Summer 2015

A Psychometric Examination of Prosocial Behavior Across Cultural Contexts

Thomas Knight

Follow this and additional works at: https://dsc.duq.edu/etd

Recommended Citation

This immediate access is brought to you for free and open access by Duquesne Scholarship Collection. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of Duquesne Scholarship Collection. For more information, please contact phillips@duq.edu.
A PSYCHOMETRIC EXAMINATION OF PROSOCIAL BEHAVIOR ACROSS CULTURAL CONTEXTS

A Dissertation

Submitted to the School of Education

Department of Counseling, Psychology, and Special Education

Duquesne University

In partial fulfillment of the requirements for

the degree of Doctor of Philosophy

By

Thomas R. Knight

August 2015
DUQUESNE UNIVERSITY

SCHOOL OF EDUCATION
Department of Counseling, Psychology, and Special Education

Dissertation

Submitted in partial fulfillment of the requirements
for the degree
Doctor of Philosophy (Ph.D.)

School Psychology Doctoral Program

Presented by:

Thomas R. Knight
B.A. Criminology, Indiana University of Pennsylvania, 2003
M.S. School Psychology, California University of Pennsylvania, 2005

July 7, 2015

A PSYCHOMETRIC EXAMINATION OF PROSOCIAL BEHAVIOR
ACROSS CULTURAL CONTEXTS

Approved by:

______________________________, Chair
Tammy L. Hughes, Ph.D.
Fr. Martin A. Hehir Endowed Chair for Scholarly Excellence, Professor and Chair
Department of Counseling, Psychology, and Special Education
Duquesne University

______________________________, Member
Kara E. McGoey, Ph.D.
Associate Professor
Department of Counseling, Psychology, and Special Education
Duquesne University

______________________________, Member
Gibbs Y. Kanyongo, Ph.D.
Associate Professor
Department of Educational Foundations and Leadership
Duquesne University
ABSTRACT

A PSYCHOMETRIC EXAMINATION OF PROSOCIAL BEHAVIOR ACROSS CULTURAL CONTEXTS

By

Thomas R. Knight

August 2015

Dissertation supervised by Tammy Hughes, Ph.D.

Acts of aggression and violence within the school setting have compelled researchers and professionals to develop and implement interventions designed to cultivate student safety. Recently, the focus of these interventions has been on prosocial behaviors, which are broadly defined as acts intended to benefit others above oneself and can include actions such as helping and cooperating (Batson & Powell, 2003). From a theoretical perspective, previous researchers (Piaget, 1932; Kohlberg, 1984) have generally postulated that moral reasoning and its corollaries (e.g., prosocial behavior) generally develop according to a prescribed trajectory. Yet, additional research in this area has supported the notion that cultural and contextual influences play an important role in one’s proclivity to act prosocially (e.g., Ellis & Boyce, 2008; Snarey, 1985). It is therefore imperative that culture and context be considered when designing interventions and assessments related to prosocial behaviors.
The *Be a Safety Kid* curriculum provides students in kindergarten through 8th grade with direct instruction intended to teach prosocial behaviors through the concept of *responsible reporting*, which entails communicating information regarding potentially dangerous situations. As part of the curriculum, the S.T.A.R. Instrument is utilized as a pre- and post-test assessment of the extent to which the curriculum has influenced the students’ knowledge of prosocial behavior, anticipated performance of prosocial behavior, and feeling of school connectedness/safety. The current study sought to ascertain the extent to which the S.T.A.R. instrument measures the constructs of knowledge, performance, and school connectedness/safety similarly across cultural contexts by examining and comparing the underlying factor structure of the S.T.A.R. Instrument post-test results from students of different states, races, community types, and socioeconomic statuses. Results of confirmatory factor analyses (CFA) for 1st and 2nd grade suggest that the S.T.A.R. instrument does not consistently align with the constructs designed through its creation. The conclusions of this study will add to the existing literature base on prosocial behaviors and further highlight the importance of considering cultural and contextual factors when developing prosocial interventions and accompanying assessments.
DEDICATION

This dissertation is dedicated to my father, Dr. Thomas A. Knight, who, despite being taken from us too soon, managed to instill in me the importance of education, determination, and self-betterment. You have always led by example and I am truly grateful for the life lessons you have taught me. I would also like to dedicate this dissertation to my mother, Nancy, who has consistently shown support, love, and encouragement during this journey. Lastly, this dissertation is dedicated to my wife, Lisa, and my daughter, Ella, who have been by my side every step of the way. I am beyond grateful for the love and support you two have shown me through this process. Your understanding and flexibility have been invaluable. It’s been a long road, but I am finally done! I love you all!
ACKNOWLEDGEMENT

I would like to acknowledge and thank the Chair of my dissertation Committee, Tammy Hughes, Ph.D., who has consistently shown support and provided me with gentle nudges when needed. Your guidance and wisdom have been invaluable during this process. I would also like to acknowledge the members of my Committee, Gibbs Kanyongo, Ph.D. and Kara McGoey, Ph.D. for their unwavering support and insightful feedback.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>iv</td>
</tr>
<tr>
<td>Dedication</td>
<td>vi</td>
</tr>
<tr>
<td>Acknowledgement</td>
<td>vii</td>
</tr>
<tr>
<td>List of Tables</td>
<td>xi</td>
</tr>
<tr>
<td><strong>CHAPTER I: INTRODUCTION</strong></td>
<td></td>
</tr>
<tr>
<td>Definitions</td>
<td>1</td>
</tr>
<tr>
<td>Development of Prosocial Behaviors</td>
<td>2</td>
</tr>
<tr>
<td>Cultural and Contextual Influences</td>
<td>4</td>
</tr>
<tr>
<td>Assessment and Intervention</td>
<td>7</td>
</tr>
<tr>
<td>Be a Safety Kid Curriculum</td>
<td>8</td>
</tr>
<tr>
<td>The S.T.A.R. Instrument</td>
<td>9</td>
</tr>
<tr>
<td>Significance of the Problem</td>
<td>9</td>
</tr>
<tr>
<td>Theoretical Basis</td>
<td>10</td>
</tr>
<tr>
<td>Existing Literature</td>
<td>12</td>
</tr>
<tr>
<td>Problem Statement</td>
<td>14</td>
</tr>
<tr>
<td>Research Questions</td>
<td>14</td>
</tr>
<tr>
<td>Summary</td>
<td>15</td>
</tr>
<tr>
<td><strong>CHAPTER II: LITERATURE REVIEW</strong></td>
<td></td>
</tr>
<tr>
<td>Evolution of the Construct of Prosocial Behavior</td>
<td>18</td>
</tr>
<tr>
<td>Contemporary Characterizations of Prosocial Behavior</td>
<td>22</td>
</tr>
<tr>
<td>Development of Prosocial Behaviors in Children</td>
<td>29</td>
</tr>
<tr>
<td>Culture, Moral Reasoning, and Prosocial Behavior</td>
<td>33</td>
</tr>
<tr>
<td>Family Values and Goals</td>
<td>36</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 1. Demographic Data for Schools.................................................................65
Table 2. Descriptive Statistics for Eliminated Data by School..........................85
Table 3. Descriptive Statistics for Gender and Grade by School ....................86
Table 4. Group Composition...................................................................................87
Table 5. CFA Assumption Statistics......................................................................92
Table 6. Comparison of CFA Model Fit Indices for States.................................94
Table 7. Squared Multiple Correlations for States..............................................95
Table 8. Comparison of CFA Model Fit Indices for Race...............................96
Table 9. Squared Multiple Correlations for Race.............................................96
Table 10. Comparison of CFA Model Fit Indices for SES .............................97
Table 11. Squared Multiple Correlations for SES...........................................97
Table 12. Comparison of CFA Model Fit Indices for Community Type ............99
Table 13. Squared Multiple Correlations for Community Type.......................99
CHAPTER I
INTRODUCTION

Despite measures increasingly being taken by schools to reduce incidents of violence, there have been approximately 30 instances over the past three decades in which students have perpetrated the act of shooting classmates and teachers in what is typically referred to as a ‘rampage’ school shooting (Kalish & Kimmel, 2010). Although schools are typically one of the safest places for children to be, tragedies that occurred in communities such as Columbine, Colorado and Padukah, Kentucky underscore the fact that these institutions are not impervious to significant acts of violence. In fact, from July 1, 2010 to June 30, 2011, there were 31 school-associated violent deaths, with 25 being homicides and 6 being suicides (Robers, Kemp, & Truman, 2013). Furthermore, these authors indicate that, during the 2009-2010 school year, 85 percent of public schools had reported that one or more crime incidents had occurred, while 75 percent of public schools reported one or more violent incidents of crime (Robers et al., 2013). These statistics, as well as the horrific images depicted by the media in the wake of these catastrophic events, undoubtedly speak to the necessity of finding ways to prevent such senseless acts of violence. To that end, there has been an increased societal focus on prosocial behaviors and curricula designed to teach and reinforce prosocial acting.

Definitions

Although definitions of the term, prosocial behavior, have varied greatly within the literature from a historical standpoint, the contemporary understanding of these behaviors is that prosocial behavior is an overarching construct that is multifaceted in nature and a concept used to describe behaviors intended to benefit others (Eisenberg & Fabes, 1998). Prosocial behaviors were initially viewed as an all-inclusive concept, without recognizing that there were various
subsets of behaviors that comprise the broad category of prosocial behaviors (Greener, 2000). However, there is now a greater understanding that prosocial behavior encompasses an array of behaviors such as helping, sharing, rescuing, comforting, protecting someone from harm, and feeling empathy and sympathy for others (Radke-Yarrow & Zahn-Waxler, 1986; Zeldin, Savin-Williams, & Small, 1984). There is also an indication that prosocial behavior is suggestive of an underlying altruistic personality trait (Eisenberg, Pasternack, Cameron, & Tyron, 1984; McKinley & Carlo, 2007) and that these behaviors are manifested as a result of an intrinsically motivated concern for others (Hay, 1994; Hay & Pawlby, 2003). Individuals who use more prosocial methods for solving conflicts with others generally offer more effective and relationship-enhancing resolutions rooted in persuasion rather than aggression (Mayeux & Cillessen, 2003). The fact that these behaviors may mediate potentially aggressive situations highlights the importance of studying the development of these behaviors.

**Development of Prosocial Behaviors**

The study of prosocial behavior and its association to child development is reported to have originated in the early 1900’s with William McDougal (1908) and Louis Murphy (1937). While McDougal is credited with composing the first text in social psychology, Murphy began empirically exploring prosocial behaviors by studying sympathy in preschool children. After Murphy’s research, attention to these behaviors diminished until 1964 when the murder of Kitty Genovese generated substantial interest in the bystander effect and why people help, or fail to help, others in particular situations. Although these behaviors were not initially connected to a specific underlying theory, Piaget (1965) and Kohlberg (1984) later postulated that these behaviors are largely rooted in processes related to moral reasoning.
Piaget (1932) proposed a cognitive-developmental theory of moral reasoning in which he put forth the idea that moral judgments progress in stages that are related to changes that occur with regard to the child’s cognitive development. His two stage theory was comprised of *heteronomous morality* (roughly from age 4 to 7) and *autonomous morality* (beginning around age 10). Upon reaching the second stage, Piaget argued that children have moved beyond viewing justice and rules as unchangeable and removed from the individual, and that they are now capable of focusing more on intentions of actions than consequences. Piaget emphasized the role that cognitions play in the development of moral reasoning and suggested that cognitive processes assist with emotional expression and experience. He acknowledged that cognitions are an integral part of development and that emotions are a necessary impetus for actions. Here, the idea was that prosocial behaviors are the result of complex interactions between thoughts and feelings.

Although Piaget provided the theoretical foundation for the study of prosocial behaviors, Kohlberg (1984) built upon the work of Piaget and provided his own archetype for understanding the development of moral reasoning. Kohlberg’s interpretation continued to recognize the relationship between prosocial behaviors and moral reasoning. However, Kohlberg expanded upon Piaget’s work by subdividing Piaget’s two stage theory into a model that was comprised of three levels and six stages. Kohlberg’s theory of moral reasoning was not only concerned with the manner in which children act in certain situations, but also the reasons and justifications children provided for their actions.

The work of Piaget and Kohlberg demonstrated that cognitive development serves as an important factor when considering moral reasoning and prosocial behaviors. However, these theoretical models, especially the one proposed by Kohlberg, largely subscribe to a universalistic
approach in which it is assumed that moral reasoning develops in the same way for all individuals. Despite these claims, there is evidence to suggest that cultural and contextual factors can serve as moderating variables that influence not only an individual’s moral development, but also the extent to which he or she engages in prosocial behaviors.

**Cultural and Contextual Influences**

From a cross-cultural perspective, differences in moral reasoning between cultures have been well documented. Through extensive research of moral reasoning in 27 different countries, Snarey (1985) found that, although Stages 1 through 4 of Kohlberg’s theory were likely universal, folk and village groups failed to exhibit any higher level moral reasoning related to Stages 5 or 6. Moreover, Snarey presented evidence suggesting that Kohlberg’s theory also failed to capture higher moral ideals embraced by some cultures. Additional cross-cultural research indicated substantial differences in moral reasoning when comparisons were made between Western and non-Western cultures. Results of this research largely suggested that Western cultures tended to give more priority to justice expectations than interpersonal expectations and that Westerners displayed moral reasoning that was more rooted in self-interest than altruism (Keller, Edelstein, Krettenauer, Fu-xi, & Ge, 2005; Miller & Bersoff, 1992). Research has also shown that cultural and contextual factors can impact prosocial development and influence an individual’s proclivity to engage in prosocial acting.

For children, goals, values, and ideals are often communicated by parents and ultimately shape the manner in which children develop (Eisenberg, Fabes, & Spinrad, 2006). Here, the question centers on the extent to which the family’s belief systems can influence a child’s prosocial behaviors. Research has demonstrated that children whose parents hold prosocial values tend to engage in higher levels of prosocial acting (Calderón-Tena, Knight, & Carlo,
Family values can also include concepts such as religious beliefs and political preferences, which have also been shown to impact prosocial acting.

Research in this area has suggested that liberals tend to endorse a greater sympathy for those less fortunate than them when compared to conservatives (Zettler & Hilbig, 2010) and that they tend to self-report higher levels of altruistic behaviors (Zettler, Hilbig, & Haubrich, 2011). Additionally, Democrats endorsed higher levels of prosocial acting and favored intrinsic values in comparison to Republicans (Sheldon & Nichols, 1994). Regarding religion, the general consensus is that there is a positive correlation between religion and reported prosocial behaviors (Furrow, Kind, & White, 2004; Saroglou, Trompette, Verschueren, & Dernell, 2005), although some (Galén, 2012) have argued that the relationship between religion and prosociality is a fallacy and a phenomenon produced by secular psychological effects. Based upon this discussion, it may be hypothesized that children growing up in religious, Democratic families may display more prosocial behaviors than a child raised in a non-religious, Republican household. Regardless of family beliefs, there is evidence that contextual factors, such as community type, can also influence the propensity to act prosocially.

There is overwhelming empirical evidence that acts of helping and prosociality are exhibited at a higher rate in rural areas when compared to urban areas (Amato, 1983; Kamal, Mehta, & Jain, 1987; McMahon et al., 2013; Steblay, 1987; Wilson & Kennedy, 2006). It has been hypothesized that issues related to violence, overcrowding, and quality of housing, which are often found in urban settings, increase an individual’s stress and produce more self-serving behaviors and less prosocial behaviors (Denhardt & Glasser, 1999). For children, it is important to consider that geographic location may play an integral role in the extent to which they engage
in prosocial acting. Relatedly, race and socioeconomic status have also been found to exert some force on prosocial behavior.

A review of the literature reveals limited empirical evidence related to the impact of race on prosocial behavior. As stated by Wentzel, Filisetti, and Looney (2007), the effects of race on prosocial behavior are poorly understood due to the fact that theoretical models that discuss them are infrequent. For example, Caucasian students in a middle school where the majority (87%) were Caucasian were reported by peers to engage prosocial behavior within the classroom more often than were African American students in a middle school where the majority (92%) were African American (Wentzel, 2002). However, elementary-age Caucasian and African American students’ reports of the frequency with which peers engaged in prosocial behavior produced no significant group differences when the racial majority was considered (Kistner, Metzler, Gatlin, & Risi, 1993). Wentzel, Filisetti, and Looney (2007) found that Caucasian students were reported to engage in prosocial behavior significantly more than the African American students. However, Caucasian females reported more self-focused reasons for behavior than what their African American counterparts did. Additional analyses based upon racial classroom composition suggested significant differences were found in the Caucasian-majority classrooms, where Caucasian students were reported to engage in prosocial behaviors more often than African Americans. Yet, reported prosocial behaviors did not significantly differ as a function of race in the classrooms where African Americans were the majority. The authors posit that these findings may be related to the fact that African American students who attend schools where they are the racial minority may not experience the same differentiated crowd structures that Caucasian students do and, in turn, are not affected by peer norms in the same way or to the same degree.
Like urban communities, lower socioeconomic status (SES) is generally associated with limited resources, greater exposure to threat, and a decreased sense of personal control over situations (Piff, Kraus, Cote, Cheng, & Keltner, 2010). As such, one may expect that individuals from lower SES would be less inclined to act prosocially and, instead, put self-interests over the welfare of others. Yet, research conducted by Piff et al. (2010) indicates that this is indeed not the case. Across four separate studies, the researchers found that those participants from a lower SES proved to be more generous, charitable, trusting, and helpful when compared to their higher SES counterparts. The researchers suggested that those from a lower SES are more inclined to act prosocially due to a greater obligation to egalitarian values and feelings of compassion. Thus, while one may expect these individuals to act in a manner that reflects self-interest, these results show that the adversity associated with low SES actually serves as an impetus for acting prosocially.

Although psychological theories provide a general overview of how development occurs, consideration of cultural and contextual factors produces the awareness that there is a complex interaction between biology and context that may contribute to vast differences in prosocial acting among children. In addition to understanding the manner in which prosocial behaviors develop, it is also important to be able to assess these behaviors effectively so that appropriate interventions can be identified and implemented.

Assessment and Intervention

Despite the significant role that prosocial behaviors can play in mediating potentially aggressive acts, there are a limited number of assessments designed to evaluate these behaviors. In fact, a search of the Mental Measurements Yearbook for the term prosocial only reveals three assessments that specifically measure the construct of prosocial behavior: The Infant-Toddler
and Brief Infant Toddler Social Emotional Assessment (ITSEA/BITSEA; Briggs Gowan & Carter, 2006); the Louisville Behavior Checklist (LBC; Miller, 1984); and the Social Competence and Behavior Evaluation (SCBE; LaFreniere & Dumas, 1995). While other measures of prosociality have been developed through the course of single research studies, these instruments lack strong psychometric properties and, in some cases, have no data regarding the reliability and validity of the instrument.

The paucity of available assessments intended to measure prosocial behaviors can make intervention planning difficult. If strengths and needs cannot be effectively identified with regard to prosocial development, then developing appropriate interventions can become problematic. Given the positive implications of prosocial behaviors, schools have begun to implement intervention programs focusing on prosocial behaviors, with most programs having an accompanying assessment to determine program effectiveness and student progress. The specific program of interest here is the Be a Safety Kid curriculum and associated S.T.A.R instrument.

**Be a Safety Kid Curriculum**

The *Be a Safety Kid* curriculum is a school-based violence intervention program designed around the concept of “responsible reporting”, which entails communicating information that could prevent an injury or crime (Safety Kids, 2005). The program is designed for grades kindergarten through eighth grade and is intended to change attitudes, teach responsibility, and promote accountability. Through the “Ima S.T.A.R.” unit of this curriculum, the goal is to break the mindset of “snitching” and encourage students to make good decisions when they see potential dangers within their environment. With this program, acts of responsible reporting are viewed as prosocial behaviors. In addition to the curriculum units, the *Be a Safety Kid*
curriculum also includes the S.T.A.R. assessment, which is designed to measure the effectiveness of the program.

**The S.T.A.R. Instrument**

The S.T.A.R. instrument is designed to be administered as a pre- and post-test for purposes of assessing three constructs: knowledge of prosocial behaviors and responsible reporting; performance of prosocial behaviors and responsible reporting; and the feeling of school connectedness/safety. Comparing scores from before and after the implementation of the curriculum provides data regarding the extent to which the curriculum was effective in producing change in the three constructs. The S.T.A.R. instrument serves as the sole indicator of progress with the *Be a Safety Kid* curriculum, with post-test scores potentially indicating the need for further prosocial interventions with some students. Thus, in addition to measuring progress, data from the S.T.A.R instrument is also used for decision making and planning purposes for students who are continuing to show deficits in prosocial functioning.

**Significance of the Problem**

Knowing the role that cultural and contextual factors can play in the development and demonstration of prosocial behaviors, it is conceivable that elements of context and culture can impact how specific prosocial behaviors are defined. With regard to the S.T.A.R. instrument, this may further suggest that cultural and contextual factors may affect how the S.T.A.R instrument is measuring behaviors associated with the three constructs (knowledge of prosocial behaviors and responsible reporting; performance of prosocial behaviors and responsible reporting; and the feeling of school connectedness) being assessed. Here, the concern is the extent to which the S.T.A.R. instrument ignores any culturally specific definitions of these constructs when comparing five states (Arizona, Florida, Georgia, Pennsylvania, and Wisconsin).
and also comparing differences between high/low SES, race, and urban/rural communities. For example, is the S.T.A.R. instrument measuring knowledge of prosocial behaviors the same way in Arizona as it does in Wisconsin?

Due to the fact that the S.T.A.R instrument is the sole indicator of progress in relation to the Be a Safety Kid curriculum, it is of great importance to determine the extent to which this instrument is capable of measuring constructs similarly across states and cultural contexts. Although previous research has examined the validity and reliability of the S.T.A.R. instrument at the local level of implementation, between-state comparisons of data and scores derived from this assessment cannot be confidently made until it has been determined that constructs are being measured in a similar fashion across these states. Further statistical analysis of the S.T.A.R. instrument will provide additional information regarding its utility and the degree to which comparisons between states and cultural contexts are possible. Ultimately, confirmation that this instrument assesses constructs comparably will permit valid comparisons of performance across states and provide support for the S.T.A.R. instrument’s use as a tool that can be utilized beyond assessment at the local level.

**Theoretical Basis**

The nature of this dissertation is largely rooted in theories of test construction. The term test construction is a broad concept used to refer to qualities of an assessment that are closely tied to reliability and validity. Consideration of test construction properties allows determinations to be made with regard to the extent to which an assessment is actually measuring what it purports to measure and whether it does so accurately. Furthermore, analysis of test construction elements can provide an indication regarding an instrument’s usefulness in measuring learning. Although there are many elements of test construction, those most relevant
Construct validity answers questions related to what scores on an assessment mean or signify and what the scores tell us about an individual (Thorndike, 1997). Since the term construct refers to a grouping of variables (Drummond, 2004), consideration of how a construct is operationally defined is vital for determining construct validity. Content validity refers to the degree to which items within an assessment are related to the concept being measured (Beins, 2004), while internal consistency is concerned with the extent to which items on an assessment produce consistent scores in the measurement of specific constructs (Tang & Cui, 2013). The standardization/norming sample, while not necessarily related to test construction, has significant implications in terms of being able to demonstrate that an instrument is suitable for use with various populations (Beins, 2004). For example, an assessment normed with an all middle-class, Caucasian population may produce a lack of confidence in the instrument’s ability to effectively assess the same constructs with lower-class, Hispanic subjects.

As has been discussed previously, cultural and contextual elements influence the extent to which children may engage and in prosocial behaviors. From a test construction perspective, however, elements related to context and culture may also impact how an assessment measures behaviors associated with specific constructs. While consideration of test construction elements is important to this study and provides important insight regarding the utility of the S.T.A.R. instrument, test construction elements alone do not provide sufficient information to adequately determine if the S.T.A.R. instrument defines constructs in a manner that is similar across states.
In other words, this study is seeking to determine whether the S.T.A.R. instrument ignores, or fails to effectively measure, any culturally specific definitions of the three constructs (knowledge, performance, and school connectedness) measured.

**Existing Literature**

Existing peer-reviewed literature has not addressed the specific empirical questions proposed by this study. However, two previous dissertations researched the *Be a Safety Kid* curriculum for purposes of yielding information related to validity of the S.T.A.R. assessment and the overall effectiveness of the curriculum. Graham (2012) found that, with a population of kindergarten through third grade students, the *Be a Safety Kid Curriculum* produced significant improvement in prosocial knowledge and the ability to demonstrate prosocial behaviors. No significant improvement in the feelings of school connectedness and safety were noted. Conversely, Martin (2011) found that, with a population of seventh and eighth grade students, the *Be a Safety Kid* curriculum did not produce significant improvement in prosocial knowledge and the ability to demonstrate prosocial behaviors.

With regard to validity of the S.T.A.R. instrument, results of factor analyses conducted by Martin (2011) suggested that the S.T.A.R. instrument aligned with the major constructs of knowledge and school connectedness in a sample of 7th and 8th grade students. However, only two questions on the S.T.A.R. instrument were associated with the construct of performance. Additionally, three items on this instrument were found to not align with any factors, which suggests that either these items do not measure any type of prosocial behavior or that they measure all three constructs equally. Overall, this author concluded that the S.T.A.R instrument was a valid measure of constructs for a majority of the questions created through theoretical and empirical analyses.
Graham (2012) conducted similar statistical analyses with a population of kindergarten through 3rd grade students. The results indicated that the instrument did not consistently align with the constructs of knowledge, performance, and school connectedness through face validity. Furthermore, analyses suggested a lack of differentiated constructs on the S.T.A.R. instrument, which may have been produced by low variance values in the instrument. Despite this, analyses suggest that the instrument was shown to be a stable measure of constructs overtime when pre- and post-tests were considered.

Beyond the abovementioned dissertations, most cross-cultural work on prosocial behaviors has subscribed to an etic framework, which assumes a culturally neutral perspective and also postulates that the constructs being measured have relevance across all cultures. According to Rubin and Menzer (2010), “The etic framework may cause researchers to operationally define (and thus assess) constructs in the same ways (with the same methods and measures) across cultures” (p. 1). Consequently, this may lead to neglecting culturally specific definitions of constructs when developing assessments. Within the domain of prosocial behaviors, this suggests that the definitions of helping or sharing may differ between cultural contexts and may cause inaccurate measurement of these constructs by assessments that assume a universal definition of constructs. Although it is also entirely possible that these constructs are conceptualized similarly between cultures, confirmation would require empirical analysis (Rubin & Menzer, 2010). Thus, empirical analysis is needed to determine if the constructs measured by the S.T.A.R. instrument are indeed universal between the five states, or if definitions of these constructs vary by location.
Problem Statement

The purpose of this exploratory research study is to determine if the S.T.A.R. instrument is measuring the three constructs (knowledge of prosocial behaviors and responsible reporting; performance of prosocial behaviors and responsible reporting; and the feeling of school connectedness) similarly across five states. This examination will also lead to additional information on the general utility of the S.T.A.R. instrument and its ability to similarly measure constructs across race, community type, and socioeconomic status.

Research Questions

The research questions for this study are provided below. Due to the fact that this is an exploratory study, no specific hypotheses regarding the research questions have been generated.

Research Question 1
Does the S.T.A.R. instrument measure students’ knowledge of prosocial behaviors, students’ anticipated performance of prosocial behavior, and students’ perception of school safety similarly across five states?

Research Question 2
Does the S.T.A.R. instrument measure students’ knowledge of prosocial behaviors, students’ anticipated performance of prosocial behavior, and students’ perception of school safety similarly across race?

Research Question 3
Does the S.T.A.R. instrument measure students’ knowledge of prosocial behaviors, students’ anticipated performance of prosocial behavior, and students’ perception of school safety similarly between low SES and high SES students?
Research Question 4

Does the S.T.A.R. instrument measure students’ knowledge of prosocial behaviors, students’ anticipated performance of prosocial behavior, and students’ perception of school safety similarly between students from urban and rural environments?

Summary

Violence within the school setting has been one of the main impetuses for finding appropriate interventions to curtail acts of violence, which have largely focused on prosocial behaviors. The term prosocial behavior is used to describe behaviors that are intended to benefit others and can include things such as helping, sharing, and rescuing. Although the study of prosocial behaviors originated with William McDougal and Louis Murphy, later work by Piaget and Kohlberg helped to lay the theoretical foundation for these behaviors. Most theories suggest that prosocial behaviors develop according to specific trajectories, but there is research suggesting that cultural and contextual factors can also influence an individual’s propensity to act prosocially.

In terms of interventions, the Be a Safety Kid curriculum is a school-based violence prevention program designed around the concept of “responsible reporting”. This curriculum can be used with kindergarten through 8th grade students and includes the S.T.A.R. instruments for purposes of serving as a pre- and post-test. Data was collected with students from five different states. Due to the fact that the S.T.A.R instrument is the sole indicator of progress in relation to the curriculum, it is of great importance to determine the extent to which this instrument is capable of measuring constructs similarly across states. In essence, the purpose of this research is to determine if the S.T.A.R. instrument is measuring the three constructs (knowledge of prosocial behaviors and responsible reporting; performance of prosocial behaviors...
and responsible reporting; and the feeling of school connectedness) similarly across five states and across cultural contexts as they relate to race, high/low SES, and urban/rural communities.
CHAPTER II

LITERATURE REVIEW

Prosocial behaviors encompass a wide range of actions aimed at benefitting others above oneself and can include actions such as helping others, rescuing others, or cooperating with others (Batson & Powell, 2003). These behaviors are of great importance, especially for children, as research has demonstrated a link between prosocial behavior and the development of social competence and acceptance by the child’s peer group (Denham & Holt, 1993). In essence, those who engage in behaviors intended to benefit others are more widely accepted by their peers and display more adept interpersonal skills. Furthermore, there is evidence to suggest that the propensity to display prosocial behaviors may serve as a protective factor and decrease the likelihood of externalizing problems, such as aggression and delinquency, later in life (Hastings, Zahn-Waxler, Robinson, Usher, & Bridges, 2000). The positive outcomes associated with prosocial acting are likely due to the fact that individuals who use more prosocial methods for solving conflicts generally offer more effective and relationship-enhancing resolutions rooted in persuasion rather than aggression (Mayeux & Cillessen, 2003).

The purpose of this chapter is to provide the reader with a framework for understanding prosocial behaviors by reviewing relevant literature. This chapter will also explore the theoretical underpinnings of prosocial behaviors, the developmental trajectory of prosocial behaviors in children, and cultural considerations related to these behaviors. Focus will then shift to a discussion of commonly used methods for assessing prosocial behaviors. Information related to test construction elements will then be presented, as the empirical focus of this dissertation is ultimately on the effectiveness of an assessment designed to assess prosocial behaviors.
Evolution of the Construct of Prosocial Behavior

Focus on prosocial behaviors originated with William McDougall (1908), who is credited with composing the first text in social psychology (Batson & Powell, 2003). Emphasis on these behaviors resulted from McDougall’s poignant statement that, “The fundamental problem of social psychology is the moralization of the individual by the society into which he is born as a creature in which the non-moral and purely egotistical tendencies are so much stronger than any altruistic tendencies” (p. 16). Here, it would seem that McDougall recognized early on that the display of altruistic behaviors is often obstructed by society’s proclivity to focus on individual needs and desires. Although his statement was thought provoking, it appeared to go unaddressed by a field that was not yet convinced on the importance of prosocial behaviors and altruism. In addition to this statement, McDougall also proposed that prosociality was the product of emotions shaped by instinctual parenting during early interactions. Lacking an audience ready to focus on prosocial behaviors, however, the topic faded into the minutiae of social psychology until the 1930’s.

At that time, Lois Murphy (1937) began exploring sympathy in preschool children and found that children, as young as two years of age, had the ability to display care and defense for peers and were also able to provide warning to others when faced with some kind of danger. In light of current definitions of prosocial behavior, which will be discussed later, it appears that Murphy attempted to categorize prosocial behaviors in a manner similar to contemporary researchers. Despite this, her research, as well as other research in the field, tended to focus only on the concept of sympathy. Consequently, research during this time produced an understanding of prosocial behaviors that was limited in scope. However, a significant event in the 1960’s would reignite the focus on helping behaviors.
Kitty Genovese was a New York City woman who was brutally stabbed and fatally wounded in 1964 (Batson & Powell, 2003). The process of her murder, which lasted for more than half an hour, was witnessed by 38 of her neighbors in the Kew Gardens area of Queens, New York (Batson & Powell, 2003). Initial reports suggested that, despite her screaming and pleading for help, no one intervened or felt compelled to at least notify the police. This incident sparked outrage at a national level and served as an impetus for trying to determine why people act, or fail to act, in particular situations. What followed was an increase in research dedicated to finding an answer to this question.

Shortly after Kitty’s murder, Darley and Latané (1968) conducted a study designed to gather more information related to bystander intervention and diffusion of responsibility during emergency situations. Subjects overheard an epileptic seizure and then were either led to believe that they were the only ones who heard it or that 1 or 4 unseen others were also present. As expected, the perceived presence of other individuals reduced the subject’s feelings of personal responsibility and lowered their speed of reporting. Although the authors point out that inaction by bystanders is often explained by apathy, anomie, or alienation, this research suggests that an individual’s propensity to act may be more related to the bystander’s response to other observers rather than indifference to the victim. Despite these findings being expected, it might be posited that more bystanders would equate to a greater possibility of intervention. Superficially, this makes sense. However, this general assumption likely fails to consider the complex processes at work during an emergency situation, which can include individuals’ cognitions and levels of moral reasoning.

During the 1960’s and 1970’s, an abundance of research on prosocial behaviors was conducted using the paradigms outlined by Kohlberg and Piaget with regard to internal moral
conflicts (Kohlberg, 1984; Piaget, 1965). According to Piaget’s cognitive-developmental theory of moral reasoning, he concluded that children go through two discrete stages in how they consider morality (Santrock, 2006). In the first stage, which occurs from age 4 to 7 and is termed *heteronomous morality*, children tend to view justice and rules as unchangeable and removed from the control of individuals. In the second stage, *autonomous morality*, the child becomes aware that rules and laws are made by people and that individuals should consider the actor’s intentions when judging an action (Santrock, 2006). It is noted that the second stage generally begins around 10 years of age and that, from between 7 and 10 years of age, children enter a transition period that is characterized by some features of both stages. Piaget also argued that, as they develop, children become more sophisticated in the manner in which they think about social matters, especially in relation to cooperation with others.

As children enter the autonomous stage and begin to focus more on intentions than consequences, Piaget (1932) emphasizes a greater role of cognition by proposing that cognitive processes assist with emotional expression and experience, which provide feedback with regard to whether the child or other individuals need to adjust or continue behavior that is goal directed. This would seem to suggest that the analysis of intentions underlying prosocial behavior must first ensure that children are being assessed in a manner that is consistent with their developmental level and, ideally, measured after the child has consistently demonstrated that he or she has entered the autonomous stage. Assessment under the aforementioned conditions would likely yield better information due to the fact that children could more precisely describe their moral reasoning with a greater consideration of intentions and without such a fervent focus on rules.
Building upon the work of Piaget, Kohlberg (1984) proposed a theoretical model of moral reasoning that was comprised of three levels and six stages. The first level, which is the *premoral* level, encompasses the first and second stages. Within the first stage, the child demonstrates a punishment and obedience orientation, while the second stage resembles naïve instrumental hedonism (Levine, 1976). Once the child moves into the second level, which is termed, *morality of conventional role conformity*, the child begins to enter stage three, and the orientation shifts to having good relations with others and obtaining the approval of others. In stage four, which is also at the second level, the child is focused on authority, fixed rules, and the maintenance of social order. The third and final level, *morality of self-accepted moral principles*, includes the fifth and sixth stages. Stage five brings about the view that the right action is defined in terms of individual rights, while stage six entails an orientation towards universal ethical principles and the respect for human beings as individual persons (Moroney, 2006).

The work of Piaget and Kohlberg provided the groundwork for how prosocial behaviors are understood. Currently, there is an implicit recognition that there are multiple variables that are predictive of prosocial behaviors. In the context of this discussion, moral identity, which consists of cognitive schemas related to the mental representation of one’s moral character, is found to predict a variety of prosocial behaviors across various settings (Winterich, Aquino, Mittal, & Swartz, 2013). Such a relationship further delineates the role that cognitions play in prosocial behaviors. As would follow, research steered by the cognitive-developmental approach places a greater emphasis on cognitive-motivational elements of a person’s behavior and describes altruistic actions in relation to cognitive motives (Penner, Dovidio, Piliavin, & Schroeder, 2005). However, other theorists subscribe to a social learning approach in studying
these behaviors, which entails greater concentration on overt, observable behaviors and less attention to cognitive processes (Eisenberg, 1982).

Regardless of the theoretical orientation utilized to investigate these behaviors, incidents such as the one involving Kitty Genovese highlight the importance of ascertaining why individuals help, or fail to help, others in emergencies. Although theorists and researchers have demonstrated interest in studying prosocial behaviors, it appears that little effort was made to actually define these behaviors. Rather, they were referred to holistically as “prosocial” or “bystander” behaviors. However, as researchers continued to explore these behaviors, there became a greater awareness that various subsets of behaviors coalesced to form the general construct of prosocial behaviors. The next section is dedicated to exploring how these behaviors have been defined over the years.

**Contemporary Characterizations of Prosocial Behavior**

Great variability exists with regard to how this term is defined. The absence of agreement regarding what prosocial behaviors actually are is complicated by the fact that the word *prosocial* cannot be found in many dictionaries and that it was actually developed by social scientists to serve as an antonym to the word *antisocial* (Batson & Powell, 2003). Consequently, those seeking a definitive answer to what these behaviors are will likely not even be able to find the answer within the most trusted source of lexicon knowledge. Despite this, however, existing literature authored by some of the most prominent individuals in the domain of prosocial behaviors provide a relatively solid framework for facilitating our understanding of what these behaviors entail.

The general consensus is that prosocial behavior is an overarching construct that is multifaceted in nature and a concept used to describe behaviors intended to benefit others.
(Eisenberg & Fabes, 1998). Given the internal locus of these behaviors, early theorists such as Darley, Latané, and Wispé pointed out that effectively studying these behaviors may be difficult due to the challenge of quantifying them. Furthermore, Zeldin, Savin-Williams, and Small (1984) indicated that operationally defining these behaviors is problematic given the fact that they represent a variety of behaviors associated with the concern for others and also entail adherence to the norm of social responsibility. The authors also propose that an absence of research conducted in naturalistic settings has exacerbated problems related to operationalization for investigational purposes. In the past, researchers have studied prosocial behaviors as an all-inclusive concept, without the recognition of the fact that there are various subsets of behaviors that comprise the broad category of prosocial behaviors (Greener, 2000). However, there is now a greater understanding that prosocial behavior encompasses an array of behaviors such as helping, sharing, rescuing, and comforting (Zeldin et al., 1984).

One of the most common concepts associated with prosocial behavior is altruism. At times, the word *altruism* has been utilized for the purpose of referring to a subset of prosocial behaviors. Examples of this may include instances of self-sacrificial helping or helping others when the situation is void of obvious external rewards (Batson & Powell, 2003). However, these authors emphasize the fact that this type of usage may be inappropriate given the fact that altruism is a motivational concept in the sense that it refers to the motivation to increase the welfare of other individuals. They further suggest that there is no one-to-one correspondence between altruism and prosocial behavior. That is, prosocial behavior does not need to be motivated by altruism and altruistic motivation does not necessarily produce prosocial behavior. This is further clarified by Anker and Feeley (2011) who make the distinction that, while altruistic behaviors are never intended to be self-serving, prosocial behaviors may vary with
regard to what motivates individuals to assist others. For example, prosocial behaviors may have an egotistic motivation (i.e., help others to benefit oneself) or an altruistic motivation (i.e., helping others without concern of self-rewards).

With the understanding that whether prosocial behaviors are altruistic or not is dependent upon the underlying motivation of the action (Eisenberg & Mussen, 1989), it should be noted that it is typically impossible to differentiate between actions that are altruistically motivated and actions that are motivated by less righteous concerns (Eisenberg, Fabes, & Spinrad, 2006). Thus, although there is a qualitative difference between prosocial and altruistic behaviors, researchers often focus on the broader domain of prosocial behaviors due to the difficulty in parsing out the differences. Despite these distinctions and difficulties, it would seem logical to suggest that altruism provides a foundation for prosocial behaviors due to the intrinsically motivated origin to help others with no concern for personal consequences. In addition, Eisenberg et al. (2002) indicate that psychological processes such as moral reasoning, sympathy, and empathy may also serve as an impetus for prosocial behavior. In that regard, consideration of these additional traits increases the already expansive inventory of prosocial behaviors used to define the need to help others.

In the attempt to define the correlates and structure of prosocial behaviors, Carlo and Randall (2002) produced one of the most cited studies on this topic. In their research with 249 college students (104 males, 145 females; mean age = 19.89 years), participants completed the Prosocial Tendencies Measure (PTM), which measures six types of prosocial behavior: altruistic, compliant, emotional, dire, public, and anonymous. The authors also had participants complete measures of sympathy, perspective taking, personal distress, social desirability, global prosocial behaviors, social responsibility, ascription of responsibility, vocabulary skills, and prosocial
moral reasoning. Correlations were then calculated between the six factors of the PTM and the variables derived from the other measures. Overall, results suggested that prosocial behavior is best conceptualized as a set of discrete, albeit interrelated, dimensions rather than a global construct. For example, participants who endorsed more helping in public contexts were more likely to: display more concern for their own needs, have a stronger desire to obtain the approval of others, demonstrate less sensitivity to the needs of others, ascribe responsibility to others, and engage in less complex forms of reasoning and perspective taking. Despite these findings, one must leave open the possibility that additional types of prosocial behaviors exist and that this study does not provide an exhaustive list of the various forms prosocial behavior can take.

Carlo and Randall’s effort to quantify prosocial behaviors is unquestionably noteworthy. Unfortunately, however, the limitations of their study tend to be representative of most research conducted on prosocial behaviors. To a large degree, research in this domain has predominately concentrated on behaviors elicited from environmental prompts with samples that were primarily late adolescents. Consequently, the results of these studies lack generalizability and tend to produce narrow definitions of prosocial behaviors that are derived from non-naturalistic settings. However, some studies have attempted to correct these flaws in an effort to yield a better understanding of prosocial behaviors.

Jackson and Tisak (2001) attempted to broaden the definition of prosocial behaviors by conceptualizing these behaviors into four types: helping (responses to people who have incurred unintentionally produced negative consequences), sharing (giving up one’s own resources to benefit another), cooperating (coordinating actions to obtain a specific goal) and comforting (actions taken to improve the overall mood of another person). In their study with 83 children (7-12 years of age) from working and middle-class families, the participants were read a set of
stories involving same gender friends. Each story was then followed by a set of questions that focused on expected behavior (would the child act prosocially), obligation (does the child feel it is wrong to not act prosocially), self-evaluation (how the child would feel about himself or herself if he or she did not act prosocially), and peer evaluation (how the child would have felt if his or her peers thought he or she was mean for not helping). The subjects’ responses to the questions were then scored on a scale from 1-4. Results indicated that children differentiated among the prosocial behaviors and that age differences were noted. For example, the 7- and 8-year olds were less likely to comfort, think it was wrong to not comfort, and were less concerned about negative self-evaluation when the prosocial behaviors involved helping sharing, or cooperating. However, the 9- and 10-year-olds were even less likely to comfort a friend, more likely to feel good for not comforting a friend, and less concerned with others thinking they were mean for not helping.

Although this study provides insightful findings given the focus on prosocial behavior in children, the authors note that some children, especially the younger ones, may have struggled to understand the questions due to how they were worded. Additionally, the children were not asked to provide any rationale for their evaluations. Failure to do so squandered the opportunity to gain awareness of the cognitive processes the children utilized to arrive at their decisions. Consideration of these variables may have generated richer data and provided a glimpse into the internal workings of prosociality.

Iannotti (1985) conducted a similar study for the purposes of categorizing prosocial behaviors. With 52 preschool children (21 girls and 31 boys, mean age = 59 months), Iannotti assessed prosocial behaviors using naturalistic observation, teacher ratings, and structured measures of perspective taking, empathy, and prosocial behaviors. As with Jackson and Tisak’s
study, prosocial behaviors were categorized into four domains: sharing, helping, cooperating, and comforting. Correlations were then calculated to determine the relationship between selected observational categories and laboratory measures. Results indicate that rates of prosocial behaviors were actually quite low. This seems to suggest that these types of behaviors tend to occur less in naturalistic settings where there is an absence of specific cues that would be found within a laboratory setting. Results also suggested that different categories of prosocial behaviors tended to be unrelated within the same setting and that there was not a consistent relationship between perspective taking and acting prosocially. The authors further point out that, for preschoolers, the ability to understand the cognitions or views of someone else does not guarantee prosocial behavior.

Unlike previous studies, Avgitidou (2001) researched prosocial behaviors by relying on a more qualitative method via the use of an ethnographic approach. To do so, the author selected two kindergarten classrooms from a middle class community, which each had 20 children (mean age = 4 years, 10 months). Methods included daily observations, semi-structured interviews with the children, and sociometric techniques at the beginning and end of the observation periods. Although Avgitidou did not seek to quantify prosocial behaviors as others had done, he found that one of the commonly accepted peer culture concerns is prosocial behavior in the form of empathy. This entailed looking after the children in the room who are younger or smaller. In essence, children recognized that peers were not able to do certain things and subsequently helped them. Furthermore, findings suggest that the nature of relationships among children can serve as a motive for displaying advanced prosocial behaviors designed to understand the other’s perspective and sustain the relationship. Despite not attempting to define types of prosocial
behaviors, the discoveries within the context of this study facilitate the understanding of antecedents to prosocial behavior.

Even further removed from the laboratory setting was a work produced by Penner, Dovidio, Piliavin, and Schroeder (2005). In their effort to define prosocial behaviors, these researchers conducted a thorough review of the literature to produce a multilevel perspective on prosocial behavior. Their examination led to the identification of three levels of analysis of prosocial behavior: the “meso” level (the study of helper-recipient dyads in the context of a specific situation); the “micro” level (the study of the origins of prosocial tendencies and the causes of variation in these tendencies); and the “macro” level (the study of prosocial behaviors that occur within the context of organizations and large groups). Based upon the differentiation of these levels, the authors ultimately suggest that a better understanding of prosocial behaviors may be obtained through paying more attention to proximal causes of prosociality. Additionally, these authors propose that, rather than viewing prosocial behaviors such as helping, cooperating, and volunteering as endpoints, these behaviors should be viewed as parts of an ongoing process. For instance, this could involve the understanding that prosocial behavior may be a component of forgiveness, which is a contributor to stable relationships, and a key influence of reconciliation, which can bolster collective identity and action.

Although existing studies have made a valiant attempt to define and categorize prosocial behaviors, more research is needed in this area to not only clarify the nature of these behaviors, but to also further analyze cognitive processes and mediating variables associated with prosociality. What is clear at this point, however, is the importance and implications of these behaviors. Future research should attempt to move beyond the current theoretical understanding of these behaviors and potentially study these behaviors as part of an ongoing process as
Development of Prosocial Behaviors in Children

Prosocial behaviors serve great importance to the social and emotional development of the child. There is now an understanding that children who engage in prosocial behaviors tend to have greater social competence and are also more readily accepted by their peer group (Denham & Holt, 1993). At a very early age, research has shown that newborn infants display some form of empathic response as evidenced by their reactive crying to hearing the cries of another infant (Martin & Clark, 1982). Additionally, Dondi, Simion, and Caltran (1999) found that newborn infants appear to display more distress to the cries of another infant than to their own and that this may suggest some type of biological predisposition to experience a basic form of empathy. This presupposition has been questioned by other researchers (e.g., Eisenberg & Lennon, 1983) who suggest that, rather than representing an empathic response, infants may simply find another infant’s cry to be more unpleasant than their own.

At approximately 6 months of age, infants will, at times, react to the cries of another child, but will more often than not either ignore the cry or simply gaze at the other child (Hay, Nash, & Pedersen, 1981). When the child reaches 38 to 61 weeks of age, the child may continue to cry when another child cries, but they will occasionally display incongruous behaviors such as laughing or smiling when another child is crying (Zahn-Waxler & Radke-Yarrow, 1982). As the child enters toddlerhood between 12 to 18 months of age, the child begins to purposefully react to the negative emotions of others. Although some may continue to display distress reactions, other children will begin to engage in rudimentary prosocial behaviors such as a pat on the back or saying that it will be ok (Zahn-Waxler, Robinson, & Emde, 1992).
Hoffman (2000) suggests that the heightened ability to respond to another’s emotional distress is associated with the development of self-awareness and is also likely influenced by the fact that the child now views others as psychological agents. By the end of their second year of life, children begin to use words connected with emotions and become more adept at effectively reading the emotional states of others (Nichols, Svetlova, & Brownell, 2010). Despite this ability, however, naturalistic observations of children reveal exceptionally low rates of responding to the distress of others (Lamb & Zakhireh, 1997). Similarly, Caplan and Hay (1989) found that, although children may be aware of another’s distress, they are not usually responsive to it. Such a phenomenon may suggest that there are other factors influencing whether a child chooses to act prosocially when confronted with the distress of another. For example, Demetriou and Hay (2004) posit that a child’s decisiveness to act may be impacted by contextual factors such as his or her familiarity with the peer or the environment in which the distress occurs (i.e., day care vs. home).

Around age 4 or 5, children continue to build upon their capacity to understand emotions in others. Their skill in identifying emotional expressions and situations strengthens and they are much better equipped to engage in dialogue related to their emotions and the emotions of others (Roberts & Strayer, 1996). Children also begin to develop perspective taking skills during this time, which serves as an essential element in prosocial behaviors (Malti, Gummerum, & Buchmann, 2007). In a meta-analysis conducted by Eisenberg and Fabes (1998), there was evidence that prosocial behaviors generally increased during the preschool and early school-age years. They also found that school-age children demonstrated a higher rate of prosocial behaviors than what preschool children did. This seems to provide some confirmation that, as children grow older and develop increasingly complex psychological processes, the inclination
for prosociality grows. There also appears to be stability in the level of children’s prosocial behaviors, as children who entered kindergarten evidencing a certain degree of helpful behavior were found to complete primary school with similar levels (Cote, Tremblay, Nagin, Zoccolillo, & Vitaro, 2002).

Later in childhood, children continue to expand their repertoire of skills related to prosocial behaviors and emotional competence. At approximately 11 years of age, children can move beyond viewing a person or object in an exclusively positive or exclusively negative manner (Hay & Cook, 2007). In theory, this may lead to an increase in prosociality since it may be assumed that, as a result of the ability to utilize a mixed-valence approach, children begin to integrate cognitive, emotional, and interpersonal factors into the analysis of situations. In other words, they no longer view people or objects with a black and white perspective. In the mid to latter part of adolescence, there is evidence that children of this age exhibit more prosocial tendencies that what elementary school students do (Fabes, Carlo, Kupanoff, & Laible, 1999). However, other research has found that, while adolescents do tend to engage in more prosocial responding than do younger children, this was not true of all types of prosocial behaviors. For example, studies have shown that behaviors related to sharing and donating (but not helping), increased in adolescents, while there was evidence to suggest that helping victims of aggression tended to decline across adolescence (Eisenberg & Fabes, 1998; Jacobs, Vernon, & Eccles, 2004).

In an effort to provide a more structured approach to understanding how prosocial behaviors develop, Hoffman (1982, 2000) proposed a four-level paradigm that elucidates the emergence of these behaviors. In the first stage, infants display basic empathic responses that are expressed through global empathy. Since they are unable to differentiate between their own
distress and that of another, infants will often react by crying when they hear another baby crying. At approximately the end of their first year, children enter the second level and have begun the process of understanding that they are separate beings from others. However, this sense of separation is very rudimentary and children will therefore continue to respond to the distress of another by crying. When the child enters his or her second year of life, he or she will start to react to other’s distress by utilizing physical contact as a modality of reassurance. At this point, Hoffman posits that the child has entered the third stage. During the fourth stage, which occurs later in their second year, the toddler displays skills related to perspective taking and has the conception that the feelings of others may differ from his or her own. As a result, the nature of their prosocial responding is less egocentric and more oriented towards the true needs of others. What Hoffman makes clear within his theory is the role that cognitive maturation plays in prosocial behaviors. He advances the view that there is a positive relationship between cognitions and prosociality in the sense that, as the child’s cognitive processes become more complex, he or she is better able to respond to others’ distress with concern.

The general consensus in the literature is that prosocial behaviors increase, both in frequency and intricacy, as the child becomes older. Within the context of this discussion, research suggests that empathy, perspective taking, and cognitions are highly influential to prosocial acting and are directly tied to the developmental trajectory of these behaviors. The research examined in this section is largely concerned with psychological processes involved in prosocial behaviors. However, if one is to have a comprehensive understanding of prosocial behaviors, it is important to move beyond concern for only biological determinants by exploring the cultural components that may directly impact an individual’s inclination to act prosocially.
Culture, Moral Reasoning, and Prosocial Behavior

As defined by Matsumoto (1997), culture encompasses the set of behaviors, beliefs, values, and attitudes that are communicated from one generation to the next. Culture has been recognized for its significant role in human behavior for many years and dates back to the times of Hippocrates (Dona, 1991). However, as a whole, psychological theories tend to ignore the role that culture plays in the development of important psychological processes. This may lead to a “one-size fits all” approach that fails to account for how external forces may impact internal development. According to Ellis and Boyce (2008), new empirical evidence and theoretical considerations indicate that the development of children cannot be accounted for by either contextual or biological factors. Rather, the authors posit that acknowledging the interaction between biology and context/culture is essential to understanding child behavior and development. The position of Ellis and Boyce is supported by research confirming that elements of culture can influence prosocial behaviors and an individual’s sensitivity to act prosocially. Unfortunately, most cross-cultural work on prosocial behaviors has subscribed to an etic framework, which is based upon a culturally neutral perspective wherein cultural impacts receive little consideration (Rubin & Menzer, 2010).

As discussed previously, moral reasoning and development play an integral role in prosocial behaviors. Although Kohlberg’s (1984) theory was previously discussed in relation to how moral reasoning develops, his theory is also a salient illustration of the extent to which psychological theories overlook culture. Kohlberg (1971) emphatically posited that moral judgment develops in basically the same way across cultures and poignantly stated that almost all individuals in all cultures use the same basic moral categories, concepts, principles, and values. Furthermore, he advanced that all individuals in all cultures go through the same sequence of
development, but that there is some variation in the rate and terminal point of development. Despite his claims of universality, there is a body of evidence to support that fact that his presuppositions were not entirely correct.

In 1985, Snarey found partial support for Kohlberg’s position by confirming that Stage 1 to Stage 4 were in existence virtually universally when the age range and sample size of the study were taken into consideration. However, in his analysis of 27 different countries, Snarey also found that Stages 5 and 6 were not as universal as Kohlberg initially claimed. For example, Snarey indicated that all folk and village cultural groups failed to exhibit any Stage 5 or Stage 6 postconventional reasoning. Conversely, there was also an indication that Kohlberg’s theory, according to Snarey, failed to capture higher moral ideals that are embraced by some cultures. Based upon this evidence, Snarey concluded that, “The failure to find a particular stage in all studies of a particular type of culture group could indicate that the stage is culture specific” (p. 204).

There has been additional research conducted in an effort to determine differences in moral reasoning between cultures. Miller and Bersoff (1992) conducted a two-session study that examined Indian and American adults’ and children’s’ reasoning about moral dilemmas related to conflicts between interpersonal and justice expectations. Results indicated that the majority of Indians gave priority to interpersonal expectations, while most Americans tended to give priority to justice expectations. Furthermore, Indians were more inclined to categorize their conflict resolutions in moral terms, whereas Americans were more apt to choose resolutions that were self-benefitting. The authors postulate that Indians appear to possess a postconventional moral code in which interpersonal responsibilities are given equal importance to justice obligations and, in some instances, actually take precedence over justice obligations. Additionally, they
state that a personal morality of interpersonal responsiveness is associated with cultural views that are highly rights-oriented, such as those emphasized in the United States.

Research conducted by Keller, Edelstein, Krettenauer, Fu-xi, and Ge (2005) also found differences in moral reasoning from a cross-cultural perspective. In their comparison of a Western culture and an Asian culture, they found that the Asian culture’s moral reasoning was rooted more in altruism, whereas the Western culture displayed moral reasoning that was related to self-interest. Additionally, the authors found that, when presented with moral dilemmas, the Asian culture interpreted the dilemma as a conflict between altruism and friendship, while those from the Western culture interpreted the situation as a conflict between self-interest and friendship. Such differences were found across ages and suggest noteworthy differences in moral development between cultures.

Research has shown differences in moral reasoning between Western and non-Western cultures and has also demonstrated that Kohlberg’s theory may not be as universal as once was previously thought. Since moral reasoning plays a role in prosocial behaviors, it is also necessary to explore the potential impact that culture has on prosocial development and the propensity to engage in prosocial actions. Given the significant diversity within the United States, failure to consider how cultural factors impact prosocial behaviors produces an incomplete conceptualization of these behaviors as a whole and, as will be discussed later, also has implications for the assessment of these behaviors. Here, culture is concerned with concepts such as values, religion, political affiliation, community type, race, and socioeconomic status, which may influence prosociality. Although an exhaustive discussion of all cultural factors that could potentially influence prosocial behaviors is beyond the scope of this paper, an examination of several select factors is provided in order to facilitate a more holistic understanding of
prosocial behaviors and to further illustrate the interaction between cultural elements and prosocial behaviors.

**Family Values and Goals**

According to Eisenberg, Fabes, and Spinrad (2006), an important element of the self is one’s values. These values are often communicated by the family and help shape the manner in which a child develops. Research has demonstrated that children whose parents report holding prosocial values have been rated by their peers as also engaging in higher levels of prosocial acting (Hoffman, 1975). This was supported by a study conducted by Calderón-Tena, Knight, and Carlo (2011) who found that parenting behaviors that promoted prosocial tendencies generally produced children who were more likely to engage in prosocial behaviors. However, some research (Turner & Harris, 1984) has found no relationship between parental emphasis of prosociality and the extent to which their children respond in a prosocial manner. Other research, such as the study conducted by Bryant and Crockenberg (1980), has indicated that, while there is a relationship, it is not as prominent as what some studies have shown.

Despite the various findings with regard to family values and children’s prosocial behaviors, it is difficult to deny the role that family socialization practices play in the role of social, emotional, and behavioral development. However, the fact that some research has not shown a link between these variables seems to reaffirm the role that internal processes play and appears to stress nature over nurture. This may be supported by instances in which children, raised by parents who engage in mainly self-serving behaviors, ultimately display higher than expected levels of prosocial behaviors. Family values and goals can also encompass other belief systems held by parents that tend to be passed down to children, such as political orientation and
religious beliefs. Here, the question centers on the extent to which these different beliefs can impact prosocial acting.

**Political Orientation.** Theoretically, an individual’s political affiliation and preferences may impact the likelihood of prosocial acting. Ergo, it is also possible that children growing up in families with certain political orientations may be more, or less, inclined to engage in prosocial acting. There is some evidence that political preference and voting habits may be rooted in an interpersonal system that directly influences a person’s social interactions in groups and dyads (Van Lange, Bekkers, Chirumbolo, & Leone, 2012). For instance, liberals tend to endorse a greater sympathy for those less fortunate than them when compared to conservatives (Zettler & Hilbig, 2010) and also indicate higher levels of self-reported altruistic behavior (Zettler, Hilbig, & Haubrich, 2011). Despite this potential connection between politics and prosocial behaviors, only two studies have directly examined the relationship between these variables.

Sheldon and Nichols (1994) sought to identify the nature of the relationship between social value orientation and political preference using a sample of law students in the U.S.A. Overall, the results indicated that 49 out of 64 (77%) individuals who identified themselves as engaging in high levels of prosocial behavior also identified themselves as a Democrat, rather than a Republican (23%). Furthermore, Republicans were consistently found to favor extrinsic values (e.g., money, popularity, image), while Democrats, as a whole, tended to be higher on intrinsic values such as intimacy, help, and growth. In a similar study, Van Lange et al. (2012) set out to determine if there was a difference in prosocial behavior between liberals and conservatives. With their sample of 401 participants (194 men, 207 women), which were mainly university students, the investigators administered a questionnaire designed to measure social
value orientation and the need for structure. Overall, results indicated that those with a more liberal orientation were more likely to demonstrate other-regarding behaviors than the conservatives. To help bolster their results, the authors conducted three other independent studies: two in Italy and one in the Netherlands. Comparable results were obtained in those studies, which further suggests that, as a whole, liberals are more likely to engage in prosocial behaviors than their conservative counterparts.

Existing research clearly highlights the fact that political preferences can be instrumental in the display of helping behaviors. It can be reasoned from this discussion that children who grown up in households that are predominately liberal or Democratic may be more inclined to demonstrate prosocial behaviors. At the present time, however, there is no direct empirical evidence to suggest that this is, or is not, the case.

**Religion.** Potentially, elements associated with various religions may impact the development and engagement in prosocial behaviors. As with political affiliation, it is also possible that children growing up in religious households have a greater propensity for engaging in prosocial behavior. Research up to this point has not conducted comparison studies to determine if there are differences in prosociality between religious groups. For example, no studies have examined whether or not Catholics are more likely to help than Baptists. Rather, the majority of research conducted on this topic has simply focused on whether or not religious people are more likely to be prosocial than individuals who are not religious. Even so, the literature in this area is limited to only a few empirical studies.

Overall, studies examining the relationship between religion and prosocial behaviors have found that there is a positive correlation between religiosity and helping behaviors. Results from research conducted by Furrow, Kind, and White (2004) revealed that there was a positive
relationship between religious self-understanding, personal meaning, and prosocial personality. Similarly, Saroglou, Pichon, Trompette, Verschueren, and Dernelle (2005) found that religious individuals endorsed higher levels of altruistic behavior and empathy, and were also perceived by others as being prosocial. It is noted, however, that results from these studies were largely obtained through self-report. As such, the extent to which social desirability impacted responses is called into question. In theory, individuals who view themselves as religious may simply associate religion with helping behaviors and subsequently endorse items reflecting prosociality when in fact they do not actually engage in those behaviors. Thus, the absence of rigorous methods in these studies undoubtedly poses concerns related to validity.

Galen (2012) has argued against the relationship between religion and prosocial acting and postulates that the religious prosociality hypothesis is simply a fallacy. He, too, makes mention of the heavy reliance on self-report measures when studying this relationship and concludes that the observed effects are not real. Rather, the results only reflect stereotypes and are likely produced by secular psychological effects. Saroglou (2012), in his response to Galen, defends the use of self-report measures and indicates that social desirability should not be seen as a response bias, but as an element of personality. Saroglou further contends that, on the basis of existing evidence, it is premature and theoretically problematic to conclude that the religion-prosociality link is false or that it exists only in the minds of religious people.

Community Type

It is possible that prosocial behaviors are also impacted by the type of community in which an individual resides. For children, the question is whether or not living in a certain community type can impact the extent to which they engage in prosocial behaviors. Here, the differentiation is between those who live in a rural environment versus those who live in an
urban setting. Typically, research has shown that urban communities often experience more concerns related to violence and quality of housing than what those in a rural setting do (Denhardt & Glaser, 1999). Additionally, overcrowding may be a concern, in which individuals have a sense of always being surrounded by others. Taken as a whole, these elements of urban living can produce a high level of stress and possibly result in more self-serving behaviors and less prosocial acting. Conversely, rural neighborhoods tend to present with much lower levels of these problems. These communities are often close-knit, which may lead to a greater cohesion among citizens. Thus, each type of community introduces characteristics that may either increase, or decrease, the propensity for prosocial acting.

In a study conducted by McMahon et al. (2013), the authors were concerned with determining the extent to which exposure to community violence contributed to the development of prosocial or aggressive behaviors. Through the use of longitudinal multilevel modeling, the study investigated predictors of self-, peer-, and teacher-reported aggressive and prosocial behavior in a sample of 266 urban, African American adolescents over two years. As was expected, results indicated that adolescents who were exposed to high levels of violence and aggressive beliefs were less likely to engage in prosocial behavior and were more likely to evidence aggression. More importantly, however, the researchers found that the presence of violence prevention skills was related to fewer instances of aggression and a higher incidence of prosocial behaviors. Thus, although the results suggest that urban adolescents may be more predisposed to avoid prosocial acting due to environmental factors, the findings emphasize the importance of designing and implementing interventions intended to teach violence prevention skills to help mitigate negative environmental influences.
Amato (1983) undertook an examination of helping behavior in urban and rural communities by conducting a series of six studies in a randomly selected sample of 55 Australian towns and cities stratified by geographical isolation and size of population. Each study involved a different contrived situation for the purposes of eliciting possible prosocial acts. For example, one study examined helping behaviors that resulted from an individual dropping envelopes, while another involved responses to an individual with a hurt leg that appeared to be bleeding. Overall, results indicated a negative association between population size and helping in four of the manufactured situations. The fifth situation produced a positive association between size and helping, while the sixth resulted in no directional association. Similar to the study conducted by McMahon et al., these results generally suggest that an individual’s proclivity for engaging in helping behaviors is influenced by the nature of the community in which they reside.

Steblay (1987) moved beyond the work of Amato by conducting a meta-analysis of helping behaviors in rural and urban environments. To do so, she conducted an extensive search for relevant literature which ultimately produced a final sample that included 32 published reports, 2 conventional papers, and 1 unpublished report. Contained within her sample were a total of 65 tests of the hypothesis that “country people are more helpful than urban people.” Statistical analyses for this meta-analysis included calculating Z-scores, effect size, and fail safe N. Results of the 65 analyses produced an effect size of 0.29 and indicated a greater prosocial response in the rural condition that was unlikely to be produced by chance alone. Overall, this would suggest that helping is more likely to occur in nonurban areas. However, the authors note that differences were only found when the population size approached 300,000 people.

Kamal, Mehta, and Jain (1987) were also interested in studying the differences in prosocial acting in urban and rural settings. Fifty postgraduate male Indian students (aged 22-25
years) were identified as the sample. In this study, participants were asked to imagine themselves in either an urban or rural setting and then subsequently respond to four requests for help either over the phone or in a face-to-face interaction. Results indicated that subjects who imagined a rural setting were higher in helping behaviors and that prosocial acting was lower for urban settings for both telephone and face-to-face requests for help. Despite these findings, there are serious methodological concerns with this study. Given the reliance on imagination, it would not seem feasible for subjects to put themselves in a mindset that would truly mimic immersion in a rural or urban setting. Furthermore, generalizability is limited as a result of the all-male, Indian sample.

In a slightly different examination of helping as it concerns rural and urban settings, Wilson and Kennedy (2006) conducted a study to assess how urban and rural individuals would respond to a women in need who was either dressed professionally or casually. The sample of convenience consistent of 40 men and 40 women, with 10 people from each sex assigned to each condition. The conditions were: rural and professional; urban and professional; rural and casual; and urban and casual. To elicit helping behaviors, the 21 year old confederate dropped an envelope within close proximity to each target helper. The time it took for the target to either pick up the envelope or point out that it had been dropped was recorded. Although no differences were found with regard to the type of attire the women was wearing, analyses of the data revealed that response times for helping were much lower in rural settings than urban settings. Furthermore, it was found that men helped the confederate more often than women.

Although the majority of studies have demonstrated that prosocial tendencies are higher in rural settings, only one study was identified that concluded the opposite. Weiner (1976) conducted a two-part laboratory study designed to assay the interrelation between variables as
they related to measures of helping behavior in artificial situations. The sample consisted of 133 American-born subjects (67 females, 66 males; mean age = 19.99) from the University of Hawai‘i’s Manoa and Kauai campuses. Each subject was randomly assigned to one of the conditions intended to prompt prosocial responses. In this study, results indicated that subjects who were raised in a rural setting tended to engage in helping behaviors to a lesser degree than what those from an urban community did. The authors conclude that those from rural settings may not have been afforded the same types of socialization experiences than their urban counterparts, which may explain the results. It would seem from these findings that, contrary to the consensus in the field, the nature of urban settings may actually provide additional opportunities for socialization that could potentially foster the development of prosociality.

The preponderance of evidence with regard to the type of community indicates that rates of prosocial behavior are higher in rural settings when compared to urban environments. Although many researchers have reached this conclusion, few have moved beyond these findings to identify why this may be the case. Future research in this area should attempt to take a closer look at specific variables within each of these settings to provide clarification. Additionally, given the results of Weiner’s (1976) study, prospective research may wish to more closely examine the potential benefits that socialization experiences in urban settings may have on prosocial development. Even though there is a general understanding that the violence, housing issues, and overcrowding within urban settings may inhibit prosocial responses, the identification of additional protective factors may be of great utility when designing interventions with urbanites.
Race

A review of the literature reveals limited empirical evidence related to the impact of race on prosocial behavior. As stated by Wentzel, Filisetti, and Looney (2007), the effects of race on prosocial behavior are poorly understood due to the fact that theoretical models that discuss them are infrequent. These authors note that few researchers have studied the prosocial behavior of African American adolescents and that, for the few studies that have been conducted, the results have been conflicting. For example, Caucasian students in a middle school where the majority (87%) were Caucasian were reported by peers to engage prosocial behavior within the classroom more often than were African American students in a middle school where the majority (92%) were African American (Wentzel, 2002). However, elementary-age Caucasian and African American students’ reports of the frequency with which peers engaged in prosocial behavior produced no significant group differences when the racial majority was considered (Kistner, Metzler, Gatlin, & Risi, 1993).

Given the dearth of race-related studies in the domain of prosocial behavior, as well as the conflicting results of previous studies, Wentzel, Filisetti, and Looney (2007) set out to examine the relationship between race and prosociality. To do so, they utilized a sample of 339 sixth-grade and eighth-grade students from a middle school in a suburban, predominately middle-class community in a mid-Atlantic state. Forty-nine percent were Caucasian, 44% were African American, and the remainder of students were of other ethnic groups. The researchers examined peer- and teacher-reported prosocial behavior in relation to prosocial goals, self-processes, and contextual cues.

Results indicated that Caucasian students were reported to engage in prosocial behavior significantly more than the African American students. However, Caucasian females reported
more self-focused reasons for behavior than what their African American counterparts did. Additional analyses based upon racial classroom composition suggested significant differences were found in the Caucasian-majority classrooms, where Caucasian students were reported to engage in prosocial behaviors more often than African Americans. Yet, reported prosocial behaviors did not significantly differ as a function of race in the classrooms where African Americans were the majority. The authors posit that these findings may be related to the fact that African American students who attend schools where they are the racial minority may not experience the same differentiated crowd structures that Caucasian students do and, in turn, are not affected by peer norms in the same way or to the same degree. Thus, while initial results suggested significant differences based only on race, further exploration revealed that contextual factors may play a significant role in the relationship between race and prosocial behavior. Given the limited nature of research on race and prosocial acting, additional studies are needed to further elucidate the nature of the relationship between these variables.

**Socioeconomic Status (SES)**

Lower SES is generally associated with limited resources, greater exposure to threat, and a decreased sense of personal control over situations (Piff, Kraus, Cote, Cheng, & Keltner, 2010). As such, one may expect that individuals from lower SES would be less inclined to act prosocially and, instead, put self-interests over the welfare of others. Yet, research conducted by Piff et al. (2010) indicates that this is indeed not the case. In this study, the researchers recruited 115 participants from the Research Participation Program for undergraduate students and the Research Subject Volunteer Program at the University of California, Berkely. Participants ranged in age from 18-59 years, with 53 being Asian American, 28 being European American, and the remaining 15 being either African American, Latino/a, or another ethnicity. Nineteen
participants did not report their ethnicity. After the participants completed an online questionnaire regarding demographic information such as age, gender, subjective SES, and religiosity. Approximately one week after completing the questionnaires, the participants were called into the lab, where they participated in several games designed to measure prosocial behavior.

Across four separate studies, the researchers found that those participants from a lower SES proved to be more generous, charitable, trusting, and helpful when compared to their higher SES counterparts. The researchers suggested that those from a lower SES are more inclined to act prosocially due to a greater obligation to egalitarian values and feelings of compassion. Thus, while one may expect these individuals to act in a manner that reflects self-interest, these results show that the adversity associated with low SES actually serves as an impetus for acting prosocially. Unfortunately, and despite the intriguing relationship between SES and prosocial behavior as identified by Piff et al. (2010), there are no other empirical studies examining the impact of SES on prosocial acting. Thus, as with race, additional research is needed in this area.

Overall, evidence presented thus far suggests that, while prosocial behaviors typically develop according to a specific trajectory, various contextual and cultural factors can influence an individual’s proclivity to engage in prosocial acting. Consideration of these external influences ultimately yields a much more comprehensive perspective of prosocial behaviors and reaffirms the potential complex interaction between biology and culture. Although having an adequate understanding of the biological and cultural aspects of prosocial development is important, being able to effectively measure these behaviors is also essential.
Assessment of Prosocial Behaviors

Given the importance of prosocial behaviors, it would seem necessary to have assessments that adequately tap these skills and provide information regarding where a child is functioning in terms of prosocial abilities. However, there are few formal assessments available for assessing the behaviors and other methods are often technically inadequate (Greener, 2000). Assessment of prosocial behaviors within the laboratory setting is often unnatural due to the contrived situations presented to participants. However, although there are concerns related to the ecological validity of assessing prosociality in this manner, obtaining data on these behaviors in natural settings presents the problem of the target individual modifying their behavior as a result of being observed (Eisenberg, 1982). Information derived from self-reports of prosocial behavior may not be entirely accurate due to social desirability, memory of events, or unconscious psychological needs creating misrepresentations (Eisenberg, 1982). Overall, all methods for measuring prosocial behaviors have some limitations that introduce significant implications for validity and reliability.

A search of the literature reveals that there is significant overlap between the types of measures used to assess prosocial behaviors. These methods often include questioning the child about his or her motives for engaging in certain behaviors (Damon, 1977), obtaining data on prosocial responding through the utilization of peer ratings (Tisak & Ford, 1986), or presenting various scenarios to a child for purposes of changing their motivation to act and then determining the circumstances in which a child will behave prosocially (Bar-Tal, Raviv, & Lewis-Levin, 1980). Quantification of prosocial behaviors is often achieved through the use of global assessments.
Global assessments attempt to quantify prosocial behaviors by measuring actions across situations and motivations. Historically, such measures have included self-report, peer ratings, questionnaires, situational tests, and observations. Although data obtained through self-report measures tends to be consistent with peer ratings, scores acquired through self-report are often more reliable than peer ratings and are therefore most often used within the context of research concerned with these behaviors (Greener, 2000). Questionnaires are frequently used as an assessment instrument and ask questions related to how the individual perceives the nature and extent of their prosocial acting. However, a key concern with this type of instrument is that individuals, whether consciously or not, may attempt to convey a higher level of prosociality than what exists in reality (Eisenberg & Mussen, 1989). Furthermore, if the individual is able to surmise the construct being measured through the questionnaire, he or she may be more likely to endorse items consistent with positive behaviors to improve how he or she is viewed by others (Greener, 2000).

As is commonly understood in psychology, the best assessments for a given construct are those that have strong psychometric properties and can consistently and accurately measure the targeted behaviors. In terms of prosocial behaviors, there is a general lack of current, normative instruments designed to assess prosocial behaviors. Some of the instruments were developed in the 1970’s and 1980’s and have not been revised since their inception. Given the ever-changing nature of society, assessments designed almost 40 years ago may not be the best indicators of prosocial behaviors in this day and age. Furthermore, these assessments are difficult to find and are not readily available for use.

A search using the term “prosocial” in the Mental Measurements Yearbook database returned three assessments that produce scores directly related to prosociality. The Infant-
Toddler and Brief Infant-Toddler Social Emotional Assessment (ITSEA/BITSEA; Briggs Gowan & Carter, 2006) is an individually administered assessment designed to identify children who may have social-emotional and behavioral delays or deficits related to social-emotional competence. It can be used for children who are 12-0 months to 35-1 months and includes items intended to assess prosocial peer relations. The Social Competence and Behavior Evaluation (SCBE; LaFreniere & Dumas, 1995) is a rating scale designed to assess patterns of social competence, affective expression, and adjustment difficulties and can be used with children aged 30 to 70 months. Among other scores, the SCBE provides data related to the child’s tendency to behave in ways that are prosocial or egotistical. The Louisville Behavior Checklist (LBC; Miller, 1984) can be used with children aged 4 to 17 years and is completed by the child’s parents. The LBC is designed to facilitate the communication of parental concerns regarding their child and also produces scores directly related to prosociality. Currently, however, the LBC is not available for purchase. In addition to these assessments, there are also several measures that have been developed through the course of research studies.

The Child Behavior Scale (Ladd & Profil et, 1996) was normed on a population of 399 children and includes a subscale specifically designed to measure prosocial behaviors in children ages four through six years old. The Missouri Children’s Behavior Checklist (Sines, 1988) was developed to allow parents and teachers to rate children’s behaviors related to aggression, inhibition, activity level, somatization, sociability, and depression. Although the sample size for the norming was moderate in size, it is noted that the entire sample was Caucasian and that little information is available regarding the psychometric properties of this instrument. The Nonverbal Measure of Children’s Frustration Response (Vondracek, Stein, & Friedrich, 1973) is intended to measure aggression, prosocial behaviors, and avoidance in children who are aged
three though six. However, as with the previous measure, the sample was composed entirely of Caucasian children. Furthermore, internal consistency ranges between 0.37 and 0.80, which raises concerns related to how this measure is assessing constructs. The Questionnaire on Emotional Instability, Prosocial Behavior, and Aggression (Pastorelli, Barbaranelli, Cermak, Rozsa, & Caprara, 1997) can be completed by parents, peers, or the subject of investigation. It is designed for children 7 through 15 years of age and was normed using 1,313 Caucasian children. It is noted, however, that test-retest reliability, inter-rater reliability, and internal consistency was not assessed. One of the strongest measures of prosocial behaviors appears to be the Peer Nomination Instrument (Crick & Grotpeter, 1995). It was normed using 1,166 Caucasian and African American children in several towns in Illinois. Among other behaviors, this instrument directly measures prosociality in children aged 9 through 12 years old. Psychometrically, this assessment reports reliability that is moderate to strong.

What is clear from a review of the available instruments available for assessing prosocial behaviors is that there is a general lack of adequate instruments intended to tap these behaviors. Existing instruments are fraught with concerns related to psychometric properties and were largely normed in specific geographic locations with Caucasian samples. Furthermore, the majority of these instruments are not readily available for use. Although the importance of prosocial behaviors is evident, methods for assessing these behaviors, as a whole, are not capable of yielding valid and reliable information regarding prosocial acting in children.

Assessment of Prosocial Behaviors and Culture

Up to this point, it has been determined that certain aspects of culture can impact the extent to which an individual will engage in helping behaviors. Although some researchers (e.g., Hoffman, 2000) postulate that there is a universal human capacity for prosocial behavior,
Trommsdorff, Friedlmeier, and Mayer (2007) indicate that such a universal ability cannot predict universality of prosocial acting in different cultural contexts because prosociality includes evaluations rooted in moral beliefs and cultural values. Furthermore, there are documented differences in prosocial behavior across cultures (Eisenberg, Fabes, & Spinrad, 2006) and the construct of prosocial behavior is composed of a variety of actions that make generalization difficult (Trommsdorff, Friedlmeier, & Mayer, 2007).

As indicated previously, most cross-cultural work has subscribed to an etic framework, which assumes a culturally neutral perspective and also postulates that the constructs being measured have relevance across all cultures. According to Rubin and Menzer (2010), “The etic framework may cause researchers to operationally define (and thus assess) constructs in the same ways (with the same methods and measures) across cultures” (p. 1). Consequently, this may lead to neglecting culturally specific definitions of constructs when developing assessments. Within the domain of prosocial behaviors, this suggests that the definitions of helping or sharing may differ between cultural contexts (e.g., rural vs. urban, high SES vs. low SES) and may cause inaccurate measurement of these constructs by assessments that assume a universal definition of constructs. Although it is also entirely possible that these constructs are conceptualized similarly between cultures, confirmation would require empirical analysis (Rubin & Menzer, 2010).

Overall, current assessments do not readily lend themselves to determining quantifiable impacts of culture. However, there are certain psychometric properties of assessments that allow us to assume that assessments are measuring prosocial behaviors in a similar manner across cultural contexts.
Test Construction

The term test construction is a broad concept used to refer to psychometric qualities of an assessment that are largely related to reliability and validity. In essence, the focus here is if the assessment is actually measuring what it purports to measure and whether it does so accurately. Although there are many elements of test construction, construct validity, content validity, internal consistency, and standardization/norming sample are those most relevant in determining the extent to which an assessment is capable of measuring constructs in a similar manner across groups and contexts. Additionally, analysis of these elements permits the determination of an assessment’s usefulness in its ability to measure learning. Given the influence of culture on prosocial acting, consideration of test construction elements is essential to measure and inform the usefulness of decision making.

Construct Validity

Within the context of test construction, construct validity answers the questions, “What do scores on this test mean or signify? What does the score tell us about an individual? Does it correspond with some meaningful trait or construct that will help us understand the person?” (Thorndike, 1997, p. 157). According to Borsboom, Mellenbergh, and Van Heerden (2004), construct validity is grounded in the assumptions that (a) the construct refers to an existing phenomenon and (b) the phenomenon produces variations in responses within the real world and within the phenomenon’s measures. Here, the term construct refers to a grouping of variables or behaviors that compose observable patterns of behavior (Drummond, 2004). It is noted that the construct itself is not measurable; rather, the behaviors that have been determined to be reflective of that construct are what is measured. In essence, construct validity is intimately related to how
well the operational definitions and procedures represent the hypothetical constructs that are the target of measurement (Beins, 2004).

Despite the importance of demonstrating that an assessment is measuring what it purports to measure, no simple metric can be utilized for purposes of quantifying the degree to which a measure is construct valid (Westen & Rosenthal, 2003). However, construct validity is often established by calculating correlations between a certain measure of a construct and a number of other measures that should be related to it (convergent validity) or have little association with it (discriminant validity) (Westen & Rosenthal, 2003). Thus, construct validation is accomplished by demonstrating that a construct is related to other variables that it should be and not related to variables that it shouldn’t be. Ultimately, verifying construct validity leads, in part, to the assumption that an assessment is truly measuring what it was designed to measure and that it will do so in a similar manner for all individuals assessed.

For purposes of this paper, construct validity would be concerned with how well assessments are measuring the construct of prosocial behavior. However, as suggested above, each measure of prosocial behavior will likely vary in terms of how prosocial behavior is operationalized. Given the current state of affairs with regard to available instruments for assessing prosocial behaviors, however, demonstrating construct validity may present somewhat of a challenge. As has been discussed previously, many measures of prosocial behaviors are technically inadequate. This is further complicated by the inherent difficulties in effectively quantifying prosocial behaviors within the contrived scenarios of a laboratory. Moreover, the complex and multifaceted nature of prosocial behavior introduces the influence of internal psychological processes that existing assessments do not reflect.
Content Validity

According to Beins, (2004) content validity refers to the extent to which the items within an assessment relate to the concept being measured. With this type of validity, the focus becomes more about individual items on an assessment and less about a large constructs. Within the discussion of prosocial behaviors, it should be expected that items on instruments designed to measure these behaviors are indeed related to prosociality. For example, items related to helping or sharing would most likely be content valid within the domain of prosocial behaviors, while items related to eating habits would likely have little association with prosocial behaviors.

Although there is no one method for establishing content validity, procedures typically involve using a panel of experts to review the assessment, analyzing existing data, calculating the content validity index, and deriving a factorial validity index (Rubio, Berg-Weger, Tebb, Lee, Rauch, 2003). If an assessment has been determined to have adequate content validity, it could be assumed that the items on the assessment will measure the same construct, in the same manner, for each individual who completes the measure.

Internal Consistency

Internal consistency deals with psychometric properties of an assessment that describe the degree to which items on an instrument produce consistent scores in their measurement of a specific construct (Tang & Cui, 2013). With regard to the measurement of prosocial behaviors, internal consistency would be concerned with, for example, the extent to which items designed to measure “sharing” are correlated with each other. For instance, if a respondent conveyed agreement with the statements “I enjoy sharing with others” and “I like to let others borrow my belongings”, but disagreed with a statement such as “I hate sharing”, this may be suggestive of internal consistency. Statistically, internal consistency is most often reported by using
Cronbach’s Alpha (Streiner, 2003). Here, higher correlations between test items will ultimately produce a high value of alpha, which is indicative of internal consistency (Streiner, 2003). Although high correlations are desirable, having correlations between items be too high (> .95) may not be beneficial and may suggest that items are overly redundant (Drummond, 2004).

**Standardization/Norming Sample**

Although the standardization sample is not necessarily related to test construction, the nature of this sample has significant ramifications for being able to show that the instrument is suitable for use on various populations (Beins, 2004). For example, an assessment normed with all middle-class, Caucasian subjects may produce a lack of confidence in the instrument’s ability to assess the construct when the subject is a Hispanic American of a low socio-economic status. Regardless of whether the construct is prosocial behaviors, intelligence, or academic functioning, it is vital that the sample be representative of the population intended to be assessed. Sattler (2001) reports that there are three main factors to consider when evaluating a norm group: representativeness, size, and relevance. He indicates that the norm group should closely match the major demographics of the population as a whole, which can include gender, geographic region, ethnicity, and socio-economic status. Regarding size, Sattler suggests that the number of subjects in the norm group should be of a size that produces stable test scores and that ensures adequate representation of the subgroups in the population. Regarding relevance, Sattler posits that, in order to effectively interpret an individual’s score, he or she must be compared to the most relevant norm group. Generally speaking, having a norm group that meets these criteria leads to more confidence in the instrument’s ability to effectively measure constructs across a wide variety of people and contexts. Ultimately, having reliable and valid methods to assess
prosocial behaviors leads to a greater ability to identify deficits and plan interventions designed to ameliorate weaknesses related to prosocial functioning.

**Be a Safety Kid**

Although there are several intervention programs designed to address prosocial behaviors within the school setting, this paper is specifically concerned with the *Be a Safety Kid* curriculum. This program is a school-based violence intervention program intended to foster an educational setting where students feel protected from the threat of violence (Safety Kids, 2005). Overall, *Be a Safety Kid* is built around the concept of “responsible reporting”, which entails communicating information that may prevent injury or a crime (Safety Kids, 2005). Through the “Ima S.T.A.R.” unit, the program attempts to break the mindset of “snitching” and encourage students to make good decisions when they see things within their environment that are potential dangers. This unit is designed for students enrolled in kindergarten through eighth grade and intended to change attitudes, teach responsibility, and promote accountability (Safety Kids, 2005).

For purposes of determining effectiveness, *Be a Safety Kid* incorporates the S.T.A.R. instrument, which is designed to assess knowledge of prosocial behaviors and responsible reporting, performance of prosocial behaviors and responsible reporting, and the feeling of school connectedness (Safety Kids, 2005). This instrument is administered as a pre- and post-test measure to assess the extent to which the curriculum may have impacted a student’s knowledge of prosocial behaviors and responsible reporting. Additionally, the instrument assesses changes related to the students’ propensity to engage in prosocial acting and the student’s perception of how safe the school environment is. Due to the differences in developmental processes across age, two different test versions were created to ensure that
questions on the assessment were consistent with the child’s developmental level. For kindergarten through third grade, the items are largely subjective, skill-based questions, whereas the assessment for fourth through eighth grade includes questions dealing with self-reflection, performance, and school connectedness.

**Culture, Test Construction, and the S.T.A.R. Instrument**

As previously indicated, the S.T.A.R. instrument was developed for purposes of measuring three specific constructs related to prosocial behaviors, which are knowledge of prosocial behaviors and responsible reporting, performance of prosocial behaviors and responsible reporting, and the feeling of school connectedness. The S.T.A.R. instrument serves as the sole indicator of the extent to which students have benefited from the implementation of the *Be a Safety Kid* curriculum. Results obtained from this instrument are utilized for decision making and ultimately allow a determination to be made with regard to the magnitude of learning that occurred. With this type of curriculum project, it is not only important that the curriculum be effective, but it is also essential that the assessments being used to measure learning are efficacious.

Prior studies (Martin, 2011; Graham, 2012) have examined the effectiveness of the *Be a Safety Kid* curriculum in five states (Arizona, Florida, Georgia, Pennsylvania, and Wisconsin). Graham (2012) found that, with a population of kindergarten through third grade students, the *Be a Safety Kid Curriculum* produced significant improvement in prosocial knowledge and ability to demonstrate prosocial behaviors. No significant improvement in the feelings of school connectedness and safety were noted. Conversely, Martin (2011) found that, with a population of seventh and eighth grade students, the *Be a Safety Kid* curriculum did not produce significant improvement in prosocial knowledge and the ability to demonstrate prosocial behaviors. In
addition to effectiveness, Martin (2011) and Graham (2012) also analyzed the validity of the S.T.A.R. instrument with specific populations.

With regard to validity of the S.T.A.R. instrument, results of factor analyses conducted by Martin (2011) suggested that the S.T.A.R. instrument aligned with the major constructs of knowledge and school connectedness in a sample of 7th and 8th grade students. However, only two questions on the S.T.A.R. instrument were associated with the construct of performance. Additionally, three items on this instrument were found to not align with any factors, which suggests that either these items do not measure any type of prosocial behavior or that they measure all three constructs equally. Overall, this author concluded that the S.T.A.R instrument was a valid measure of constructs for a majority of the questions created through theoretical and empirical analyses.

Graham (2012) conducted similar statistical analyses with a population of kindergarten through 3rd grade students. The results indicated that the instrument did not consistently align with the constructs of knowledge, performance, and school connectedness through face validity. Furthermore, analyses suggested a lack of differentiated constructs on the S.T.A.R. instrument, which may have been produced by low variance values in the instrument. Despite this, analyses suggested that the instrument was shown to be a stable measure of constructs overtime when pre- and post-tests were considered.

While previous studies have found mixed results regarding the effectiveness of the Be a Safety Kid Curriculum and the validity of the S.T.A.R. instrument, no studies have sought to determine whether the S.T.A.R. instrument is measuring these three constructs similarly across these five states. Furthermore, no studies have examined the potential impact of SES, race, or community type (rural vs. urban) on the manner in which the S.T.A.R. instrument measures
these constructs. Since cultural factors have been shown to impact an individual’s propensity to engage in prosocial behaviors, it is possible that cultural differences between the schools in the five aforementioned states could also affect how the S.T.A.R. instrument is measuring behaviors related to these constructs. However, there is currently no direct empirical evidence to either prove or disprove that the behaviors associated with the constructs measured by the S.T.A.R. instrument are equally relevant across the states and cultures in which the program was implemented. In the absence of empirical confirmation, an exploratory study is needed to conclude, via the utilization of statistical analyses, the extent to which the S.T.A.R. instrument assesses these constructs in a similar manner across the five states and across cultural contexts.

Although it has been determined that cultural factors exert an influence on prosociality, the previous discussion of test construction would suggest that, with strong psychometric properties, an assessment should be able to measure constructs similarly regardless of cultural elements. Specifically related to the S.T.A.R. instrument, however, there is simply no existing evidence to indicate that this is, or is not, the case. This exploratory study will facilitate a better understanding of the S.T.A.R. instrument and also elucidate the degree to which this instrument overlooks any culturally-specific definitions of the three constructs. To do so, however, requires statistical analyses that move beyond test construction.

In conducting this study, conclusions can also be drawn in terms of the relative utility of the S.T.A.R. assessment. Currently, the S.T.A.R. instrument can be likened to a curriculum-based measure designed to assess the effects of this curriculum at a local level. However, if the results of this study suggest that the S.T.A.R. instrument measures the constructs similarly across five states, this may speak to a broader utility of this instrument and potentially elevate it to a level beyond just a curriculum-based measure. For example, the Dynamic Indicators of Basic
Early Literacy Skills (DIBELS; University of Oregon Center on Teaching and Learning, 2014), which is now a widely used assessment of early literacy skills, was first purposed as a curriculum-based assessment to be used at a local level. However, through an ongoing series of studies and analysis of data from the DIBELS, comparisons can now be made between students who are in different locations. In essence, the research has indicated that the DIBELS is measuring constructs (e.g., nonsense word fluency, oral reading fluency, etc.) in the same way across contexts. Presently, such comparisons cannot be made with the S.T.A.R. assessment given the lack of data confirming that this is possible. Thus, if the results of this study suggest that the S.T.A.R. instrument measures constructs similarly across the five states, a broader utility of this instrument may be realized. Again, however, statistical analyses are needed to either support or refute the S.T.A.R. instrument’s ability to measure constructs similarly.

**Research Questions**

**Research Question 1**

Does the S.T.A.R. instrument measure students’ knowledge of prosocial behaviors, students’ anticipated performance of prosocial behavior, and students’ perception of school safety similarly across five states?

**Research Question 2**

Does the S.T.A.R. instrument measure students’ knowledge of prosocial behaviors, students’ anticipated performance of prosocial behavior, and students’ perception of school safety similarly across race?
Research Question 3

Does the S.T.A.R. instrument measure students’ knowledge of prosocial behaviors, students’ anticipated performance of prosocial behavior, and students’ perception of school safety similarly between low SES and high SES students?

Research Question 4

Does the S.T.A.R. instrument measure students’ knowledge of prosocial behaviors, students’ anticipated performance of prosocial behavior, and students’ perception of school safety similarly between students from urban and rural environments?

Summary

A review of the existing literature regarding prosocial behaviors reveals that these behaviors have significant implications for a child’s social competence and the extent to which a child is accepted by his or her peer group (Denham & Holt, 1993). However, existing research focusing on prosocial behaviors presents with significant methodological limitations concerning sampling and measurement techniques. Furthermore, few studies have attempted to ascertain the complex underlying psychological processes that may impact one’s proclivity to act prosocially. These problems are further complicated by the broad nature of prosocial behaviors and the fact that there is no consistency in defining the term prosocial behaviors.

Regarding measurement of prosocial behaviors, typical assessments include naturalistic observation, laboratory studies, and global assessments. Although laboratory settings minimize confounding variables, the contrived nature of these environments makes it difficult to conclude that the behaviors observed in the laboratory will resemble those that occur in a natural setting. While global assessments allow these behaviors to be quantified, there is a lack of valid and
reliable instruments at the present time. Naturalistic observations provide real world insight about prosocial behaviors, but provide little quantifiable data.

Although psychological theories suggest the same developmental trajectory for prosocial behaviors, there is evidence to suggest that cultural elements can influence the extent to which an individual will engage in these behaviors. Specifically, community type, political affiliation, family values, SES, race, and religion have all been found to impact prosociality. At face value, this may present difficulties regarding the ability to adequately assess prosocial behaviors. However, assessments that maintain strong psychometric properties and have standardized with a sample representative of the population should be able to overcome the influences of culture and demonstrate the ability to measure prosocial constructs similarly across individuals and contexts. Ultimately, this aids in intervention planning.

The Be a Safety Kid curriculum has been developed as a violence prevention program, with focus on increasing prosocial acting in the form of responsible reporting. By using the S.T.A.R. instrument as a pre- and post-test measure, data can be obtained with regard to changes in knowledge of prosocial behaviors and responsible reporting, performance or prosocial behaviors and responsible reporting, and the feeling of school connectedness. The focus of this paper is to determine to extent to which the S.T.A.R instrument is capable of measuring these constructs in the same manner among a sample of kindergarten through eighth grade students given the potential influence of cultural elements. Doing so will allow a determination to be made with regard to the extent to which the S.T.A.R. instrument overlooks any culturally specific definitions of the constructs. Additionally, this study will allow conclusions to be made with regard to whether or not the S.T.A.R. instrument has a greater utility beyond a local curriculum-based measure.
CHAPTER III

METHODS

This chapter will describe the specific procedures utilized to investigate the research questions presented in chapter two. The purpose of this study was to determine the extent to which the S.T.A.R. instrument measures the constructs of knowledge of prosocial behaviors and responsible reporting, performance of prosocial behaviors and responsible reporting, and the feeling of school connectedness similarly across the states of Georgia, Pennsylvania, Wisconsin, Florida, and Arizona. Additionally, this study examined the extent to which the STAR instrument measured these constructs similarly across race, SES, and community type (i.e., rural vs. urban). The participants for the present study will first be discussed, with details regarding the manner in which they were recruited. Next, the measures that operationalize each construct related to the research questions are outlined. The procedures used for administering the measures and collecting data are then discussed, with specific information related to the technical qualities of the instruments used to collect the data. Lastly, the steps of data analyses for the present study will be described.

Participants

Due to the fact that this study consisted of a secondary analysis of a pre-existing database, the actions taken by the owner of the Be a Safety Kid curriculum to secure participants will be explained. Schools across the continental United States received requests to receive and implement the Be a Safety Kid curriculum according to the guidelines set forth by the developer of the curriculum. Additionally, attempted recruitment occurred during national conferences and promotional events when school personnel would inquire about the curriculum. In all, 92 requests were sent to schools within the continental United States. If a school’s personnel
determined that they did not wish to participate, they were asked to complete a questionnaire designed to collect information regarding the reasons for not wanting to participate. When school administrators agreed to participate, they were informed that students would be administered a pre-test and post-test, with the pre-test occurring at the beginning of the year before the Be a Safety Kid curriculum was implemented and the post-test occurring at the end of the year once the curriculum delivery was completed. Prior to the implementation, schools were able to decide if they wanted the curriculum to be implemented throughout the entire school, only in certain grades, or only in certain classrooms.

**Participant Characteristics**

Significant effort was expended in trying to obtain a national sample of students in kindergarten through 8th grade who were from rural, urban, and suburban school districts. Out of the 92 requests sent, 10 schools from 7 different school districts across the United States committed to implementing the curriculum and the administration of the pre- and post-test measures. None of the schools who chose to not participate returned information regarding their rationale for not participating. The sample for the current study includes 8 schools from 6 of the 7 school districts who agreed to participate in the following states: Wisconsin, Georgia, Florida, Pennsylvania, and Arizona. This study will focus on data collected from 1st and 2nd grade students who completed the Be a Safety Kid curriculum. There are approximately 349 first grade students and 326 second grade students. The data utilized for this study was collected between 2008 and 2010. A description of each school’s demographic data is present in Table 1, which was obtained from the National Center for Education Statistics (NCES, 2014).
<table>
<thead>
<tr>
<th>School</th>
<th>Type/Locale</th>
<th>Title One</th>
<th># Students</th>
<th># Free/Red Lunch</th>
<th>Racial Enrollment</th>
</tr>
</thead>
<tbody>
<tr>
<td>PA #1</td>
<td>Public/Urban</td>
<td>Yes</td>
<td>248</td>
<td>156</td>
<td>Asian – 1; Black – 136; White – 94</td>
</tr>
<tr>
<td>PA #2</td>
<td>Public/Urban</td>
<td>Yes</td>
<td>346</td>
<td>216</td>
<td>Asian – 4; Black – 152; White – 157;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>American Indian – 3; Hispanic – 4</td>
</tr>
<tr>
<td>PA #3</td>
<td>Public/Urban</td>
<td>Yes</td>
<td>458</td>
<td>183</td>
<td>Asian – 7; Black – 294; White – 148;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>American Indian – 1</td>
</tr>
<tr>
<td>PA #4</td>
<td>Public/Rural</td>
<td>Yes</td>
<td>391</td>
<td>260</td>
<td>Asian – 2; Black – 4; White – 350;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>American Indian – 2; Hispanic – 32</td>
</tr>
<tr>
<td>GA</td>
<td>Public/Urban</td>
<td>No</td>
<td>1,080</td>
<td>85</td>
<td>Asian – 228; Black – 54; White – 714;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>American Indian – 3; Hispanic – 52</td>
</tr>
<tr>
<td>AZ</td>
<td>Public/Urban</td>
<td>Yes</td>
<td>787</td>
<td>771</td>
<td>Black – 11; White – 19; American Indian</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>– 2; Hispanic – 755</td>
</tr>
<tr>
<td>FL</td>
<td>Public/Urban</td>
<td>Yes</td>
<td>441</td>
<td>423</td>
<td>Black – 368; White – 1; Hispanic – 70</td>
</tr>
<tr>
<td>WI</td>
<td>Private/Rural</td>
<td>No</td>
<td>209</td>
<td>12</td>
<td>Asian – 14; Black – 3; White – 184;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Hispanic - 8</td>
</tr>
</tbody>
</table>
Intervention

The curriculum that is the focus of this study is the Be a Safety Kid curriculum. The overarching aim of this program is to foster a sense of safety and security within the school setting, where students are not concerned about the threat of school violence. Through this curriculum, students are taught requisite skills and behaviors for preventing harm and violence. Additionally, students learn how to develop attitudes that are characteristic of prevention and prosocial styles of interaction (Safety Kids, 2005). This school-wide program is predicated upon the concept of “responsible reporting”, which entails appropriate communication of information when a potentially dangerous situation is evident. By conveying this information, a “responsible reporter” attempts to avert harm to others caused by a hazardous or unsafe situation. If an incident has already occurred, the “responsible reporter” tries to prevent harm to additional individuals by informing school personnel of the initial situation. This program is also designed to convey the fact that “responsible reporters” should not be perceived in a negative manner due to the fact that individuals who cause harm to others should be held responsible for their behaviors (Safety Kids, 2005).

The Be a Safety Kid curriculum incorporates activities, concepts, and objectives that align with age-appropriate development of skills and knowledge of prosocial behaviors for students in kindergarten through eighth grade. The curriculum includes resources designed to offer additional reinforcement via materials that promote the use of skills during the entire school day and facilitate the application of behaviors across various settings within the school. Students are also encouraged to utilize their thinking skills and senses for purposes of acting responsibly under the pretense that all students should be responsible, which, in turn, produces acceptance for the idea that peers must be held accountable for their actions (Safety Kids, 2005). This
curriculum includes additional instructional reinforcement materials, such as worksheets, role-play activities, and hypothetical scenarios, which are utilized upon the completion of each lesson. These materials are intended to provide real-world examples of behaviors and also meant to test skill knowledge related to concepts previously taught.

Instrumentation

Creation of the S.T.A.R. Instrument

In an attempt to create a technically adequate and comprehensive assessment instrument for exploring the fidelity of the Be a Safety Kid curriculum, variables were evaluated for their impact on children’s skill development. This process produced questions on the S.T.A.R. instrument that related to three specific areas, which were knowledge, performance, and school connectedness. Knowledge questions on the instrument were intended to evaluate the acquisition of information in relation to pre-determined lesson objectives for each grade level. Questions related to performance were designed to measure the propensity for engaging in prosocial behaviors, while the school connectedness questions were aimed at measuring the perception of safety within the confines of the school setting. Due to the large variation in development between younger students and older students, two distinct versions of the S.T.A.R. instrument were created for purposes of facilitating the measurement of similar skills in a manner consistent with the developmental levels of the children.

When determining the most appropriate ways to assess children, it is important to ensure that assessment techniques are consistent with how children learn at certain ages. Younger children, for example, tend to learn best through concrete, hands-on, interactive, and experiential methods as opposed to activities that emphasize abstract reasoning and paper-pencil tasks (Bredekamp & Rosegrant, 1995). Such methods for younger children can include things such as
object manipulation, building/creating, movement, and talking/singing. In terms of conveying information, younger children are more capable of reporting information that is subjective in nature rather than objective (Stone & Lemanek, 1990). As such, the developers of the Be a Safety Kid curriculum decided that the pre- and post-test questions for students in kindergarten through 3rd grade would be limited to subjective, skill-based questions. The pre- and post-test questions for students in 4th through 8th grade included questions related to performance, self-reflection, and school connectedness, which are better suited to children in this age group as opposed to younger children.

Due to the difference in abilities between 3rd and 4th grade students in the school environment, a younger version of the S.T.A.R. instrument was developed for use specifically with the kindergarten through 3rd grade population (Safety Kids, 2005). This version was aligned with the Be a Safety Kid curriculum and focused on acquisition of knowledge and utilization of these skills within the school setting. Since reading abilities are still emerging in younger children, some authors (Reynolds & Richmond, 1978; Stone & Lemanek, 1990) have posited that elementary students should have test questions read to them. This may lead to greater confidence that the test is measuring what it was deigned to assess and that the child’s reading abilities have not impacted test performance. As such, the pre- and post-tests in the Be a Safety Kid curriculum are administered orally to those students in kindergarten to 3rd grade. On the older version of the S.T.A.R. instrument for children in 4th through 8th grade, the emphasis is not only on skill acquisition and performance, but also on the utilization of these skills on a consistent basis. Given that older students often present with better developed reading abilities, the pre- and post-test for these students is in written form. Both the younger and older versions of the S.T.A.R. instrument also include questions intended to assess students’ perception of
school safety within the school environment and the students’ capability to form meaningful relationships with the educational structure, including staff in the school.

A frequently used method in psychological and educational measurement is the Likert scale. Assessments that utilize Likert scale questions often have instructions that are easily understood by students, which leads to minimal instruction time for the individual proctoring the assessment (Guyatt, Townsend, Berman, & Keller, 1987; Jaeschke, Singer, & Guyatt, 1999; Vickers, 1999). Younger students are better able to understand and accurately answer Likert scale questions in comparison to other question formats, such as the continuous rating scale (Shields, Cohen, Harbeck-Weber, Powers, & Smith, 2003). From a developmental perspective, the fact that a child’s language, writing, and reading skills are not fully developed often produces limitations in terms of how well children are able to comprehend and respond to self-report inventories. Thus, for purposes of measuring skills and the understanding of the Be a Safety Kid curriculum, it was determined that Likert scale questions would be the most appropriate.

Although the Likert scale has many inherent benefits and positive attributes, there is some evidence that this type of scale can also be unclear to those completing the questions. Specifically, an excessive number of response categories may make it challenging for the respondent to choose between the responses, while an insufficient number of categories may not afford the respondent adequate choices (McCormack, Horne, Sheather, 1988; Vickers, 1999). The Likert scale may cause respondents to select a rating that is slightly more positive or negative than neutral since the construction process of the Likert scale has a tendency to reduce the selection of neutral responses and increase the selection of choice that are more extreme (Vickers, 1999). As such, questions contained within the pre- and post-test questions of the Be a Safety Kid curriculum use a gradient scale that ranges from Always, Often, Sometimes, and
Never, which eliminates any neutral options. Furthermore, the Likert scale for the younger children included only two options, which were Yes or No. The basis for this decision was that younger children may struggle to distinguish between four options on the Likert scale. For example, they may have difficulty differentiating between Always and Often and between Sometimes and Never. Presenting the younger students with only two options provides a clearer delineation between the two choices, which will yield a better indication of the extent to which the students have acquired and retained the necessary skills.

Prior to administering Likert scale test questions to students, it is necessary to determine if the questions are age-appropriate in terms of vocabulary and reading levels (Stone & Lemanek, 1990). With the Be a Safety Kid curriculum, both versions of the S.T.A.R instrument were analyzed for readability using a formula created during the development of the instrument. Specifically, OKAPI was utilized to determine readability, which is an Internet application used to develop reading probes for purposes of curriculum-based assessment. As a web-based application, OKAPI permits users to insert and format text samples that will ultimately serve as a set of Examiner and Student Curriculum-Based Assessment reading probes. With regard to the S.T.A.R. instrument pre- and post-tests, the Spache or Dale-Chall Readability formulas were calculated for each scenario on the instrument. The Spache Readability Formula is generally used to determine the text difficulty for 3rd grade and below (Spache, 1953), while the Dale-Chall Readability Formula is typically used for 4th grade and above (Dale & Chall, 1948). Thus, the Spache formula was used for the kindergarten through 3rd grade version of the S.T.A.R. instrument and the Dale-Chall formula was used for the 4th through 8th grade version. Readability for each version was determined to be age-appropriate for the population with which it would be used.
The developers of the *Be a Safety Kid* curriculum were cognizant of the fact that younger children may not be able to maintain interest on a test for extended periods of time. However, Harter and Pike (1984) found that the interest of younger children can be improved by incorporating a pictorial format (e.g., cartoon drawings) in the assessment. They further posit that this type of format helps to clarify the responses and make the verbal material more concrete. This led the developers of the *Be a Safety Kid* curriculum to use both pictorial and verbal representations for skills questions on the younger version of the S.T.A.R. instrument. For example, on the kindergarten through 3rd grade pre- and post-tests of the S.T.A.R. instrument, the YES or NO response options were paired with a “thumb up” or “thumb down” picture.

In developing questions related to skills acquisition on the S.T.A.R. instrument, the developer incorporated vocabulary and scenarios directly from the curriculum so that the students would be familiar with the test content and vocabulary. The questions were taken verbatim from the curriculum and were related to the lessons taught at each grade level, which produced a slightly different instrument for each grade. The questions were designed to be concrete in nature and avoid any opinion based inquiries. According to the developer, the knowledge-based questions were created by an elementary education teacher and were evaluated by a school psychologist and special education specialists to ensure face validity.

With the performance related questions, two areas were examined to determine their role in a child’s tendency to engage in prosocial behaviors on a regular basis. One area dealt with the ability to demonstrate helping behaviors even when the student was not directly involved in situations that were violent or potentially hazardous. This type of helping behavior is generally described in the literature as bystander or prosocial behavior. In specific relation to school
violence, bystanders may be those students who witness acts of physical aggression, such as fights. It is noted, however, that these situations are not confined only to acts of physical aggression, but also include instances in which the bystander may have information suggesting that future violence is likely to occur (Stueve, Dash, O’Donnell, Tehranifar, Wilson-Simmons, Slaby, & Link 2006). Bystanders are generally not passive observers due to the fact that their action or inaction can influence whether and how volatile situations evolve. According to Twemlow, Fonagy, and Sacco (2004), bystanders are intrinsically motivated by being helpful and do not seek attention for their actions. Given the significance of bystander behaviors, the developers of the curriculum determined that it was vital to include questions related to the extent to which students felt comfortable engaging in prosocial behaviors intended to prevent violence in the educational setting.

Several quintessential features of the bystander role are identified in the literature. Darley and Latane (1968) proposed several important steps in the process of deciding to be a prosocial bystander, which includes noticing what is occurring, labeling it as a problem where help is required, taking responsibility, determining what actions to take, and feeling confident that one has the skills to take action and can do so in a safe manner. Ajzen (2002) proposed another paradigm that emphasizes how people consider the benefits and costs of various courses of action, how they evaluate the normative expectation of others, and how they gauge their ability to act. These models facilitate the delineation of specific areas of inquiry that are germane to the assessment of a student’s propensity to be a responsible reporter when he or she is involved in a violent or potentially violent situation. These principles were also a vital aspect in the Be a Safety Kid curriculum instruction aligning with the core concept of “responsible reporting.” To that regard, the questions on the S.T.A.R. instrument were designed to measure
the likelihood of a student appropriately reporting unsafe situations and how comfortable or fearful they would be in doing so.

In an effort to integrate research models, the creators of the curriculum decided to ask students questions related to why they were unwilling to report potentially dangerous situations. These reasons were derived from research that explained the possible contextual factors that may preclude prosocial acting. For example, in situations that are less serious and more ambiguous, bystanders are often slower to notice warning signs and are therefore less likely to intercede (Latane & Nida, 1981; Shotland & Goodstein, 1984). Additionally, the presence of several bystanders can cause bystanders to misperceive or miscalculate the severity of a situation, and bystander involvement may be hindered by the degree of intimacy or relational distance between the victim and aggressor (Stueve et al., 2006). Research has also demonstrated that, when compared to strangers, socially cohesive groups of bystanders are more likely to react to emergency situations, which further supports the need for a normative environment that promotes social responsibility (Horowitz, 1971; Latane & Nida, 1981; Rutkowski, Gruder, & Romer, 1983). Ultimately, the abovementioned reasons were utilized as a guide for purposes of identifying research-based answer options related to why students would elect to not make prosocial decisions.

Also essential to the performance of prosocial behaviors is the function of the broader social context factors in the development of prosocial behaviors across the lifespan (Carlo & Randall, 2001). For example, the feeling of bonding and connectedness in the school setting are social context factors that should be considered since they can potentially influence prosocial acting. On the S.T.A.R. instrument, questions related to school connectedness were adapted from various measures used in prior research studies. The Sense of Community Scale, which
was created by Unger and Wandersman (1982) has been previously utilized in research with college students and is a short, three-item measure containing the following questions: “Do you feel a sense of community with other people on campus?”; “How important is it to you to feel a sense of community with people on this campus?”; and “Some people care a lot about the kind of campus they live on. For others, the campus is not important. How important is what the campus is like to you?” On the S.T.A.R. instrument, these questions were adjusted to better correlate to the school setting by using school and school personnel, such as administrators and teachers, as the main focus of the questions. School connectedness questions on the S.T.A.R. instrument were also derived from the Prevention Scale, which is a 13-item measure created for use in the program evaluation of the Mentors in Violence Prevention Program (MVP; Katz, 1995). This scale was intended to evaluate an individual’s self-efficacy with regard to the prevention of gender violence; however, the questions are more oriented to a school violence perspective with a focus on the extent to which students believed they had control over violence in the school setting. These questions were also modified to better associate with the K-8 school environment.

For purposes of evaluating the integrity and utility of the Be a Safety Kid curriculum in the school, a teacher questionnaire was also developed as part of the program. Questions concerning school connectedness were adapted from the student questionnaire and additional questions were included to obtain information related to the pros and cons of the curriculum and the general ease of implementation. Additionally, various assessment techniques from a teacher instrument were integrated into the curriculum in order to evaluate the efficacy of a bullying prevention program (Edmondson & Hoover, 2008). The Be a Safety Kid creators indicate that information related to the perception of student behavior, reported implementation of curricular
lessons, and resulting changes to the atmosphere of the school were elicited subsequent to the curriculum’s implementation.

**Measures**

Students who participated in the *Be a Safety Kid* curriculum received early violence intervention for the entire school year. Teachers and students completed the self-report survey both before and after the curriculum’s implementation. For purposes of this study, the S.T.A.R. instrument serves as the primary measure from which data was derived. In order to align with the theoretical constructs detailed in the creation of the S.T.A.R. instrument, test items in the kindergarten, 1st, 2nd, and 3rd grade instruments were separated into questions related to the following three areas: knowledge of prosocial behaviors, performance of prosocial behaviors, and feeling of school connectedness. The instrument has a total of 10 questions, with questions 1 through 8 measuring knowledge and performance of prosocial behaviors. On the kindergarten, 1st, and 2nd grade instruments, the first question assesses which sense (e.g., hearing or seeing) the students feel they are using to analyze the situation. Questions 2 through 8 are largely knowledge questions derived from the *Be a Safety Kid* curriculum, with one (kindergarten) or two (1st, 2nd, and 3rd grades) questions measuring performance. According to the curriculum developer, the knowledge questions were intended to assess whether students effectively learned and acquired basic information through the curriculum, while performance questions attempted to ascertain the likelihood of demonstrating these behaviors and the reasoning for becoming actively involved in a potentially dangerous situation. On the 3rd grade instrument, three of the first eight questions were designed to assess performance of prosocial behaviors (questions 1, 3, and 4), with the other five questions assessing knowledge. Questions 9 and 10 on the kindergarten, 1st, 2nd, and 3rd grade S.T.A.R. instrument were designed to measure the extent to
which students felt safe and connected in the school setting. It is noted that, although the STAR instrument is very similar for those students in grades K-3, the instrument for 1st and 2nd grade is exactly the same, while the kindergarten and 3rd grade instruments are slightly different in terms of question wording and content. Thus, given the research questions of this study and the comparisons being made, it was necessary to only use data from the 1st and 2nd grade STAR instrument.

The S.T.A.R. instrument for 4th through 8th grade consists of 20 questions, with questions 1 through 5 measuring knowledge of prosocial behaviors. As with the kindergarten through 3rd grade S.T.A.R. instruments, knowledge questions were taken directly from the instruction delivered through the Be a Safety Kid curriculum. Questions 6 through 12 were designed to measure the student’s ability to use their knowledge to perform prosocial behaviors, while questions 13 through 20 assess the student’s perception of safety and connectedness in the school environment.

Information regarding psychometric properties of the S.T.A.R. instrument is limited to two exploratory dissertation studies conducted by Martin (2011) and Graham (2012). With regard to validity of the S.T.A.R. instrument, results of factor analyses conducted by Martin (2011) suggested that the S.T.A.R. instrument aligned with the major constructs of knowledge and school connectedness in a sample of 7th and 8th grade students. However, only two questions on the S.T.A.R. instrument were associated with the construct of performance. Additionally, three items on this instrument were found to not align with any factors, which suggests that either these items do not measure any type of prosocial behavior or that they measure all three constructs equally. Overall, this author concluded that the S.T.A.R instrument was a valid
measure of constructs for a majority of the questions created through theoretical and empirical analyses.

Graham (2012) conducted similar statistical analyses with a population of kindergarten through 3rd grade students. The results indicated that the instrument did not consistently align with the constructs of knowledge, performance, and school connectedness through face validity. Furthermore, analyses suggested a lack of differentiated constructs on the S.T.A.R. instrument, which may have been produced by low variance values in the instrument. Despite this, analyses suggested that the instrument was shown to be a stable measure of constructs overtime when pre- and post-tests were considered.

Both Martin (2011) and Graham (2012) report that face validity was established by asking experts in the fields of school psychology, intervention implementation, and violence prevention for their opinions of the S.T.A.R. instrument during the development. The creators of the curriculum, as well as multiple school psychologists, police office, statistics professors, teachers, children, and principals, provided input and corrections regarding the details of the instrument and its alignment with theoretical constructs.

**Research Design**

This study utilized a pre-test post-test quasi-experimental design, which consisted of nonrandomized groups. Such designs are frequently used in behavioral research in order to compare groups and/or measure change resulting from experimental treatments (Dimitrov & Rumrill, 2003). With a quasi-experimental approach, the researcher has little or no control over the provision of treatments or other factors being examined, and participants are not randomly assigned to groups.
Procedures

After the *Be a Safety Kid* curriculum was approved by the school districts, the curriculum was provided to each district by the developer of the program. The curriculum was subsequently implemented by the school districts as a general education practice and each school district had discretion over the implementation of any portion of the curriculum. Participation was voluntary and the districts could, at any time, withdraw participation by simply not completing the forms. Every school district also had the option of returning demographic information and de-identified teacher, parent, and child data forms. For those that did return information, consent to compare de-identified data from the school district was assumed. Introductory material provided by the developer of the curriculum outlined the purposes of comparisons in the event that districts decided to provide their information to the creator. All data provided to the developer was in aggregate form in order to preclude the ability to identify any parent, child, teacher, or administrator. Thus, no names were included.

Students participated in sessions once per week, which were integrated into the regular curriculum throughout the school year. The study was not perceived to have caused physical, social, legal, economic, or psychological harm to any participants. Since the curriculum is considered typical educational practice, it was considered to have no more than a minimal risk to students. As with any instruction related to prosocial behaviors, there is a possibility that some students may experience a sense of discomfort as a result of discussing controversial or intrusive personal information. In order to address any problems, monitoring of the instruction occurred and, if needed, supports were offered by the district’s curriculum leader. If it was determined that action was warranted, the developer of the curriculum worked with the school district to establish appropriate support at the local level. Additionally, each curriculum administrator was
provided with necessary training. Overall, the benefits of this study were determined to outweigh any potential risks. Approval for this study was granted by Duquesne University’s Institutional Review Board (IRB), as well as by administrators and school boards from each school.

Each participant completed the pre-test during the first week of the school year, while the post-test was completed during the last week of the year. Teachers completed the teacher form and assigned each student an anonymous identification number for purposes of organizing the pre- and post-test measures. Instruction with the curriculum took place once per week during the school day for one hour. Upon completion of the curriculum and administration of the post-test S.T.A.R. instrument, the completed measures were then returned to the Be a Safety Kid developer by the district’s curriculum leader.

**Data Analysis**

The owner of the Be a Safety Kid curriculum collected all data and only de-identified data was provided to the primary researcher. Participants were assigned an identification number for purposes of protecting their privacy and facilitating the ability to match pre- and post-test scores for each student. Statistical analyses for this study entailed first aligning questions from the S.T.A.R. instrument by content area (i.e., knowledge, performance, and school connectedness). Five separate confirmatory factor analyses (one for each state) were then performed using post-test scores on the instrument for purposes of determining if the underlying factor structure of the S.T.A.R. instrument varies between states. To further examine cultural and contextual factors, the schools were then categorized according to race (i.e., majority White, majority Black, and majority Hispanic), SES (i.e., high SES and low SES), and community type (i.e., rural and urban) and separate confirmatory factor analyses were conducted with each grouping to
determine the underlying factor structure. Comparisons regarding the factor structure were then made.

In terms of categorizing schools into State groups, schools were simply assigned to the group that represented the state in which they were located. Regarding race, schools were assigned to either a Caucasian group, African American group, or Hispanic group depending on the race that represented the majority of the student population within the school (i.e., over 50%). In terms of community type, schools were assigned to either a Rural or Urban group depending on the locale designated by NCES (2014). For example, schools classified as being situated in cities or suburbs were assigned to the Urban group, while schools in rural areas and towns were assigned to the Rural group. This follows the methodology set forth by NCES (2014) in classifying schools as either rural or urban. Lastly, schools were categorized as either High-SES or Low-SES based upon the percentage of students who receive a free or reduced lunch. According to the United States Department of Agriculture (USDA, 2002), schools that serve at least 40% or more of school lunches to children who qualify for free or reduced price meals are classified as “severe need.” As such, the 40% cutoff was utilized for purposes of classifying schools into SES groups by dividing the total number of students in the school by the total number of students eligible for free or reduced lunches.

Research Questions

Research Question 1

Does the S.T.A.R. instrument measure students’ knowledge of prosocial behaviors, students’ anticipated performance of prosocial behavior, and students’ perception of school safety similarly across five states?
**Research question 1 statistical analysis.** In this study, confirmatory factor analysis was used to determine the stability of content areas (i.e., knowledge, performance, school connectedness) across states using post-test data from the S.T.A.R. instrument.

**Research Question 2**

Does the S.T.A.R. instrument measure students’ knowledge of prosocial behaviors, students’ anticipated performance of prosocial behavior, and students’ perception of school safety similarly across race?

**Research question 2 statistical analysis.** In this study, confirmatory factor analysis was used to determine the stability of content areas (i.e., knowledge, performance, school connectedness) across race using post-test data from the S.T.A.R. instrument.

**Research Question 3**

Does the S.T.A.R. instrument measure students’ knowledge of prosocial behaviors, students’ anticipated performance of prosocial behavior, and students’ perception of school safety similarly between low SES and high SES students?

**Research question 3 statistical analysis.** In this study, confirmatory factor analysis was used to determine the stability of content areas (i.e., knowledge, performance, school connectedness) across SES using post-test data from the S.T.A.R. instrument.

**Research Question 4**

Does the S.T.A.R. instrument measure students’ knowledge of prosocial behaviors, students’ anticipated performance of prosocial behavior, and students’ perception of school safety similarly between students from urban and rural environments?

**Research question 4 statistical analysis.** In this study, confirmatory factor analysis was used to determine the stability of content areas (i.e., knowledge, performance, school
connectedness) across community type (i.e., rural vs. urban) using post-test data from the S.T.A.R. instrument.

**Summary**

This study consisted of a secondary analysis of preexisting data. To obtain the initial data that was analyzed in this study, schools across the United States received requests to implement the curriculum according to the curriculum guidelines. Out of 92 requests, 10 schools from 7 different school districts agreed to implement the curriculum. The curriculum is designed to foster a sense of safety and security within the school setting, where students are not concerned about the threat of school violence. The program is based upon the concept of “responsible reporting”, which entails appropriate communication of information when a potentially dangerous situation is evident. The curriculum includes activities, concepts, and objectives that align with age-appropriate development of skills and knowledge of prosocial behaviors for students in kindergarten through eighth grade.

As part of the curriculum, students are administered the S.T.A.R. instrument as a pre- and post-test to assess knowledge of prosocial behaviors, performance of prosocial behaviors, and the perception of school connectedness. The pre-test is administered before the implementation of the curriculum, while the post-test is administered at the conclusion of the program. In order to ensure developmental appropriateness, two versions of the S.T.A.R. assessment were created: one for kindergarten through 3rd grade and one for students in 4th through 8th grade. Existing information regarding the psychometric properties of the S.T.A.R. instrument suggests that it is a generally valid measure of the constructs of knowledge, performance, and school connectedness.

This study utilized a pre-test post-test quasi-experimental design, which consisted of nonrandomized groups. Statistical analyses entailed aligning questions from the S.T.A.R.
instrument by content area and the performing separate confirmatory factor analyses using data from each state. Analyses also focused on categorizing schools based on race, SES, and community type and then conducting confirmatory factor analyses for each grouping. Comparisons regarding the factor structure were then be made to determine the extent to which the structure is similar.
CHAPTER IV

RESULTS

This chapter will present the results of all analyses that were described in chapter three. Descriptive statistics will first be reported for all variables in this study and then statistical assumptions for the statistical tests are examined for purposes of confirming the appropriateness of running the key analyses for each research question. Finally, results of the analyses for each research question are provided.

Descriptive Statistics

Descriptive statistics are utilized for purposes of summarizing and describing data. Frequencies and percentages were used for purposes of describing participant characteristics in relation to age and grade. The S.T.A.R. instruments for first and second grade can be found in Appendix A and B, respectively.

Missing Data

Data was collected from 645 students from eight schools in six school districts across the United States. Since this study only utilized post-test data, any cases with any missing data from the post-test were removed from the data set via list-wise deletion, which resulted in a total of 42 cases (6.5%) being deleted. Explanations were not provided to the researcher regarding the reasons for missing data. Although alternatives such as mean substitution and maximum likelihood estimate can also be used to address missing data, list-wise deletion was determined to be an appropriate technique for this study. With the exception of the school in Arizona, which had roughly one-third of the data eliminated due to missing data, the total number of subjects deleted for all other schools was minimal (under 15%) and did not produce a substantial reduction in sample size across the groups. Due to the fact that all groups produced for this
study were derived from the individual schools, descriptive statistics for eliminated data by school is provided in Table 2. The final sample size, used for all analyses in the current study, was 603.

Table 2

Descriptive Statistics for Eliminated Data by School

<table>
<thead>
<tr>
<th>School</th>
<th>Original N (%)</th>
<th>Missing N (%)</th>
<th>% Eliminated</th>
</tr>
</thead>
<tbody>
<tr>
<td>PA #1</td>
<td>112 (17.4%)</td>
<td>5 (11.9%)</td>
<td>4.46%</td>
</tr>
<tr>
<td>PA #2</td>
<td>125 (19.4%)</td>
<td>11 (26.2%)</td>
<td>8.80%</td>
</tr>
<tr>
<td>PA #3</td>
<td>24 (3.7%)</td>
<td>1 (2.4%)</td>
<td>4.16%</td>
</tr>
<tr>
<td>PA #4</td>
<td>90 (13.9%)</td>
<td>1 (2.4%)</td>
<td>1.11%</td>
</tr>
<tr>
<td>Georgia</td>
<td>139 (21.6%)</td>
<td>9 (21.4%)</td>
<td>6.47%</td>
</tr>
<tr>
<td>Arizona</td>
<td>35 (5.4%)</td>
<td>11 (26.1%)</td>
<td>31.42%</td>
</tr>
<tr>
<td>Florida</td>
<td>55 (8.5%)</td>
<td>2 (4.8%)</td>
<td>3.63%</td>
</tr>
<tr>
<td>Wisconsin</td>
<td>65 (10.1%)</td>
<td>2 (4.8%)</td>
<td>3.07%</td>
</tr>
<tr>
<td>Total</td>
<td>645 (100%)</td>
<td>42 (100%)</td>
<td>6.51%</td>
</tr>
</tbody>
</table>

*Note: % eliminated represents the percentage of original cases per school that were eliminated.*

Participant Characteristics

The final sample consists of 603 students who completed the post-test S.T.A.R. instrument after receiving the *Be a Safety Kid* curriculum. Of the 603 participants, 277 (46%) were male and 326 (54%) were female. Regarding grade level, 332 (55%) of the 603 participants were first graders, while 271 (45%) were second graders. Information regarding the participants’ race and age was not gathered. A detailed description of the gender and grade composition for each individual school is provided in Table 3.
Table 3

*Descriptive Statistics for Gender and Grade by School*

<table>
<thead>
<tr>
<th>School</th>
<th>Gender</th>
<th>Gender N (%)</th>
<th>Grade</th>
<th>Grade N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PA #1</td>
<td>Male</td>
<td>38 (36%)</td>
<td>First</td>
<td>44 (41%)</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>69 (64%)</td>
<td>Second</td>
<td>63 (59%)</td>
</tr>
<tr>
<td>PA #2</td>
<td>Male</td>
<td>60 (53%)</td>
<td>First</td>
<td>46 (40%)</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>54 (47%)</td>
<td>Second</td>
<td>68 (60%)</td>
</tr>
<tr>
<td>PA #3</td>
<td>Male</td>
<td>11 (49%)</td>
<td>First</td>
<td>0 (0%)</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>12 (50%)</td>
<td>Second</td>
<td>23 (100%)</td>
</tr>
<tr>
<td>PA #4</td>
<td>Male</td>
<td>45 (51%)</td>
<td>First</td>
<td>44 (49%)</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>44 (49%)</td>
<td>Second</td>
<td>45 (51%)</td>
</tr>
<tr>
<td>Georgia</td>
<td>Male</td>
<td>53 (41%)</td>
<td>First</td>
<td>74 (60%)</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>77 (59%)</td>
<td>Second</td>
<td>54 (40%)</td>
</tr>
<tr>
<td>Arizona</td>
<td>Male</td>
<td>15 (63%)</td>
<td>First</td>
<td>12 (50%)</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>9 (37%)</td>
<td>Second</td>
<td>12 (50%)</td>
</tr>
<tr>
<td>Florida</td>
<td>Male</td>
<td>25 (47%)</td>
<td>First</td>
<td>53 (100%)</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>28 (53%)</td>
<td>Second</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Wisconsin</td>
<td>Male</td>
<td>29 (46%)</td>
<td>First</td>
<td>33 (51%)</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>34 (54%)</td>
<td>Second</td>
<td>32 (49%)</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>603</td>
<td></td>
<td>603</td>
</tr>
</tbody>
</table>

As indicated previously, the groups created for purposes of the analyses were derived from the individual schools. The schools that comprise each individual group, as well as the descriptive statistics, are presented in Table 4.
Table 4

*Group Composition*

<table>
<thead>
<tr>
<th>Group Name</th>
<th>Group Composition</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pennsylvania</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PA #1</td>
<td></td>
<td>107</td>
</tr>
<tr>
<td>PA #2</td>
<td></td>
<td>114</td>
</tr>
<tr>
<td>PA #3</td>
<td></td>
<td>23</td>
</tr>
<tr>
<td>PA #4</td>
<td></td>
<td>89</td>
</tr>
<tr>
<td><em>Total</em></td>
<td></td>
<td>333</td>
</tr>
<tr>
<td>Florida</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Florida</td>
<td></td>
<td>53</td>
</tr>
<tr>
<td><em>Total</em></td>
<td></td>
<td>53</td>
</tr>
<tr>
<td>Arizona</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arizona</td>
<td></td>
<td>24</td>
</tr>
<tr>
<td><em>Total</em></td>
<td></td>
<td>24</td>
</tr>
<tr>
<td>Wisconsin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wisconsin</td>
<td></td>
<td>65</td>
</tr>
<tr>
<td><em>Total</em></td>
<td></td>
<td>65</td>
</tr>
<tr>
<td>Georgia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Georgia</td>
<td></td>
<td>128</td>
</tr>
<tr>
<td><em>Total</em></td>
<td></td>
<td>128</td>
</tr>
<tr>
<td>High SES</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Georgia</td>
<td></td>
<td>128</td>
</tr>
<tr>
<td>Wisconsin</td>
<td></td>
<td>65</td>
</tr>
<tr>
<td>PA #3</td>
<td></td>
<td>23</td>
</tr>
<tr>
<td><em>Total</em></td>
<td></td>
<td>216</td>
</tr>
<tr>
<td>Low SES</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PA #1</td>
<td></td>
<td>107</td>
</tr>
<tr>
<td>PA #2</td>
<td></td>
<td>114</td>
</tr>
<tr>
<td>PA #4</td>
<td></td>
<td>89</td>
</tr>
<tr>
<td>Florida</td>
<td></td>
<td>53</td>
</tr>
<tr>
<td>Arizona</td>
<td></td>
<td>24</td>
</tr>
<tr>
<td><em>Total</em></td>
<td></td>
<td>387</td>
</tr>
</tbody>
</table>
Table 4 Continued

<table>
<thead>
<tr>
<th>Group Name</th>
<th>Group Composition</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rural</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wisconsin</td>
<td>65</td>
</tr>
<tr>
<td></td>
<td>PA #4</td>
<td>89</td>
</tr>
<tr>
<td></td>
<td><strong>Total</strong></td>
<td><strong>154</strong></td>
</tr>
<tr>
<td>Urban</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Georgia</td>
<td>128</td>
</tr>
<tr>
<td></td>
<td>PA #1</td>
<td>107</td>
</tr>
<tr>
<td></td>
<td>Arizona</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>Florida</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>PA #2</td>
<td>114</td>
</tr>
<tr>
<td></td>
<td>PA #3</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td><strong>Total</strong></td>
<td><strong>449</strong></td>
</tr>
<tr>
<td>Caucasian</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PA #2</td>
<td>114</td>
</tr>
<tr>
<td></td>
<td>PA #4</td>
<td>89</td>
</tr>
<tr>
<td></td>
<td>Georgia</td>
<td>128</td>
</tr>
<tr>
<td></td>
<td>Wisconsin</td>
<td>65</td>
</tr>
<tr>
<td></td>
<td><strong>Total</strong></td>
<td><strong>396</strong></td>
</tr>
<tr>
<td>African American</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PA #1</td>
<td>107</td>
</tr>
<tr>
<td></td>
<td>PA #3</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>Florida</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td><strong>Total</strong></td>
<td><strong>183</strong></td>
</tr>
<tr>
<td>Hispanic</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Arizona</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td><strong>Total</strong></td>
<td><strong>24</strong></td>
</tr>
</tbody>
</table>

88
Statistical Assumptions

Once the groups were created based upon state, community type (rural/urban), race (Caucasian/African American/Hispanic), and socioeconomic status (high SES/low SES), confirmatory factor analysis (CFA) was used to determine the stability and validity of content areas (i.e., knowledge, performance, and school safety) across the post-test for each group. CFA is widely utilized for purposes of determining the extent to which specified relationships between individual scale items and latent (i.e., unmeasured) factors are supported in a sample (Graham, 2012). Furthermore, CFA is of great utility for identifying method effects, analyzing construct validity, creating and revising measurement instruments, and assessing factor invariance across time and groups (Brown, 2006). Conversely, exploratory factor analysis (EFA) is more advantageous when assessing relationships without a predetermined theoretical or empirical foundation or when no previous assumptions regarding the data have been formulated. Thus, given the present investigation regarding the S.T.A.R. instrument and the fact that a theoretical paradigm has already been established, CFA was determined to be the preferred approach in analyzing the data. When utilizing CFA, it is suggested that at least two items be contained within each factor; although three factors are more frequently recommended (Kline, 2005). With the first and second grade S.T.A.R. instruments, each factor (i.e., knowledge, performance, and school safety) had at least two items, which would generally satisfy this recommendation. It is noted that the ratio of items to factors is less critical as the sample size becomes larger (Kline, 2005).

The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett’s test of sphericity were conducted on the post-test sample for each group for purposes of determining the appropriateness of the factor analysis. The KMO statistic is measure of sampling adequacy for
the overall model and also for each individual variable that summarizes the size of the partial correlations in relation to the original, zero-order correlations (Kaiser, 1970). The KMO statistic varies between 0 and 1, where a value of 0 suggests that the sum of the partial correlations is large in relation to the sum of zero-order correlations. A value closer to 1 would indicate that the sum of partial correlations is small in relation to the sum of zero-order correlations and that factor analysis would likely produce discrete and reliable factors (Spicer, 2005). As such, KMO values close to 0 would generally suggest that factor analysis would likely not be suitable since the analyses may not yield distinct factors. For the KMO statistic, Kaiser (1970) recommends that the KMO value be at least .50. Values between .50 and .70 are described as mediocre, values between .70 and .80 are good, values between .80 and .90 are great, and values above .90 are superb (Hutcheson & Sofroniou, 1999). It is further suggested by Fei and Jixin (2011) that, when the KMO statistic is below .50, factor analysis is not appropriate. KMO statistics for this study generally fell within the acceptable range, with 8 out of the 12 groups (Pennsylvania, Florida, Georgia, High SES, Low SES, Urban, Caucasian, and African American) falling between .514 and .614. However, the groups of Arizona, Wisconsin, Rural, and Hispanic had KMO values that fell below the recommended .5 value, which ranged from .327 to .441. Given the KMO values for these 4 groups in relation to the guidance above, these groups were excluded from the statistical analyses. One likely explanation for the lower KMO values of these groups deals with sample size.

According to Mertler and Vannatta (2010), the bases for any underlying factor structure that is derived from a factor analysis are the relationships among all of the original variables. When using a smaller sample size, correlation coefficients tend to be less reliable. If less reliable correlations exist among variables and the variables are subsequently subjected to a factor
analysis, these authors posit that the resulting factors will also be less reliable. One of the common suggestions in the literature regarding the subject-to-variable ratio is the Rule of 10 (Garson, 2008; Everitt, 1975; Kunce, Cook, & Miller, 1975). Here, the guidance is that there should be at least 10 cases for each item in the instrument. Since the S.T.A.R. instrument has 10 items, there should be at least 100 cases in order for the analyses to return reliable factors. The groups of Arizona, Wisconsin, and Hispanic had sample sizes of 24, 65, and 24, respectively, which does not satisfy this recommendation. Thus, these three groups were excluded from further analyses due to the small sample size and inadequate KMO statistic, while the Rural group was excluded solely based upon the KMO statistic despite having an adequate sample. The Florida group, despite having adequate values related to KMO and Bartlett’s test of sphericity, was also excluded given the small sample size of 53.

Bartlett’s test of sphericity tests the null hypothesis that the original correlation matrix is an identity matrix (Tobias & Carlson, 1969). In other words, this test is concerned with determining if the variables in the population correlations matrix are uncorrelated (Mertler and Vannatta, 2010). When the significance value is less than the alpha level of .05, it can be assumed that there are relationships among the variables. For this study, values for Bartlett’s test of sphericity fell below .05 for all groups. However, as indicated previously, KMO values for the groups of Arizona, Wisconsin, Rural, and Hispanic were below the recommended .50 value. Fei and Jixin (2011) suggest that, even when Bartlett’s test of sphericity is within acceptable limits, groups with a KMO value below .50 are not suitable for factor analysis due to the fact that the analysis would likely not yield distinct and reliable factors. A detailed summary of the CFA assumption statistics is provided in Table 5.
Table 5

CFA Assumption Statistics

<table>
<thead>
<tr>
<th>Group</th>
<th>KMO</th>
<th>Bartlett’s Test of Sphericity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>X²</td>
</tr>
<tr>
<td>Pennsylvania</td>
<td>.524</td>
<td>144.535</td>
</tr>
<tr>
<td>Florida</td>
<td>.514</td>
<td>65.347</td>
</tr>
<tr>
<td>Arizona</td>
<td>.327</td>
<td>65.887</td>
</tr>
<tr>
<td>Wisconsin</td>
<td>.441</td>
<td>30.882</td>
</tr>
<tr>
<td>Georgia</td>
<td>.538</td>
<td>130.590</td>
</tr>
<tr>
<td>High SES</td>
<td>.576</td>
<td>181.803</td>
</tr>
<tr>
<td>Low SES</td>
<td>.559</td>
<td>108.301</td>
</tr>
<tr>
<td>Rural</td>
<td>.441</td>
<td>133.275</td>
</tr>
<tr>
<td>Urban</td>
<td>.614</td>
<td>184.517</td>
</tr>
<tr>
<td>Caucasian</td>
<td>.600</td>
<td>175.223</td>
</tr>
<tr>
<td>African American</td>
<td>.547</td>
<td>95.881</td>
</tr>
<tr>
<td>Hispanic</td>
<td>.327</td>
<td>65.887</td>
</tr>
</tbody>
</table>

Data Analysis

Research Question 1

The first research question was designed to determine the extent to which the S.T.A.R. instrument measures students’ knowledge of prosocial behaviors, students’ anticipated performance of prosocial behavior, and students’ perception of school safety similarly across the five states of Pennsylvania, Florida, Arizona, Wisconsin, and Georgia. Based upon the theoretical underpinnings of the S.T.A.R. instrument for first and second grade, the analyses should produce a three factor model in which Questions 1, 2, 3, 4, 7, and 8 align with knowledge; Questions 5 and 6 align with anticipated performance of prosocial behaviors; and Questions 9 and 10 align with the perception of school safety. Due to the fact that item wording
for the first and second grade S.T.A.R. instruments were similar, separate analyses for each grade was not required.

For purposes of evaluating overall validity, construct and face validity were assessed. Face validity was established previously via soliciting input from experts in the fields of school psychology, intervention, and child violence for their expert opinions of the S.T.A.R. instrument while it was being created. Various school psychologists, police officers, statistics professors, principals, teachers, and children, as well as the creator of the Be a Safety Kid curriculum, offered feedback regarding facets of the instrument and its alignment with theoretical constructs. For the present study, construct validity was determined by conducting confirmatory factor analysis using LISREL 9.1 (Jöreskog & Sörbom, 2014) with maximum likelihood estimation to assess for model fit. Each model was examined once using post-test data for the students who participated in the study.

The assessment of overall model fit for each group was based upon various fit indices (e.g., normed fit index [NFI], comparative fit index [CFI], root mean square error of approximation [RMSEA]). For one-time analyses, Byrne (1994) recommends using the NFI, CFI, and RMSEA for SEM reporting. Normed fit indices ≥ .95 suggest a better fit; however, NFI may be accepted if the value is greater than 0 and less than 1 (Schreiber, Nora, Stage, Barlow, & King, 2006). CFI values that are higher indicate a better fit than lower values, with values of .93 or above being desirable (Byrne, 1994). RMSEA is frequently discussed in relation to “badness of fit”, with lower values indicative of a good model fit. Values ≤ 0.5 are preferred, but values between .05 and .08 are generally regarded as acceptable (Kline, 2005). Acceptable models will also have a relative chi-square of less than 2 or 3 (Kline, 1998; Ullman, 2001) and a Goodness of Fit Index score that exceeds .90 (Byrne, 1994). Squared multiple correlations
produced by CFA should also be reported in order to determine the proportion of variance accounted for in the endogenous variables (Schreiber et al., 2006). These guidelines were used for purposes of interpreting the results of the current study.

Using the overall model fit guidelines above, results of the confirmatory factor analysis for each state (Pennsylvania and Georgia) were examined to determine adequacy of fit. The groups of Arizona and Wisconsin were excluded from analyses due to KMO values below .50, while Florida was excluded due to an inadequate sample. It is noted that, prior to conducting the CFA, any variables with zero variance were removed. No variables were removed from Pennsylvania or Georgia.

Results of the CFA suggested that states of Georgia and Pennsylvania had values for fit indices that generally fell within the unacceptable range. Georgia had an estimated CFI of 0.17, RMSEA of 0.22, NFI of 0.18, and GFI of 0.77. Pennsylvania had a CFI of 0.55, RMSEA of 0.15, NFI of 0.53, and GFI of 0.85. Across models for all states the proportion of variance accounted for by each variable was determined to be low across all variables, which suggests that the model explains little of the variation for questions on the S.T.A.R. instrument. Chi-square, degrees of freedom, CFI, RMSEA, NFI, and GFI estimates are provided for each state in Table 6. Squared multiple correlations (i.e., $R^2$) are reported in Table 7.

Table 6

*Comparison of CFA Model Fit Indices for States*

<table>
<thead>
<tr>
<th>State</th>
<th>N</th>
<th>$X^2$</th>
<th>df</th>
<th>CFI</th>
<th>RMSEA</th>
<th>NFI</th>
<th>GFI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pennsylvania</td>
<td>333</td>
<td>294.15</td>
<td>32</td>
<td>0.55</td>
<td>0.15</td>
<td>0.53</td>
<td>0.85</td>
</tr>
<tr>
<td>Georgia</td>
<td>130</td>
<td>249.43</td>
<td>32</td>
<td>0.17</td>
<td>0.22</td>
<td>0.18</td>
<td>0.77</td>
</tr>
</tbody>
</table>
Table 7

Squared Multiple Correlations for States

<table>
<thead>
<tr>
<th>Variable</th>
<th>Factor</th>
<th>Pennsylvania</th>
<th>Georgia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question 1</td>
<td>Knowledge</td>
<td>0.05</td>
<td>0.95</td>
</tr>
<tr>
<td>Question 2</td>
<td>Knowledge</td>
<td>0.05</td>
<td>0.06</td>
</tr>
<tr>
<td>Questions 3</td>
<td>Knowledge</td>
<td>0.08</td>
<td>0.10</td>
</tr>
<tr>
<td>Question 4</td>
<td>Knowledge</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Question 5</td>
<td>Skill</td>
<td>0.11</td>
<td>0.10</td>
</tr>
<tr>
<td>Question 6</td>
<td>Skill</td>
<td>0.01</td>
<td>0.11</td>
</tr>
<tr>
<td>Question 7</td>
<td>Knowledge</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Question 8</td>
<td>Knowledge</td>
<td>0.04</td>
<td>0.00</td>
</tr>
<tr>
<td>Question 9</td>
<td>Safety</td>
<td>0.10</td>
<td>0.18</td>
</tr>
<tr>
<td>Question 10</td>
<td>Safety</td>
<td>0.25</td>
<td>0.00</td>
</tr>
</tbody>
</table>

**Research Question 2**

The second research question sought to determine the extent to which the S.T.A.R. instrument measures students’ knowledge of prosocial behaviors, students’ anticipated performance of prosocial behavior, and students’ perception of school safety similarly across race. As described in the previous chapter, each school was placed into one of three categories (Caucasian, African American, Hispanic) based upon the race that was the majority (i.e., 50% or greater) of the students within that school. Confirmatory factor analysis was then conducted on the Caucasian and African American groups using the model fit guidelines previously described in relation to the first research question. As previously discussed, the Hispanic group was excluded from analyses due to the KMO statistic and inadequate sample. Removal of variables with no variance resulted in only the exclusion of Question 3 from the African American group.
Examination of the overall model fit for each group suggested that the Caucasian group had fit indices that generally fell within the unacceptable range, with an estimated CFI of 0.61, RMSEA of 0.11, NFI of 0.58, and GFI of 0.91. The African American group had fit indices that fell within the acceptable range, with a CFI of 1.00, RMSEA of 0.00, NFI of 0.23, and GFI of 0.99. As with the state analyses, squared multiple correlations were generally low across groups.

Chi-square, degrees of freedom, CFI, RMSEA, NFI, and GFI estimates are provided for each state in Table 8. Squared multiple correlations (i.e., $R^2$) are reported in Table 9.

### Table 8

*Comparison of CFA Model Fit Indices for Race*

<table>
<thead>
<tr>
<th>Race</th>
<th>N</th>
<th>$X^2$</th>
<th>df</th>
<th>CFI</th>
<th>RMSEA</th>
<th>NFI</th>
<th>GFI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caucasian</td>
<td>396</td>
<td>206.64</td>
<td>32</td>
<td>0.61</td>
<td>0.11</td>
<td>0.58</td>
<td>0.91</td>
</tr>
<tr>
<td>African American</td>
<td>183</td>
<td>8.34</td>
<td>24</td>
<td>1.00</td>
<td>0.00</td>
<td>0.23</td>
<td>0.99</td>
</tr>
</tbody>
</table>

### Table 9

*Squared Multiple Correlations for Race*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Factor</th>
<th>Caucasian</th>
<th>African American</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question 1</td>
<td>Knowledge</td>
<td>0.01</td>
<td>0.33</td>
</tr>
<tr>
<td>Question 2</td>
<td>Knowledge</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>Questions 3</td>
<td>Knowledge</td>
<td>0.66</td>
<td>N/A</td>
</tr>
<tr>
<td>Question 4</td>
<td>Knowledge</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Question 5</td>
<td>Skill</td>
<td>0.01</td>
<td>0.12</td>
</tr>
<tr>
<td>Question 6</td>
<td>Skill</td>
<td>0.00</td>
<td>0.02</td>
</tr>
<tr>
<td>Question 7</td>
<td>Knowledge</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Question 8</td>
<td>Knowledge</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Question 9</td>
<td>Safety</td>
<td>0.26</td>
<td>0.02</td>
</tr>
<tr>
<td>Question 10</td>
<td>Safety</td>
<td>0.11</td>
<td>0.00</td>
</tr>
</tbody>
</table>

### Research Question 3
The third research question aimed to determine the extent to which the S.T.A.R. instrument measures students’ knowledge of prosocial behaviors, students’ anticipated performance of prosocial behavior, and students’ perception of school safety similarly across socioeconomic status (SES). As described in the previous chapter, schools were divided into two groups (high SES and low SES) based upon the percentage of students receiving free/reduced lunches. Confirmatory factor analysis was the performed on both groups to determine overall model fit utilizing the recommendations provided for the previous two research questions. No variables for either data set lacked variance, so no variables required exclusion from the analyses.

Analysis of the model fit indices for each group revealed that both groups had indices that fell within the unacceptable range. The high SES group had an estimated CFI of 0.25, RMSEA of 0.20, NFI of 0.25, and GFI of 0.80, while the low SES group had a CFI of 0.53, RMSEA of 0.13, NFI of 0.51, and GFI of 0.88. Squared multiple correlations for all variables were low across both groups. Chi-square, degrees of freedom, CFI, RMSEA, NFI, and GFI estimates are provided for each group in Table 10. Squared multiple correlations (i.e., R²) are reported in Table 11.

Table 10

Comparison of CFA Model Fit Indices for SES

<table>
<thead>
<tr>
<th>SES</th>
<th>N</th>
<th>X²</th>
<th>df</th>
<th>CFI</th>
<th>RMSEA</th>
<th>NFI</th>
<th>GFI</th>
</tr>
</thead>
<tbody>
<tr>
<td>High SES</td>
<td>216</td>
<td>330.03</td>
<td>32</td>
<td>0.25</td>
<td>0.20</td>
<td>0.25</td>
<td>0.80</td>
</tr>
<tr>
<td>Low SES</td>
<td>387</td>
<td>268.27</td>
<td>32</td>
<td>0.53</td>
<td>0.13</td>
<td>0.51</td>
<td>0.88</td>
</tr>
</tbody>
</table>

Table 11

Squared Multiple Correlations for SES
<table>
<thead>
<tr>
<th>Variable</th>
<th>Factor</th>
<th>High SES</th>
<th>Low SES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question 1</td>
<td>Knowledge</td>
<td>0.01</td>
<td>0.03</td>
</tr>
<tr>
<td>Question 2</td>
<td>Knowledge</td>
<td>0.00</td>
<td>0.04</td>
</tr>
</tbody>
</table>

Table 11 Continued

<table>
<thead>
<tr>
<th>Variable</th>
<th>Factor</th>
<th>High SES</th>
<th>Low SES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Questions 3</td>
<td>Knowledge</td>
<td>0.00</td>
<td>0.15</td>
</tr>
<tr>
<td>Question 4</td>
<td>Knowledge</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Question 5</td>
<td>Skill</td>
<td>0.00</td>
<td>0.26</td>
</tr>
<tr>
<td>Question 6</td>
<td>Skill</td>
<td>0.00</td>
<td>0.07</td>
</tr>
<tr>
<td>Question 7</td>
<td>Knowledge</td>
<td>0.01</td>
<td>0.06</td>
</tr>
<tr>
<td>Question 8</td>
<td>Knowledge</td>
<td>0.00</td>
<td>0.04</td>
</tr>
<tr>
<td>Question 9</td>
<td>Safety</td>
<td>0.24</td>
<td>0.01</td>
</tr>
<tr>
<td>Question 10</td>
<td>Safety</td>
<td>0.01</td>
<td>0.01</td>
</tr>
</tbody>
</table>

**Research Question 4**

The fourth and final research question sought to determine the extent to which the S.T.A.R. instrument measures students’ knowledge of prosocial behaviors, students’ anticipated performance of prosocial behavior, and students’ perception of school safety similarly across community type. As discussed in chapter three, schools were divided into two groups (rural and urban) based upon population estimates of the community in which the school is situated. Confirmatory factor analysis was then performed on the Urban group and assessed in relation to the model fit guidelines discussed previously. The Rural group was excluded from analyses due to the KMO statistic previously discussed. No variables required exclusion due to having no variance.

Results of confirmatory factor analysis indicated that the Urban group had a model fit that fell within the unacceptable range, with a CFI of 0.72, RMSEA of 0.09, NFI of 0.17, and GFI of 0.97. Squared multiple correlations for all variables were low for the Urban group. Chi-
square, degrees of freedom, CFI, RMSEA, NFI, and GFI estimates are provided in Table 12.

Squared multiple correlations (i.e., $R^2$) are reported in Table 13.

Table 12

**Comparison of CFA Model Fit Indices for Community Type**

<table>
<thead>
<tr>
<th>SES</th>
<th>N</th>
<th>$X^2$</th>
<th>df</th>
<th>CFI</th>
<th>RMSEA</th>
<th>NFI</th>
<th>GFI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urban</td>
<td>451</td>
<td>159.25</td>
<td>32</td>
<td>0.72</td>
<td>0.09</td>
<td>0.68</td>
<td>0.97</td>
</tr>
</tbody>
</table>

Table 13

**Squared Multiple Correlations for Community Type**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Factor</th>
<th>Urban</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question 1</td>
<td>Knowledge</td>
<td>0.01</td>
</tr>
<tr>
<td>Question 2</td>
<td>Knowledge</td>
<td>0.01</td>
</tr>
<tr>
<td>Questions 3</td>
<td>Knowledge</td>
<td>0.07</td>
</tr>
<tr>
<td>Question 4</td>
<td>Knowledge</td>
<td>0.00</td>
</tr>
<tr>
<td>Question 5</td>
<td>Skill</td>
<td>0.08</td>
</tr>
<tr>
<td>Question 6</td>
<td>Skill</td>
<td>0.07</td>
</tr>
<tr>
<td>Question 7</td>
<td>Knowledge</td>
<td>0.00</td>
</tr>
<tr>
<td>Question 8</td>
<td>Knowledge</td>
<td>0.00</td>
</tr>
<tr>
<td>Question 9</td>
<td>Safety</td>
<td>0.17</td>
</tr>
<tr>
<td>Question 10</td>
<td>Safety</td>
<td>0.03</td>
</tr>
</tbody>
</table>

**Summary**

Results of the first research question examining the extent to which the S.T.A.R. instrument measures constructs consistently across states suggests a generally poor fit. All state groups (Pennsylvania and Georgia) had model fit indices that fell within the unacceptable range.

Regarding the second research question, acceptable model fit was indicated for the African American group, but not the Caucasian group. For the third research question, both the high SES and low SES groups had an overall fit that was considered to be unacceptable. For the fourth
research question, the model fit for the Urban group was found to be within the unacceptable range. It is again noted that the groups of Arizona, Wisconsin, and Hispanic were excluded based upon the KMO statistic and insufficient sample, while the Rural group was excluded based upon the KMO statistic alone. The Florida group was excluded based upon sample size alone. For all groups across the four research questions, squared multiple correlations suggested that the proportion of variance accounted for in the endogenous variables was low. Results suggest that the S.T.A.R. instrument does not consistently align with the constructs designed through its creation.
CHAPTER V
DISCUSSION

The results of the statistical analyses as presented in chapter four of the present study are described more fully in this chapter. First, the findings will be summarized briefly with emphasis on providing answers to the four research questions guiding the current study. Limitations to the current investigation will then be examined, followed by recommendations for future research. Lastly, conclusions and implications based upon these results are discussed.

Summary of Research Findings

Although four separate research questions were formulated for the current study, the overarching objective of this investigation was to determine the extent to which the S.T.A.R. instrument measures the constructs of knowledge, performance, and school safety similarly across cultures and contexts. To do so, the first research question sought to determine the validity and stability of the S.T.A.R. instrument across five different states (Pennsylvania, Georgia, Arizona, Wisconsin, and Florida) by measuring its ability to align with the constructs previously established during the creation of the instrument. Results indicated that, despite the fact that a three factor model was produced for each state, the overall model fit for those states was poor. It is again noted that the states of Arizona, Wisconsin, and Florida were excluded. The general conclusion is that the model fit explicates little of the variation in the data for the majority of the questions generated through empirical and theoretical analysis. Thus, the S.T.A.R. instrument was largely found to be a poor measure of the constructs across states.

The second research question was aimed at determining if the S.T.A.R. instrument measures the constructs of knowledge, performance, and school safety similarly across race. In comparing schools that were either predominately, Caucasian or African American, it is noted
that a three factor model was again produced for each group. The Hispanic group was excluded from analyses. The model fit for the Caucasian group was generally unacceptable according to specified guidelines, while the model fit for the African American group was acceptable. Furthermore, race analyses would suggest that the model explains little of the variation in the data for questions previously developed via theoretical and empirical analyses.

Similar results were found for the third research question, which sought to determine if the S.T.A.R. instrument measured the constructs of knowledge, performance, and school safety similarly across various levels of socio-economic status (SES). Analyses indicated a poor model fit for both SES groups. Once again a three factor model was produced, yet squared multiple correlations indicated that little variance was explained by the questions on the S.T.A.R. instrument.

Lastly, the final research question examined the extent to which the S.T.A.R. instrument measured the constructs of knowledge, performance, and school safety similarly across different community types (i.e., rural vs urban). An unacceptable model fit was obtained for the Urban group, while the Rural group was excluded from the analyses. As with previous comparisons, values for squared multiple correlations suggested that little variance was explained by the questions on the S.T.A.R. instrument despite the fact that a three factor model was produced.

The overall conclusion regarding analyses related to the four research questions is that the S.T.A.R. instrument does not consistently align with the constructs of knowledge, performance, and school safety. Although face validity was previously established between the test questions on the S.T.A.R. instrument and these three constructs, statistical analyses would suggest a lack of differentiated constructs on the instrument. This may be attributable to the low variance values observed in the instrument. However, the absence of cohesive constructs has
previously been demonstrated in research conducted by authors such as Eisenberg et al. (1999); Carlo and Randall (2001); Jackson and Tisak (2011); Penner et al. (2005); and Radke-Yarrow and Zahn-Waxler (1986), which affirms the general lack of consensus on the precise definition of prosocial behavior. Furthermore, results of this study, as well as previous research, further illustrates the difficulties in accurately measuring prosocial behaviors. The findings of the present study are comparable to those obtained from pilot studies conducted by Martin (2010) and Graham (2012), which also suggest a lack of differentiated constructs on the S.T.A.R. instrument.

Out of all of the groups included in the analyses, only the African American group was found to have an acceptable model fit according to the indices. This may suggest that the S.T.A.R. instrument is a more valid assessment for this group than the other groups. Yet, it is also noted that squared multiple correlations were generally low for the African American group. Thus, while an acceptable model fit was obtained, the fact remains that little variance was explicated by the questions on the instrument.

**Limitations**

Several limitations were inherent in the present research study. One of the primary limitations deals with the small sample size of some of the derived groups. Although the sample size of the total aggregate was more than adequate, creating subsets (i.e., groups) out of the data resulted in some groups having an insufficient sample size. For example, the Hispanic and Arizona groups only had an $N$ of 24, while the Florida group had an $N$ of 53. Given the nature of the statistical analyses performed, as well as the large number of variables, the small sample sizes generally prohibited running factor analyses on some groups. Due to the fact that the purpose of this study was to examine any similarities or differences of the underlying factor
structure of the S.T.A.R. instrument across groups, the exclusion of some groups made comparisons unfeasible.

Another limitation deals with the fact that the S.T.A.R. instrument is a self-report measure of prosocial behavior. Due to the fact that respondents are self-reporting the frequency with which they engage in prosocial acting, this ultimately creates an imperfect indicator of prosocial responding since individuals may attempt to portray themselves as more altruistic than what they actually are (Eisenberg & Mussen, 1989). Furthermore, self-report measures of prosocial acting often rely on assumptions that the respondent understands the construct, knows which behavior relates to the construct, and comprehends the reference points (Greener, 2000). Since the questions on the S.T.A.R. instrument were associated with contrived or theoretical situations, the questions may not be precisely replicate real-life conditions.

An additional and related limitation deals with the extent to which respondents understood the questions they were being asked to answer. Although the S.T.A.R. instrument was previously assessed for readability using the Spache Readability Formula, it is unclear as to how the overall reading abilities of the respondents may have impacted the results of the S.T.A.R. instrument. While average to above average students likely had no problems reading questions, students with below average reading skills or a specific learning disability in reading may have experienced difficulty reading and deriving meaning from the questions. As such, their responses may potentially reflect an incorrect understanding of the question or, at worst, simply reflect a guess or randomly selected answer.

The final limitation concerns the general lack of consensus regarding the specific behavioral manifestations and definitions of the overarching construct of prosocial behavior. Previous research has demonstrated the difficulty in accurately defining and measuring prosocial
behavior and the results of the present study, especially in light of the low squared multiple correlation values, would again elucidate this concern. Despite established face validity and alignment of constructs with theoretical models, the variance explained by the questions on the S.T.A.R. instrument was generally low.

**Implications**

Acts of school violence create dangerous and potentially deadly consequences for students within the educational setting. Such acts do not occur within a vacuum and, in most instances, someone else was cognizant of an individual’s plan or idea to perpetrate a violent or unsafe act prior to the act being committed. Here, the value of responsible reporting cannot be overemphasized since communicating information about a possible act of violence may prevent a disastrous situation. For that reason, interventions and curricula designed to teach prosocial behavior and responsible reporting are of significant importance. Although this study was not specifically concerned with the overall efficacy of the *Be a Safety Kid* curriculum, previous research conducted by Martin (2011) and Graham (2012) suggest that this curriculum is capable of producing positive changes with regard to increasing knowledge and anticipated performance of prosocial behavior. Yet, these studies, as well as the present study, reaffirm the inherent difficulties in accurately measuring prosocial behavior.

As discussed in previous chapters, it is important to have accurate and valid assessments for measuring skills being taught, whether these skills are related to reading, math, or prosocial behavior. However, there are very few instruments specifically designed to assess prosocial acting. This is likely due to the problematic nature of accurately defining the construct of prosocial behavior and the subsets of behavior that comprise this construct. At face value, the items on the S.T.A.R. instrument were determined to be suitable for measuring knowledge of
prosocial behavior, anticipated performance of prosocial behavior, and the perception of school safety. Yet, statistical analyses suggested that this instrument was a relatively poor measure of these constructs. The results of this study, as well as findings from previous research, strongly suggest that there is substantial room for advancement with regard to the assessment of prosocial behavior.

Results of this study, as well as existing literature regarding prosocial behaviors, yield several suggestions for how to potentially improve assessments for prosocial behavior. First, it is important to recognize that, although prosocial behaviors generally increase with age, the types of prosocial behaviors displayed as children develop are dynamic. For example, sharing and donating (but not helping) generally increase in adolescence, while helping victims of aggression generally decreases in adolescence (Eisenberg & Fabes, 1998). As such, it is important that assessments accurately capture the prosocial behaviors typically associated with various ages and ensure that the assessment accounts for this variability in prosocial responding. Secondly, it is suggested that assessments attempt to utilize response types that are more of a Likert-based scale than a simple Yes/No format. Although younger children may more easily understand Yes/No questions, such a format greatly decreases variability in responding and makes it difficult to account for subtle differences in responses since this represents an “all or nothing” approach. Another suggestion is that assessments attempt to incorporate the theoretical bases from which prosocial behaviors were derived. This entails considering the cognitive-developmental framework of Piaget (1965) and the moral reasoning paradigm proposed by Kohlberg (1984) when designing assessments. Here, it is suggested that assessments make more of an attempt to obtain valuable information on the cognitive processes that mediate prosocial acting. For example, assessments could ask the student why they responded in the manner they did.
Additionally, assessments should incorporate age-appropriate moral dilemmas that students could subsequently respond to. This may provide a more accurate glimpse of their prosocial functioning and provide a better gauge of how likely they are to respond prosocially. Lastly, assessment of prosocial behavior should entail a comprehensive consideration of the various environmental factors that may promote or inhibit prosocial acting. In the school setting, these factors can include teacher personality, school climate, classroom structure and rules, and family modeling. In essence, the question centers on the presence of factors that may make the child more or less likely to behave prosocially. Failure to consider these factors not only makes it difficult to adequately determine the impact of prosocial curricula and interventions, but may also skew results of assessments.

**Recommendations for Future Research**

The present study continues to highlight the difficulties of measuring prosocial behavior. To that regard, future research in this area should be aimed at furthering the general understanding of the complex construct of prosocial behavior. More importantly, however, research in this area should focus on identifying valid and accurate ways to measure prosocial behavior. Although a variety of interventions and curricula exist for purposes of teaching and reinforcing prosocial behaviors, the general dearth of assessments that accurately measure these behaviors is cause for concern. Without effective instruments, measuring progress and development as a result of the interventions and curricula becomes problematic.

To a large extent, the present study was concerned with the extent to which cultural and contextual factors impact the manner in which the S.T.A.R. instrument measures the constructs of knowledge, performance, and perception of school safety. However, issues with sample size made accurate comparisons among groups difficult. To that regard, future research should
continue to explore not only how culture and context influence the development of prosocial behavior, but also how they affect the manner in which prosocial behaviors are assessed. Existing research suggests that the development of prosocial behaviors is best understood by considering the complex interaction between biological and cultural factors. Logically, one could assume that consideration of the interaction between biology and culture is also vital to producing a better understanding of how to best measure this multifaceted set of behaviors. However, very few studies have attempted to explicate this relationship.

Another recommendation, born largely out of the extensive review of the existing literature on prosocial behavior, is that future research make more of a concerted effort to conduct studies in more naturalistic settings as opposed to laboratory environments. It is postulated that a more accurate understanding of these behaviors can be obtained by removing any external cues that are often utilized to elicit prosocial behaviors in laboratory settings. Furthermore, examination of prosocial behaviors in natural settings would likely provide more accurate data in terms of the frequency that individuals engage in prosocial acting.

Lastly, existing research on prosocial behavior strongly suggests that these behaviors are often mediated by complex cognitive processes. Yet, few studies have attempted to empirically examine the role that these processes play in an individual’s tendency to act prosocially. In addition to more research on culture and prosocial behaviors, additional studies are needed to further explore the role that cognitive processes play in prosocial behavior. It is noted, however, that cognitive processes are often complex and difficult to quantify. Nonetheless, additional research in this area will contribute to a more complete understanding of prosocial behaviors.
Conclusions

The tragedy caused by acts of school violence has precipitated an intense interest in finding the most effective ways to prevent these acts. Within the educational setting, efforts have largely focused on the implementation of interventions and curricula designed to teach prosocial behaviors. Literature regarding prosocial behaviors reveals that these behaviors have significant implications for a child’s social competence and the extent to which a child is accepted by his or her peer group (Denham & Holt, 1993). There is also evidence to suggest that the propensity to display prosocial behaviors may serve as a protective factor and decrease the likelihood of externalizing problems, such as aggression and delinquency, later in life (Hastings, Zahn-Waxler, Robinson, Usher, & Bridges, 2000). Despite the importance of these behaviors, there continues to be a lack of consensus with regard defining and measuring these behaviors.

From a historical viewpoint, research and theoretical perspectives have generally suggested that prosocial behaviors develop according to a specific trajectory (e.g., Piaget, 1965; Kohlberg, 1984; Darley & Latane, 1968). However, prominent researchers such as Snarey (1985), Miller and Bersoff (1992), and Keller et al. (2005) have found that the development of moral reasoning and altruistic tendencies is not as universal and prescribed as what was initially believed. Later research supported the influence of culture and context by demonstrating that factors such as religion, community type, socioeconomic status, political orientation, etc. can impact an individual’s proclivity to act in a prosocial manner. Despite these findings, most cross-cultural work on prosocial behaviors has subscribed to an etic framework, which assumes a culturally neutral perspective and also postulates that the constructs being measured have relevance across all cultures.
In terms of the assessment of prosocial behaviors, adherence to an *etic* framework can have significant implications for the accuracy with which behaviors are measured across cultures. The definitions of *helping* or *sharing* may differ between cultural contexts and may cause inaccurate measurement of these constructs by assessments that assume a universal definition of constructs. This study attempted to elucidate the extent to which culture and context impacted accurate assessment using the S.T.A.R. instrument. Although comparisons between groups was difficult due to group exclusion, the general conclusion is that the S.T.A.R. instrument does not serve as a valid indicator in relation to knowledge of prosocial behaviors, anticipated performance of prosocial behaviors, and perception of school safety.

The importance of intervention and assessment of prosocial behavior cannot be understated. However, given the role of culture in prosocial behavior, more research is needed to further clarify the specific mechanisms by which culture not only impacts the accurate definition of prosocial behavior, but also the assessment of these behaviors. Additional work in this area will ultimately lead to a better understanding of these behaviors and, hopefully, facilitate the development of more effective interventions and assessments.
References


doi:10.1207/S1532480XADS0801_3


Guyatt, G. H., Townsend, M., Berman, L. B., & Keller, J. L. (1987). A comparison of Likert and
visual analogue scales for measuring change in function. *Journal of Chronic Disease, 40*, 1129-1133.


doi:10.1146/annurev.psych.56.091103.070141


University of Oregon Center on Teaching and Learning. (2014). Dynamic Indicators of Basic Early Literacy Skills. Retrieved February 17, 2014, from https://dibels.uoregon.edu/


124


APPENDIX A

First Grade S.T.A.R. Instrument
S.T.A.R. Grade 1

Teachers read questions and directions aloud. Students circle the best answer.

You are sitting at lunch with your friends. Two students come up to your table. They tell you to give them your food or they will hurt you after school.

1. How did you know what was happening? Circle which sense you used

2. Could someone get hurt?
   - YES
   - NO

3. Should you tell an adult?
   - YES
   - NO

4. Do you know what information you need to tell an adult?
   - YES
   - NO

5. Would you tell an adult about these kids?
   - YES
   - NO
Student Name or ID Number

Date
Pre-test / Post-test (circle one)
Male / Female (circle one)

6. Would you be afraid to tell an adult?

YES NO

7. A responsible reporter is a tattletale.

YES NO

8. It is important to think about what we sense (see, hear, taste, touch, smell).

YES NO

9. Is there an adult you feel you can talk to at school when you see something bad happens?

YES NO

10. Do you feel safe at your school?

YES NO
APPENDIX B

Second Grade S.T.A.R. Instrument

S.T.A.R.  Grade 2

Teachers read questions and directions aloud. Students circle the best answer.

You are riding home on the school bus when you hear a couple kids threatening another child. They are telling her that she better give them money or they will beat her up.

1. How did you know what was happening? Circle which sense you used.

2. Could someone get hurt?
   YES  NO

3. Should an adult know about this?
   YES  NO

4. Do you know what information you need to tell an adult about something that is unsafe or wrong?
   YES  NO

5. Would you tell an adult about those kids?
   YES  NO
Student Name or ID Number __________________________

Date

Pre-test / Post-test (circle one)

Male / Female (circle one)

6. Would you be afraid to tell an adult about something someone was doing that was unsafe or wrong?

YES

NO

7. Responsible reporting is telling the teacher so another child gets in trouble.

YES

NO

8. It is good to report something if it prevents someone from getting hurt.

YES

NO

9. Is there an adult you feel you can talk to at school when you know about something unsafe or wrong?

YES

NO

10. Do you feel safe at your school?

YES

NO