models, the results of this investigation indicated that there was 10% greater error associated with use of those scales.

**Moderator Effects**

Although this investigation was able to identify seven significant moderator effects, the random-effect model was not able to detect them all. Just as the reliability estimates varied by model used to analyze reliability, so too did moderator effects.

**Moderating Effect of Scale Mean on Reliability**

The random-effect model was not able to identify any significant effect of scale mean on scale reliability, while the varying coefficient model identified a significant
effect of scale mean on the reliability estimates for all but the father permissive scale. Unfortunately, interpretation of this effect was confounding because on two of the scales (mother authoritarian and father authoritarian) an increase in scale mean resulted in a decrease in the reliability estimates of the two scales. For the mother authoritative, mother permissive, and the father authoritative scales, the higher the scale meant the higher the reliability estimate. One could only conclude that other unidentified moderating factors contributed noise to the data. Even so, it is important to remember that many of the studies did not report scale mean, so this analysis was conducted on a small subset of the total studies included in this investigation and the results may have been more consistent had data regarding scale mean been available for all of the studies included in this reliability generalization.

**Moderating Effect of Scale Variance on Reliability**

The random-effect model was also not able to identify any significant moderator effect of scale variance on scale reliability. Yet again, the varying coefficient model analysis was more sensitive and identified a significant moderating effect of variance on reliability for all but the mother authoritarian scale. For the mother authoritative, mother permissive, father authoritarian, and father permissive scales, an increase in scale variance resulted in an increase in reliability. The moderating effect of variance on the reliability of the mother permissive scale demonstrated a significant effect in the opposite direction. For this scale, as variance increases, reliability decreases. Again, some other unidentified factor must be contributing to the noise in this data, making conclusions regarding such effects more difficult to interpret.
Moderating Effects of Percent Male on Scale Reliabilities

Just as for the moderating effects of the mean and variance, the random-effect model could detect no significant effect of the percentage of male participants on reliability. And again, the varying coefficient model was able to identify a significant moderator effect of gender of participant on reliability for three of the six scales (mother authoritative, mother permissive, and father authoritarian). Results indicated that, as the percentage of male participants increased, the reliabilities of the scales decreased. It is important to note that the vast majority of the participants across studies were female. This moderating effect might be due to the differences in the way males and females interpret parental authority or to differences in how mothers and fathers might parent sons differently than daughters. McKendrick (2011) concluded that,

By gender, results of this study revealed in the sample distribution that male respondents were more likely to report perceptions of their mother (figures’) parenting style as Authoritative with higher and significantly different mean scores than females on the Parental Authority Questionnaire (PAQ). Males were also slightly more likely to perceive of their mother (figure’s) parenting style as more permissive than females; again males had higher mean scores but not significantly different from females. Female respondents were more likely however, to perceive of their mother (figures’) as Authoritarian as hypothesized. Females had a higher mean score than males on the PAQ for this parenting style. (p.43)

Russell et al. (1998) found that parents were more likely to use an authoritarian parenting style when dealing with boys and an authoritative style when dealing with girls. McKinney and Renk (2007) also concluded that “fathers and mothers may parent their sons and daughter differently” (p. 821).
Moderating Effects of the Language the Instrument

Both the random-effect model and the varying coefficient model showed significant moderating effects of the language the instrument across all six scales of the Parental Authority Questionnaire (PAQ). Across the board, reliabilities were higher when the instrument was administered in English. That said, one could not necessarily conclude that the language itself was what was contributing to the moderating effect, because this variable also represented a number of different constructs that were inherently intertwined with language.

One of these constructs was the culture of the respondent and how parental authority was viewed across the globe. Since many of the studies in which the PAQ was not administered in English were conducted in non-Western countries, cultural concepts of family structure and parental roles came into play. Chao (2001) contended that, “because the typical ways in which family members relate to each other are primarily a reflection of culture, parenting style then can also be seen as a reflection of culture” (p. 1833). Chao’s (1994, 1996, 2001) research has demonstrated that Chinese-American and European-American mothers parent and interact with the children differently. McBride-Chang and Chang (1998) expanded on Chao’s (1994) research and further explained, “A question regarding a parent’s requirement of strict obedience, which might seem controlling and dictatorial to a Westerner” (p. 423). They went on to explain,

In Chinese culture, control, care, and concern are virtually synonymous. The notion of guan [training of children in Chinese culture] encompasses a controlling parenting style as well as a high degree of mother-child interaction and physical proximity. These concepts are largely absent from Western culture. (McBride-Chang & Chang, 1998; p. 423)
In their own research, McBride-Chang and Chang (1998) concluded that, “categorizing parenting styles as authoritative, authoritarian, or permissive may not be a culturally relevant dimension of socialization in Hong Kong adolescents” (p. 430).

Newman et al. (2015) studied the parenting styles of Chinese, Turkish, and American mothers, and in part, also concluded that Baumrind’s (1971a) parenting styles may not reflect a global perspective: “the categories of Baumrind's (1971a) parenting style model do not appear to reflect the variety of parenting practices that are observed in diverse and diversifying modern cultures” (Newman et al., 2015, p. 280). The researchers also highlighted another issue that was inextricably intertwined with the language of the instrument—that of translation.

A further consideration is that the meanings of some parenting concepts are culturally relative and very dependent on the way terms and behaviors are translated and operationalized. Concepts such as individualism, parental control, or parental warmth have most often been defined and described by Westerners, often Americans, and the instruments devised to measure them reflect Western concepts. (Newman et al., 2015; p. 281)

Peña (2007) agreed, adding that, even when the instruments were faithfully translated to be a duplicate of the one in the original language, direct translation may itself be an issue.

Translation from one language to another can result in incongruity in meaning, threatening content validity of a study's methods. Functional equivalence addresses some of these threats by ensuring that the instrument and elicitation method allow examination of the same construct. This aspect of translation is often overlooked in favor of achieving uniformity in instrumentation and procedures. (p. 1257)

Peña (2007) also spoke to the issue of salience: “Disparities in salience may be due to the distinct cultural and historical ways in which concepts are interpreted by respondents.
Cultural equivalence with respect to respondents' interpretations and responses to given items needs to be explored when one is developing methods and procedures” (p. 1257).

Peña (2007) recommended that care be taken when using and instrument that needs to be translated.

For translation of instructions to participants and of instrumentation, first consider whether bias will be introduced to the study. Techniques such as translation and back-translation result in well-translated instructions and instruments, but they may not provide equally relevant methods for the populations under study. Decentering—that is, adapting the question or item so that it is culturally familiar—can also be used in conjunction with translation. (p. 1261)

There was no clear explanation for the differences in reliabilities of the scales for those studies that used the English version of the PAQ versus a translated version. In this investigation, it was clear that language of the instrument had a moderating effect on scale reliabilities.

**Moderating Effect of Developmental Level**

Both the random-effect and varying coefficient analyses showed that developmental level of the participant had a significant moderating effect on scale reliabilities. Across all the scales, the reliabilities were higher when participants were age 20 years or older. There could be several possible explanations for this effect: (a) older participants were less likely to live at home with their parents and, thus, evaluate parental authority differently than those still in the throes of adolescent battles for control; (b) older participants’ perception of how their parents’ authority styles may be clouded by time and experience; or (c) older participants might be parents themselves and appreciate how difficult the task of parenting really is. De Los Reyes and Kazdin (2005) found that, with regard to perceptions of parental authority, current perspectives affect memory
recall. Further, Hardt and Rutter (2004) noted that those recall experiences that were heavily influenced by judgment and interpretation may be different that those that actually occurred.

Another explanation might be the item language itself. Buri (1991) designed the questionnaire for administration to college students and he used that population to establish his original instrument reliability estimates. Twenty of the 30 items on each of the parent questionnaires contained the phrase: “As I was growing up . . .” (Buri, 1991; pp. 113-114). This language might be less clear to participants under the age of 20 than those over the age of 20. They might have asked themselves what time period “when they were growing up” they should choose.

Whatever the explanation, it was clear that the age of the respondents significantly affects scale reliabilities. This phenomenon needs to be investigated more thoroughly than it has been thus far.

Moderating Effect of Publication Type

Both the random-effect model and the varying coefficient model identified a significant moderator effect of type of publication on reliability. Surprisingly, research conducted for dissertations and theses exhibited higher scale reliabilities than those studies published in journal articles. There could be several possible explanations for this moderator effect: (a) research conducted at the graduate level had more faculty oversight than other research; (b) graduate students, by the very nature of graduate programs, may have been required to learn and implement more sophisticated statistical techniques; (c) in order for research projects to be approved, graduate research should have “contributed” to the body of knowledge in the discipline; or (d) all the many, many
dissertations and theses with really egregious methodological errors were weeded out during the study review process. Alternatively, not all “peer-reviewed” journals would have the same standards when it came to publishing research, nor were the stringent ethical standards Western researchers met, in place all over the world.

**Moderating Effect of Year of Publication**

Both the random-effect model and the varying coefficient model showed a significant moderator effect of year of publication on scale reliabilities. Across the board, studies published prior to 2000 yielded higher scale reliabilities than those published after. Again, there could be many possible explanations for this effect. One explanation might be the evolving English language. A review of all of the studies that used the PAQ found that researchers modified the language of the questionnaire in 160 cases. Perhaps as language has become less formal over the last three decades, item meaning was not as clear as it once was. Item language like, “As I was growing up my mother did not direct the behaviors, activities and desires of the children in the family” and “As I was growing up, I knew what my mother expected of me in the family, and she insisted that I conform to those expectations simply out of respect for her authority,” (Buri, 1991; p. 114) might not be as clear today as it was in the last century when conforming or not conforming was a cultural focus.

Another issue that might affect reliability could be the obfuscating effects of the changing modern family. Today it is not uncommon for children to live in a blended family in which they interact with both parents and step-parents (Bianchi, Raley & Casper, 2012; Wall & Gouevia, 2014). Indeed, they might alternate between family households. In such circumstances, completing a questionnaire that asks students to
evaluate their father on these items might prompt the question: “Which father should I choose?” Then there are those families headed by grandparents, foster parents, adoptive parents, or even those with two mothers or two fathers.;

Another issue to consider was that, as more researchers used the PAQ to assess parental authority for all manner of constructs, more noise was contributed to the data. This noise, due to unidentified moderating variables, could affect reliability. Whatever the reason, it was clear that studies conducted prior to the turn of the century exhibited higher scale reliabilities than those published in the new millennium.

**Limitations**

While the results of this reliability generalization indicated that the scales of the PAQ were reliable, it was important to view these results in light of the limitations of this study. One of the major limitations was the lack of data. Of the 388 studies identified that used the original PAQ with offspring, 319 studies had to be excluded from this analysis due to the methodological mistakes and/or data omissions discussed earlier in this dissertation. It was impossible to know what impact this missing data may have had on the scale reliabilities reported here.

Moreover, of the 69 studies included in this reliability generalization, none reported all of the standard demographic information and/or statistical measures such as scale mean and scale variance. It was difficult to speculate how this missing information might have affected the moderator analyses presented here.

Another limitation was that only studies published in English or Spanish were included in this dissertation. I identified 43 studies that were not published in English or Spanish that might have been included in this meta-analysis could I have understood
what they said and had been able to determine if they indeed used the PAQ and included all the necessary data. It was also impossible to know how many other studies were conducted that was not listed on English and Spanish language databases.

Finally, the changing landscape of the family in the 21st Century called into question the reliability of the scales of the PAQ and the instrument’s ability to consistently identify parental authority types in terms of current definitions. Family structure has become much more diverse today than it was three decades ago and lumping female caregivers into the “mother” category and male caregivers into the “father” category might be inadequate to reliably identify “parental” authority and could obfuscate moderator effects.

**Implications of Results of this Meta-Analysis**

The implications of this research were clear. While the six scales of the Parental Authority Questionnaire seemed to be reliable at a level acceptable for research purposes (DeVellis, 2016), scale reliabilities were not consistent across scales, groups of participants, or time. Furthermore, the scale reliabilities where not to the degree necessary to make decisions that affect individuals (Suhr & Shay, 2009). Moreover, methodological mistakes and data omissions confounded the issue with regard to scale reliabilities and modifier effects. Before researchers plan interventions based on the results of studies that seem to correlate parental authority with various other psychological, sociological, or educational constructs, care must be taken to ensure that sound methodological practices are in place throughout the research process. These practices should include calculating reliability each time an instrument such as the Parental Authority Questionnaire is administered.
This research also yielded valuable information for meta-analytic researchers by demonstrating, with actual study data, how methods of analysis in reliability generalization could differ in not only identifying reliability estimates, but variables that moderate reliability as well. Prior to this meta-analysis, the efficacy of using differing models of analysis was investigated with only simulation data. This reliability generalization illustrated these differences using “real world” data.

**Areas of Future Research**

Although each of the scales of the Parental Authority Questionnaire have been found to be reliable to varying degrees, the results of this investigation cannot speak to instrument validity. Reliability is an important component of validity, but not the only one. The significant effects of the moderating variables call into question whether the Parental Authority Questionnaire actually yields valid results. A validity generalization is needed to determine scale and instrument validity.
REFERENCES


*Psychological Reports, 74*(3 Pt 1), 712-714. Retrieved from http://doi.org/10.2466/pr0.1994.74.3.712


APPENDIX A

INSTITUTIONAL REVIEW BOARD APPROVAL
RE: Clarification re IRB
Lahman, Maria

Sent: Sunday, September 09, 2012 10:07 AM
To: Dean, Lynn
Cc: Pulos, Steven

This is correct. Thanks for checking.

Maria Lahman
Co-Chair of IRB

From: Dean, Lynn
Sent: Friday, September 07, 2012 10:43 AM
To: Lahman, Maria
Cc: Pulos, Steven
Subject: Clarification re IRB

Maria,
Just to clarify our conversation today, am I right in understanding that I don't need an IRB to conduct a meta-analysis in which I will use published data, and, if necessary, I contact researchers regarding statistical data from said studies. -Lynn

Lynn Dean
Educational Psychology Doctoral Student
School of Psychological Sciences
McKee 14, Campus Box 94
University of Northern Colorado
Greeley, CO 80639
APPENDIX B

META-ANALYSIS REFERENCES
META-ANALYSIS REFERENCES


