

10-22-2004

Helping Break the Cycle of School Violence and Aggression: A Program Evaluation of the Owning Up Curricula

Stephanie T. Mihalas
University of South Florida

Follow this and additional works at: <https://scholarcommons.usf.edu/etd>

 Part of the [American Studies Commons](#)

Scholar Commons Citation

Mihalas, Stephanie T., "Helping Break the Cycle of School Violence and Aggression: A Program Evaluation of the Owning Up Curricula" (2004). *Graduate Theses and Dissertations*.
<https://scholarcommons.usf.edu/etd/1159>

This Dissertation is brought to you for free and open access by the Graduate School at Scholar Commons. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact scholarcommons@usf.edu.

Helping Break the Cycle of School Violence and Aggression: A Program Evaluation of
the Owing Up Curricula

by

Stephanie T. Mihalas

A thesis submitted in partial fulfillment
of the requirements for the degree of
Education Specialist
Department of Psychological and Social Foundations
College of Education
University of South Florida

Major Professor: Linda Raffaele Mendez, Ph.D.
Harold Keller, Ph.D.
Jeffrey Kromery, Ph.D.

Date of Approval:
October 22, 2004

Keywords: gender-based violence, relational aggression, physical aggression,
program evaluation

© Copyright 2005, Stephanie Mihalas

Acknowledgements

The experience of proposing and defending my thesis has provided me with tremendous growth opportunities both personally and academically. I have been challenged continuously throughout the process, as I developed into a competent researcher.

However, persevering through the thesis process would have not been possible without the help of many dear people. First and foremost, I would like to thank my committee for supporting me throughout the stages of this project. Their support, time, and genuine interest in making this research project successful were insurmountable. I would also like to thank The Ophelia Project-Tampa Bay for giving me the opportunity to work on a very important piece of research that will have implications for future aggression studies.

Kudos also goes to Robin Hardesty, MaryLou, and William for supplying me with doses of reassurance that there was a light at the end of the tunnel. Extra thanks to Keith Anders for taking the time to serve as my official copywriter, confidant, and best friend when the times appeared to be very rough. Finally, I would not be where I am today without the constant love and encouragement from the two best people in the entire world: my parents.

Table of Contents

| | |
|---|-----|
| List of Tables | iv |
| List of Figures | vi |
| Abstract | vii |
| Chapter 1: Introduction | 1 |
| Brief review of relational aggression literature | 1 |
| Significance | 3 |
| Rationale for Current Study | 3 |
| Purpose of the Current Study | 4 |
| Summary and Description of the Current Study | 5 |
| Chapter 2: Review of the Literature | 7 |
| Forms of Aggression and Associated Prevalence Rates | 8 |
| Physical aggression | 9 |
| Indirect aggression | 12 |
| Social aggression | 14 |
| Relational aggression | 15 |
| Risk Factors | 16 |
| Physical aggression | 16 |
| Social and relational aggression | 18 |
| Concurrent and Future Social Adjustment Issues | 18 |
| Implications of RA for perpetrators and victims | 18 |
| Consequences of RA for victims | 18 |
| Implications for perpetrators | 24 |
| Summary | 26 |
| Assessment Tools and RA | 26 |
| Preschool-aged children | 27 |
| Elementary-aged children | 31 |
| College-aged students | 32 |
| Summary | 33 |
| Prevention and Intervention Efforts | 34 |
| Primary prevention-General aggression | 35 |
| Secondary prevention-General aggression | 40 |
| Tertiary prevention-General aggression | 42 |
| Prevention and relational aggression | 45 |
| CASS | 45 |

| | |
|--|-----|
| Allies in Action | 46 |
| Owning Up | 48 |
| Limitations and Barriers to Prevention | 51 |
| Culture | 51 |
| Community | 52 |
| School | 52 |
| Summary | 53 |
| Summary and Future Research | 53 |
| Current study | 55 |
| Chapter 3: Methods | 57 |
| Overview | 57 |
| Development of Study | 57 |
| Participants and Setting | 58 |
| Procedure | 60 |
| Planning | 60 |
| Pre-data collection | 61 |
| Data collection | 65 |
| Intervention | 67 |
| Post-data collection | 70 |
| Measures | 71 |
| Dependent Measures | 72 |
| The Aggression Questionnaire | 72 |
| The Normative Beliefs About Aggression Scale | 74 |
| Behavioral Scale-Relational Aggression | 76 |
| The Peer-Nomination Scales for Children | 78 |
| Consumer Satisfaction Survey-Student | 80 |
| Implementation Checklist | 81 |
| Independent Measure | 82 |
| Instructional Process | 82 |
| Instructional Content | 84 |
| Data Analysis | 87 |
| Chapter 4: Results | 91 |
| Overview | 91 |
| Treatment of the data | 91 |
| Measures of central tendency | 92 |
| Data quality checks-Cronbach's alpha | 101 |
| AQTotal | 110 |
| Rbttotal | 111 |
| Nbttotal | 113 |
| PNPR+ | 114 |
| PNPR- | 115 |
| PNVIC | 116 |
| Chapter 5: Discussion | 126 |

| | |
|---|-----|
| Summary of Study | 126 |
| Examination of Results | 126 |
| Notable findings from the aggression measures | 127 |
| Implementation integrity | 129 |
| Effectiveness of the intervention | 130 |
| Implications of the Results | 133 |
| Supportive and shared leadership | 134 |
| Roles and responsibilities | 134 |
| Commitment of key personnel | 135 |
| Limitations | 138 |
| Suggestions for Future Research | 139 |
| Final Thoughts | 141 |
| References | 143 |
| Appendices | 154 |
| Appendix A: Passive Parent Consent Form | 155 |
| Appendix B: The Aggression Questionnaire | 157 |
| Appendix C: The NOBAGS | 158 |
| Appendix D: Behavior Scale-Relational Aggression | 161 |
| Appendix E: The Peer-Nomination Scales for Children | 163 |
| Appendix F: Consumer Satisfaction Survey-Student | 164 |
| Appendix G: Implementation Checklist | 166 |
| Appendix H: Exit Interview | 168 |
| Appendix I: Observer Review | 169 |
| Appendix J: Project Timeline | 171 |
| Appendix K: Child Assent Form | 173 |
| Appendix L: Control School Lesson Plan Information | 174 |
| Appendix M: Correlations of Measures | 175 |

LIST OF TABLES

| | | |
|----------|---|-----|
| Table 1 | Proportions of verbal, physical, and indirect aggression | 13 |
| Table 2 | Reliabilities of NOBAGS and subscales | 75 |
| Table 3 | Internal consistencies of an inner city sample | 80 |
| Table 4 | Descriptive statistics of all participants on the pre-/post-test | 93 |
| Table 5 | Descriptive statistics for complete data subset on the pre-/post-test | 94 |
| Table 6 | Distribution of scores for complete data subset on pre-/post-test | 95 |
| Table 7 | Effect sizes for female and male participants | 96 |
| Table 8 | Distribution of scores for all participants on pre-/post-test | 98 |
| Table 9 | Distribution of scores for complete data subset on pre-/post-test | 98 |
| Table 10 | Item Total Statistics | 102 |
| Table 11 | Correlations between measures for pre- and post-test | 110 |
| Table 12 | ANOVA summary table for AQTtotal | 111 |
| Table 13 | ANOVA summary table for Rbttotal | 112 |
| Table 14 | ANOVA summary table for Nbttotal | 113 |
| Table 15 | ANOVA summary table for PNPR+ | 115 |
| Table 16 | ANOVA summary table for PNPR- | 116 |
| Table 17 | ANOVA summary table for PNVIC | 117 |

| | | |
|----------|---------------------------------------|-----|
| Table 18 | Frequency distribution-Interest | 121 |
| Table 19 | Frequency distribution-New Learning | 122 |
| Table 20 | Frequency distribution-Recommendation | 122 |

LIST OF FIGURES

| | | |
|------------|---|-----|
| Figure 1. | Stem and leaf representation of effect size | 96 |
| Figure 2. | Box and whisker plot for pre-test AQ scales | 99 |
| Figure 3. | Box and whisker plot for post-test AQ scales | 99 |
| Figure 4. | Box and whisker plot for pre-test MacDonald et al., 2000 | 100 |
| Figure 5. | Box and whisker plot for post-test MacDonald et al., 2000 | 100 |
| Figure 6. | Box and whisker plot for pre-test NOBAGS scales | 100 |
| Figure 7. | Box and whisker plot for post-test NOBAGS scales | 100 |
| Figure 8. | Box and whisker plot for pre-test Peer Nomination scale | 100 |
| Figure 9. | Box and whisker plot for post-test Peer Nomination scale | 100 |
| Figure 10. | Ordinal interaction for School 2 on Nbttotal | 114 |
| Figure 11. | Ordinal interaction for School 1 on Nbttotal | 114 |
| Figure 12. | Ordinal interaction for School 3 on Nbttotal | 114 |
| Figure 13. | Ordinal interaction on PNVIC by gender and time | 118 |
| Figure 14. | History of consumer satisfaction survey-Interest | 121 |
| Figure 15. | History of consumer satisfaction survey-New Learning | 122 |
| Figure 16. | History of consumer satisfaction survey-Recommendation | 122 |

Helping Break the Cycle of School Violence and Aggression: A Program Evaluation for the
Owning Up Curricula

Stephanie Mihalas

ABSTRACT

This study explored the effects of a gender-based violence prevention program called *Owning Up* on a group of middle school students in Tampa, Florida. The majority of participants were African-American at-risk youth who voluntarily participated in the intervention. This study is novel in that it is one of the first to empirically validate curricula that addresses an array of aggressive behaviors. A mixed methods design was used to gain insight into significant changes over time, as well as treatment integrity throughout the implementation. Findings from the study found females to be more aggressive than males on all forms of aggression across time. Additionally, results suggest that minimal gains were found between pre- and post-test administrations indicating that the intervention with this particular group may have not been effective. Cultural fit, systems-level issues, and problem-solving tactics are discussed to explain the findings. Implications for the findings for future users of this program are addressed.

Chapter 1

Introduction

A popular focus of research in the social sciences over the past several decades has been on understanding the phenomenon of aggression. Aggression is known to have a substantial negative impact on all individuals involved and therefore research has primarily focused on risk and resiliency factors, assessment tools, adjustment trajectories and effective interventions.

Although long recognized that aggression may take various forms, the majority of research to date has focused on physical and verbal forms of aggression because of their overt nature. As a result, aggression research has inadvertently taken a gender biased stance, based on the fact that boys engage in more physical and verbal aggression. However, current research has highlighted the fact that when the definitional domain of aggression is expanded to include social ostracism and rumors, girls equally aggressive as boys. The systematic study of relational aggression, a form of aggression that damages relationships through manipulation, has expanded and refocused research efforts to take a broader perspective when identifying gender and types of behavior that have negative outcomes on youth.

Brief review of relational aggression literature

The majority of research to date has primarily focused on developing and determining valid and reliable assessment tools to identify prevalence rates by gender (e.g., Crick, Casas, & Mosher, 1997; Henington, Hughes, Cavell, & Thompson, 1998). This is probably a result of trying to validate that, in fact, relational aggression is a predominant problem in our schools and communities. Researchers have used a variety

of tools including teacher, peer, and self-reports. Although each kind of assessment tool has been validated empirically, a trend has developed towards utilizing a multi-informant, multi-trait approach as a means to counteract any kind of bias by either peers, teachers, or self perceptions (McEvoy, Estrem, Rodriguez, & Olson, 2003).

Most of the assessment tools have been developed for specific age groups and are usually designed for either preschool-aged children (Crick et al., 1997), elementary-aged children (Henington et al., 1998), or college-aged students (Loudin, Loukas, & Robinson, 2003). Limitations of the research so far are a lack of tools designed specifically for middle-school populations. Findings indicate that relational aggression is evident as early as preschool-aged children and continues throughout development into college. Although clearly researchers have identified that relational aggression is prevalent in a wide range of age groups, there is a lack of agreement relative to which gender exhibits more relational aggression. Further research needs to address this by determining why individual studies are unable to replicate prevalence rate findings.

Although assessment is important in its own right, it is not a stand alone answer. Once a problem has been identified, something must be done in order to stymie and/or prevent further escalation. Therefore, design and implementation of interventions is a critical component in relational aggression research. However, to date, there are only three programs available to intervene and provide students and schools with empowering tools to overcome aspects of relational aggression. This is an area that needs further development in order to help students overcome barriers and obstacles in their development.

Significance

One may question the significance of studying a construct like relational aggression because of the longstanding societal norms placed on women's behavior. Many people are quick to write off relationally aggressive acts as fairly normative behavior of many middle school and adolescent girls. The reality is however, that relational aggression is not a behavior that should be condoned based on this unfortunate perception that many hold. In fact, relational aggression affects both boys and girls, although possibly to differing degrees and also begins as early as the preschool years. Therefore, because of its expansive nature across gender and age, it is absolutely necessary to conduct research that not only will appropriately identify students that are aggressors but also students that are victims. Current research highlights that each person that is involved either as victim or perpetrator has potentially serious future social maladjustment issues including loneliness, depression, and both internalizing and externalizing behavior (Crick et al., 2001). Based on the significance of predicted outcomes of victims and aggressors, intervention efforts must be constructed and must be empirically validated to provide students with the opportunity to mediate their cognitions and behavior, as it relates to relationally aggressive acts. If this is not accomplished, then we are not affording today's youth with positive mental outcomes that they deserve for their future.

Rationale for Current Study

As has been delineated briefly so far, relational aggression is a serious issue that affects children's social interactions in school and life. In order for schools to serve as institutions that provide children with a myriad of opportunities, it is clear that they must

engage students in programming that deals with everyday issues that children face. Ironically, there are very few programs that actually help children understand violence and aggression and promote empowerment through teaching students valuable tools for coping and adapting. Additionally, programs that are designed to deal with these kinds of serious issues are rarely empirically validated to determine their overall effectiveness. In fact, schools and non-profit organizations around the country are implementing programs with no determined real link addressed between curricula and positive benefits to youth. This is a serious problem because the effects of the programs are unknown, and therefore schools and organizations may a) spend money on ineffective programs, and b) indirectly produce negative outcomes for students. Therefore, the current study was warranted because it was another step in determining if a widely-used program was actually a successful intervention for students. More importantly, a program evaluation afforded fellow researchers and potential funders of the program to make informed decisions prior to assuming the program is effective or ineffective.

Purpose of the Current Study

In this study, the researcher examined a program called *Owning Up*. The *Owning Up* curricula is designed for both boys and girls and deals with a variety of issues including sexual assault, harassment, gender-based norms, and social and relational aggression. The short-term goals of the evaluation were to a) determine the general effectiveness of the implementation process, and b) to determine if the curriculum had positive outcome effects on participants. The long-term goal was to potentially expand the curricula across schools within Hillsborough County, Florida, such that the program

would be a county-wide and school-wide initiative to help raise awareness of violence and aggression issues.

Summary and Description of the Current Study

The current study took place in three pilot middle schools in Tampa, Florida. Participants in the study were chosen from a county-wide program called “Program to Empower Teens” which was developed for students who are designated as “at-risk” because of problems with academic performance, school behavior, and social skills. Two of the schools served as the experimental group, and the third school acted as a treatment control. All participants in the three schools completed four measures at pre- and post-test. A multi-informant approach of both peer and self-reports was utilized. The following measures were used in this study and have been validated empirically for reliability, consistency, and/or validity: The Aggression Questionnaire (Buss & Warren, 2000), The Normative Beliefs about Aggression Scale (Huessman, & Guerra, 1997), Relational Aggression Scale-unknown name (MacDonald, D’Amico, & O’Laughlin, 2000), The Peer-Nomination Scales for Children (Huessman, 2002). Additionally, in order to make a reasonably firm conclusion about the reliability of the outcome data, the researcher assessed the likelihood that the implementation process confounded the outcome results. In order to measure the integrity of the implementation process, the researcher conducted exit interviews with the facilitators of *Owning Up*, a consumer satisfaction survey for participants, a weekly intervention checklist, and random observations were made in the experimental conditions. These measures were designed by the researcher and served to highlight any process problems that needed remediation in this study or for future implementation efforts. Additionally, when issues did arise, the

researcher was able to make qualitative inferences about the data. The intervention period was intended to occur for a period of eight consecutive weeks in 50-minute intervals. Each week a new topic was broached and facilitators engaged students in discussion, journal reflections, and role-play activities. Facilitators bridged the material each week by reflecting on thoughts and feelings from the previous week and attempted to highlight critical features of past lessons and how they related to new lessons each week.

Specifically, this study addressed the following questions:

1. How well was the *Owning Up* curricula implemented?
 - a. Were there any major issues that could have impacted outcomes at pre- and post-test measures?
 - b. Did implementation differ across schools? Across facilitators?
2. What level of consistency did the multi-informant approach provide for the evaluation (i.e., what was the relationship between measures)?
3. Was there a significant change between pre- and post-test scores for the intervention group, relative to the control group?
 - a. Assuming the change was significant, what was the magnitude of the change?
4. What was the relationship between how each facilitator viewed the effectiveness of the program and the results from the outcome measures?
 - a. How did students view the program relative to how the facilitator viewed the utility of the program?
 - b. Was the outcome data consistent with facilitator and student reports?

Chapter 2

Review of the Literature

Over the past 30 years, research related to aggression has been directed at understanding developmental trajectories, functions of behavior, and outcome expectancies. Aggression, an act that is intended to injure another, has been studied extensively because of its all encompassing nature: it affects both girls and boys academically, socially, and psychologically as well as predicts future antisocial outcomes. A current resurgence of the study of aggression has been underway due to the increase in violence on school campuses across the U.S. One of the first studies to address bullying and violence found that of the 15,638 children who were polled, 36.4% carried weapons in response to verbal/physical attacks by peers within school. In addition, students who are bullied are 1.5 times more likely than those who are not to carry a weapon and are 1.7 times more likely to be involved in physical aggression (Viadero, 2003b).

In recent years, the severity and impact of various forms of aggression have begun to be identified. These findings indicate the need to facilitate programs to curb aggressive behavior in order to maintain safety and positive mental health outcomes for youth (Viadero, 2003a). A fundamental shift is occurring in the way that educators and researchers define and assess aggression. No longer is the focus solely on physical aggression. Rather, it has expanded to indirect and social forms that also have been found to predict risky and unhealthy behavior. School-wide efforts around the country are a by-product of the concern of various forms of aggression and their relative impact on youth and provide solid evidence for the need to further understand various forms of aggression

and how educators and school-related personnel may help to initiate programs to cope with its effects (Toppo, 2003).

This literature review intends to serve as a contextual foundation to understand how destructive and harmful aggressive behavior is. More importantly, the seriousness of the aggressive behaviors reviewed serve to instill a sense of urgency in engaging and implementing effective and appropriate assessments and interventions for victims and perpetrators. This will be accomplished by exploring variables related to aggression, specifically by: a) briefly defining and discussing the different forms of aggression including physical, social, and relational aggression, b) discussing the risk factors that promote aggressive behavior, c) understanding maladjustment problems for both the victim and aggressor, d) evaluating and understanding how relational aggression is assessed and its implications on prevalence rates, e) exploring intervention and prevention initiatives, f) discussing limitations and gaps in this literature base that will indicate a need for further research, and g) explaining necessity of current study and implications of evaluation.

Forms of Aggression and Associated Prevalence Rates

More contemporary research has determined the extent to which gender differences in aggressive behavior depend on the defining principles and characteristics of aggression. The standard approach of assuming that aggression is a derivative of modeling and imitation has been undermined by theories that entertain the role of social-reasoning, moral development, and social-cognition. The different theories have circuitously divided aggression into two predominantly different categories: direct and indirect. Direct aggression is behavior that is intended to injure another by hurting

someone in an overt manner, while indirect aggression is an attempt to hurt another without face-to-face conflict. For the purposes of this literature review, direct and physical aggression will be used interchangeably. Indirect aggression will be discussed as separate constructs: social and relational.

Physical aggression. The term “physical aggression” is almost synonymous with the word “boys.” Considerable attention has focused on males as being both the perpetrators and victims of aggression (Campbell & Frabutt, 1999). Historically, a large emphasis has been placed on the link between physical aggression and gender because the behavior is overt and can be observed easily by outsiders. Physical aggression, in theory, is based on different models, relying heavily on the frustration-aggression model proposed by Berkowitz (1989), which states that aggression is an inherently angry and defensive response to provocation, threat, and frustration. The question then arises to why it appears that boys, more so than girls, are apt to respond or precipitate physical aggression.

One popular rationale to account for the gender differences are social representations that men and women internalize when evaluating a physically aggressive act. Women associate negativity with physical aggression because it represents a loss of self-control (known as expressive representation). Conversely, men view aggression as a positive act of exerting control over others (known as instrumental representation). These differences in representations have been found in children as young as 7 to 8 years of age (Archer & Haigh, 1994).

Tapper and Boulton (2000) attempted to replicate findings from Archer and Haigh (1994) for three main reasons: a) the assessment tool may have not been developmentally

appropriate for younger children; b) the questions in the assessment tool were ambiguous and may have not detected direct and indirect aggression as separate constructs and c) to examine sex and age differences detected by the assessment tool. To accomplish this, 130 children from the United Kingdom were chosen to participate. Fifty seven of the children (females, $n = 29$; males, $n = 28$) were in year three of grade school (aged 7-8 years) and 73 (females, $n=37$; males, $n=36$) were in year six of grade school (aged 10-11 years). Modifications of the EXPAGG, a common tool utilized to assess physical aggression and its links to emotions, cognitions, and social values were used in this study. Specifically, three different EXPAGG-type questionnaires were administered to all participants; each one focused specifically on physical, verbal, or indirect aggression. The items on the newly designed EXPAGG questionnaires showed strong internal reliability: reliability coefficients of .85 for the physical EXPAGG, .83 for the indirect EXPAGG, and .84 for the verbal EXPAGG. The results suggest that the new tools are more developmentally appropriate for younger children. Additionally results showed a main effect size for sex, $F(3,124) = 3.14, p < .03$) as a result of internalized representations: girls showed more expressive social representations of indirect aggression and direct aggression.

Overall, results from Tapper and Boulton (2000) highlight an important point: social representations permeate across different domains of aggression. By extending the concept of social representations to other forms of aggression like relational and social, researchers may be able to explain and account for differing perspectives on prevalence rates between genders. Furthermore, targeting children early with intervention programs

may serve as a valuable tool to intercede in the formation and/or transmission of representational models of sex-linked aggressive behavior.

Gender differences in physical aggression also have been explained according to conventional rule models (moral, conventional, personal). The nature to which the rules are embedded within one's own personal cognitive network attempts to account for gender differences within physical aggression. According to Turiel (1983), moral rule models may comprise aggressive behavior because it is directly related to the rights and welfare of others; for example, those that do not value morality may be more likely to engage in aggressive acts. Crane-Ross, Tisak, and Tisak (1998) investigated the role of social domain theory further by testing three concepts in relation to aggression: a) beliefs about legitimacy, b) value placed on consequences, and c) expectations of negative effects on others. Participants included 398 adolescents ranging in age from 14-19 years old from primarily Caucasian, middle-class backgrounds. Results were based on peer and self-reported questionnaires using a Likert-type scale ranging from 1 (acceptable) – 4 (not acceptable). Results on legitimacy via self-reports concluded that females were less likely than males to consider aggression and violations to be acceptable (female, $M = 3.04$; male, $M = 2.49$). Females also were more concerned than males about negative self-evaluations (female, $M = 3.06$; male, $M = 2.58$) and placed less value on expected rewards (female, $M = 1.27$; male, $M = 1.48$). A synthesis of the results confirmed that overall males were more likely to violate norms and consequently engage in physically aggressive behavior as a by-product of the beliefs and values placed on the aggressive act. Essentially, the kinds of rules held by males predicted behavioral outcomes.

The variance in adolescents' reasoning and perceptions of moral transgression was not addressed in this study. Future research may want to assess a child's meta-cognitive awareness of contextual factors that influence the usage of rules. This may play an important role in developing intervention plans for physically aggressive children. In addition, research also may investigate the effects of various personal, developmental, and/or cultural factors to determine if adolescents' or preschoolers' beliefs on conventional and aggressive domains play a factor in engagement and prevalence.

Indirect aggression. The general theme that permeates physical aggression research is the identification of aggressive boy populations. Recently parent, teacher, and mental health professionals have recognized that girls engage in aggression, although it may not be presented in physical form. Extensive cross-cultural studies, initiated by Osterman and colleagues (1998) indicate that aggressive behavior is present across different age groups and genders. Statistical outcomes from their study are presented in Table 1. The outcomes in Table 1 identify the percentage of total aggression within the sample across gender and age. Findings indicate that girls significantly engage in more indirect aggression than boys (Osterman et al., 1998).

Table 1

Proportions of verbal, physical and indirect aggression within a sample size of 1025 girls and 1069 boys by age (Osterman et al., 1998)

| Girls: 8 years | Percentage of total aggression | Boys: 8 years | Percentage of Total aggression |
|-----------------|--------------------------------|----------------|--------------------------------|
| Verbal | 31% | Verbal | 37% |
| Physical | 14% | Physical | 37% |
| Indirect | 55% | Indirect | 26% |
| Girls: 11 years | | Boys: 11 years | |
| Verbal | 39% | Verbal | 39% |
| Physical | 20% | Physical | 38% |
| Indirect | 41% | Indirect | 23% |
| Girls: 15 years | | Boys: 15 years | |
| Verbal | 40% | Verbal | 47% |
| Physical | 8% | Physical | 33% |
| Indirect | 52% | Indirect | 20% |

A combination of high prevalence rates and detrimental effects of indirect aggression on adolescent and young girls have instilled a sense of urgency to delve more deeply into an area that has been understudied for many years (e.g., Connor et al., 2003; Galen & Underwood, 1997; Henington et al., 1998). Although the literature on indirect aggression is not as exhaustive as physical aggression, consensus findings illustrate two major points regarding aggression: 1) it is no longer exclusively related to boys and 2) physical, social, and relational aggression impact the mental health status of children and must be addressed to improve outcomes for youth.

Social aggression. One of the primary areas of research on indirect aggression is social aggression a term coined by Marion Underwood. Social aggression is a subtle behavior that is intended to damage another's self-esteem, social status or both, and may take two forms: direct and indirect. The direct form may include verbal rejection or negative facial expressions, and the indirect form may include social exclusion or rumors. In general, social aggression refers to an ongoing social interaction that degrades another's self-concept and/or social standing (Galen & Underwood, 1997). Social aggression is more prevalent in girls than boys; the reasons for this are unknown. Some theorists believe that girls view the functional aspect of verbal aggression as a form of gaining social status (Rauste-von Wright, 1989), whereas others support the notion that the nature of female relationships and bonding permits levels of self-disclosure and exclusion from cliques (Salmivalli, Kaukiainen, & Lagerspetz, 2002). Regardless, a key difference between physical and social aggression is the dyadic versus triadic relationship that must exist prior for an act to occur. Therefore, a reliance on social networks is an essential component for the employment of social aggression. Without a connection to a network, the point of a social attack is moot. Conversely, if someone is deeply entrenched within a network, the more effective and self-serving an attack may be (Xie, Cairns, & Cairns, 2002).

Xie et al. (2002) highlights an important difference between social and physical aggression: the relative trajectory of risky behavior in other areas of development. While physical aggression predicts heightened rates of school dropout and poor academic performance, social aggression may have the opposite effects; for example, teacher ratings of academic achievers are higher, children exhibit higher social intelligence

scores, and also higher levels of social network centrality (i.e., popularity). These findings have not been confirmed by other studies, and therefore replication must ensue before generalizations about developmental trajectories can be safely made.

Limitations in social aggression research include predictive factors, developmental trajectories, and the types of children that engage in social aggression. These three factors may be beneficial to assess, in order for effective intervention construction to be made possible. For further information on social aggression, refer to studies by Xie et al., (2002) and Galen and Underwood (1997).

Relational aggression. Relational aggression (RA) is an indirect and manipulative form of aggression that intends to harm others through damage to peer relationships in a way that blocks the social goals of the target peer. Unlike physical aggression which harms others through hitting or threat of physical harm, RA interferes with friendships and threatens exclusion from peer groups (Henington et al., 1998). Acts of RA may be verbal or nonverbal and may include spreading rumors as a form of retaliation, excluding others from play or encouraging others to exclude peers, in addition to social exclusion through gossip.

RA has serious implications for social and emotional maladjustment in both boys and girls across a wide range of age groups, beginning in preschool (Crick et al., 1997) and extending even into college (Loudin et al., 2003). Therefore, the latest aggression research has primarily focused on RA. The focus of research has mainly revolved around determination of the prevalence rates by gender. Additionally, a new body of literature on RA is attempting to identify concurrent and future predictors of maladjustment as well as effects on relationships with peers. Based on the significant and increasing trend of

identifying children who engage in RA (Osterman et al., 1998), it seems appropriate to review the literature to develop a comprehensive understanding of specific attributes of RA, and what, if any, gaps appear in the literature that may serve as a basis for further exploration into an area that deserves much needed attention. Additionally based on the social-maladjustment issues that take place for victims of RA, intervention exploration is necessary in order to provide aid to those in need.

Risk Factors

Physical aggression. The term “risk,” as it is related to aggression research denotes a behavior that a child displays, a condition that a child has been exposed to, or an event that will increase the likelihood of engagement in aggressive behavior at some point during development (Van Acker & Talbott, 1999). Although a causal relationship between risk factors and outcomes remains unclear, it is established that no single predictor variable for aggressive behavior exists; instead, there are multiple pathways and antecedents that intervene in producing an aggressive act (Domitrovich & Bierman, 2001). Influential risk variables mediating aggressive behavior can be categorized into individual, family, and peer influences.

At the individual level, genetic predisposition to physical health problems (i.e., asthma) and psychological problems (i.e., psychotic disorder, hostile temperament, predisposition for substance abuse) have been identified as components that contribute to aggressive behavior (Connor, Steingard, Anderson, & Melloni, 2003). Additionally, Connor (2002) found that global IQ deficits play a significant factor in antisocial youth behavior. IQ scores that are approximately 1.5 standard deviations below the norm have consistently been documented as predictive variables for physical aggression. Equally

important to the development and/or reinforcement of aggressive behavior is the impact of parenting practices and the nature of the home environment. Permissive and authoritarian styles of parenting have been shown to promote differences in levels of aggressive behavior (Rubin, Hastings, Chen, Stewart, & McNichol, 1998). Werner and Silbereisen (2003) reiterate this point by highlighting the importance of cohesion and closeness with parents as variables that are relative to delinquent behavior. Aside from direct interactions between parent and child, behavior modeled in the home environment has also been indicated as a primary means of transmission and reinforcement of aggression (Bandura, Ross, & Ross, 1961). Finally, peers play an integral part in influencing aggressive behavior (Watson, 2002). Specifically, children that are bullied (Olweus, 1997) and/or associate with aggressive children (Rodkin & Hodges, 2003) are at higher risk to engage in aggressive acts. The research on risk factors for physical aggression is hardly definitive; however, the findings of the aforementioned studies provide valuable insight into possible predictor variables for aggressive behavior. The focus on risk factors related to overt/physical forms of aggression may not generalize to other forms of aggression, and therefore must be viewed with caution when making inferences to other populations of aggressors.

However, it can be inferred that a transactional model is the most appropriate way to describe the relationship between risk factors and aggression. Coping and resilience mechanisms serve as evidence for the transactional approach because they may be able to intervene and alter risk variables (Phelps Roecker, 2001). Viewing physical aggression in this light supports the notion that interventions are possible for aggressive children if they are taught effective ways of coping.

Social and relational aggression. Thus far, examination of risk factors has been limited to physical forms of aggression. Therefore, conclusive comments on indirect aggression cannot be made at this time. Further research efforts should be directed at addressing indirect aggression and risk variables.

Concurrent and Future Social Adjustment Issues

Implications of RA for perpetrators and victims. Physical aggression research has shown that the effects are profound on those that engage in the behavior. Research within the RA domain also has a bounty of studies that have examined the implications of RA for both perpetrators and victims. Victimized children may develop derogatory attitudes toward peers, self-control problems, and/or experience unwarranted negative self-appraisals (Prinstein, Boergers, & Vernberg, 2001). Perpetrators have social problems because they lack friendships of depth that tend to serve as buffers to social maladjustment (Grotmeter & Crick, 1996). For the purpose of this section, a clear dichotomy is drawn between perpetrator and victim. However, the line is much finer in actuality—a cycle may ensue in which one that is initially in the role of a victim may transition into the status of perpetrator (Rodkin & Hodges, 2003) such that research oftentimes identifies a false dichotomy of victim and perpetrator. The widespread nature of adjustment issues makes this aspect of the literature relevant for intervention efforts. Each of the studies in the following sections of the review addresses the impact of RA on those who perpetrate it (e.g., aggressors) and those who receive it (i.e., victims).

Consequences of RA for victims. The dyadic nature of RA has provided a forum for researchers to assess the quality of friendships between victims and perpetrators. Grotmeter and Crick (1996) evaluated various aspects of a friendship in 12-year old girls

($N=315$), including intimacy levels and their perceived importance to the relationship. Self-reports including the Friendship Quality Measure (FQM) and the Importance of Friendship Quality Measures (IFQM), were used to assess components of intimacy levels: exclusivity, intimate exchange, and conflict. Univariate tests indicate that involvement with a child labeled “relationally aggressive” tends to predict more assertion of aggression within the dyad as compared to ($M = 1.93, SD = .83$) children labeled as non aggressive ($M = 1.59, SD = .76$). In addition, relationally aggressive girls report engaging in highly intimate and exclusive friendships; however, they tend to self-disclose less in the friend dyad. This is especially important because the information that a friend may disclose to an aggressor may be used against the friend in manipulation tactics and other relationally-related activities. Basically, the role that many non-aggressive children take in these dyadic relationships, adds fuel to the fire for the relationally aggressive child.

Although Grotper and Crick (1996) confirmed that the way relationally aggressive girls approach friendships is maladaptive for non-aggressive friends, information on the impact on the non-aggressor is limited. In addition, specific social adjustment difficulties were not directly addressed but alluded to as arising out of the aggressor/non-aggressor dyad. Future research may investigate how levels of satisfaction within a relationship impact social adjustment issues for the aggressor and non-aggressor.

Additionally, although Grotper and Crick (1996) did not directly address social adjustment problems for victims, other studies have pointed to victimization as being strongly linked to significant problems with both internalizing and externalizing behavior (Crick et al., 1997). This is largely in part due to internal representations held by children

and adolescents that RA is a mean and hostile act. The question that remains is why certain children place themselves in an aggressive situation knowing that it is not beneficial.

In response to this question, a study was designed to assess gender differences in peer maltreatment and the potential implications of victimization (Crick & Nelson, 2002). A multiple informants' evaluation, including the FQM, The Franke and Hymel Social Anxiety Scale, and self-reports were used for 496 children in grades three through six (grade three, $n = 114$; grade four, $n = 81$; grade five, $n = 148$; grade six, $n = 153$). A 2 x 2 ANOVA reported a significant main effect of friend victimization type, ($F(1, 306) = 4.82, p < .001$), indicating that children reported higher levels of friend relational victimization than friend physical victimization across all grades. Additionally, boys reported more physical friend victimization (11.3%) compared to girls (1.8%). However, girls reported more friend relational victimization (12.0%) compared to boys (5.6%). In addition, hypotheses were confirmed regarding victims within a friend dyad. They had more social difficulties and internalizing/externalizing problems specifically related to factors like loneliness ($F(2, 127) = 14.1, p < .001$) and psychological distress ($F(2, 127) = 6.4, p < .05$). If RA had not been addressed as a separate construct from physical aggression, it would account for failure to identify over 70% of victims. This statistic alone lends itself to the importance of evaluating RA as a separate construct.

There are two major limitations with the Crick and Nelson (2002) study related to procedures enacted for child participants. The first, is the long (two different, 60-minute sessions) time period in which children were directed to complete their assessments. It can be inferred that the time span of attention for younger children may be limited and

they may not be equipped to focus on assessment for an entire hour; for that reason, the results from their ratings may contain more error than the researchers intended. Future research may wish to replicate this study reducing the amount of time allotted for assessment so that this potential confound may be eliminated. In addition, each item on the measure was read aloud. Assuming the child was off-task or missed the questions, he/she could again have easily scored in a manner that did not accurately depict their beliefs.

A problematic feature of the victim-aggressor relationship is the stability with which victims remain in the harmful dyad relationship. As a result, the victim has an inability to gain valuable social skills and buffer support from other friends (Crick, 1996). In addition, speculations have been made that the dysfunctional dyad may serve as a prototype for other abusive relationships. Specifically, that the role the victim plays in the dyad may transfer to future abusive relationships (Coie & Dodge, 1983). The extent to which children remain resilient in these situations needs to be addressed in order to discover ways to help victims end the cycle of victimization. The nature of adolescent populations, specifically a heavy focus on social comparisons warrants future research. One could hypothesize that the rates would be even more alarming as they are in Crick and Nelson (2002) because of the adolescent's innate desire to "fit in" with peers.

Since future outcomes are significant for victims, it is extremely important for research to try to provide practical information that may be used to design early identification programs and interventions. Therefore, researchers have attended to younger children, specifically preschool populations. Crick, Casas, and Ku (1999) addressed RA and victimization in a preschool population of children (boys, $n = 67$; girls,

$n = 62$) ranging in age from three to five. The PSBS-T, PSBS-P, self-reports and Asher's (1992) Loneliness scale were used to assess aggression and social maladjustment.

Results indicated that girls were significantly more victimized than boys (girls, $M = 4.96$; boys, $M = 4.68$). This is an important finding because although there is a discrepancy in the literature regarding who is more aggressive by nature, the findings for victims of RA are consistent across studies (e.g., Crick 1996; Tomada & Schneider, 1997; Prinstein et al., 2001). Findings from Crick, Casas, and Ku (1999) are similar to Grotperter and Crick (1997), suggesting that both older and younger victims experience social-psychological adjustment problems.

According to peer reports, victimized peers showed lower levels of acceptance from other children ($F(1, 114) = 6.02$) and a main effect revealed that victimized boys ($M = 3.41$) used significantly less prosocial behavior than non-victimized boys ($M = .51$). In contrast, girl subgroups showed no difference in prosocial behavior scores. According to Coie and Dodge (1998), children's perceptions of the self as a victim might not evolve until age four or five. This may explain the smaller effect sizes—as compared to other peer victimization studies—because part of the sample population may lack the understanding on what the “victim” role actually is. This represents a potential limitation of the study and replication may provide higher effect sizes if the three-year-old population is eliminated. Overall, this study adds further support to the relative nature of negative effects of victims of RA now evident in age groups from 4-12. Based on the stability of aggression (Tomada & Schneider, 1997), the early identification and promotion of intervention strategies may help to deter negative outcomes. Future research may address this potential. Further investigations may attempt to use a direct

observation method in addition to peer and self-reports in order to provide a naturalistic account versus a contrived account of younger peer actions.

To complete the picture on victimization across the development, another group of researchers attempted to assess the relationship of RA and social-psychological adjustment in adolescents—from the standpoint of both aggressors and victims. Prinstein et al. (2001) attempted to answer questions in four related areas: 1) RA's association with depressive symptoms, loneliness, externalizing behavior and self-esteem issues; 2) the dependency of externalizing/internalizing behaviors on type of aggression (covert/overt); 3) impact of co-occurrences of multiple forms of aggression; 4) social support networks as potential buffers to victimization. The sample population was large, consisting of 566 adolescents in grades 9-12, who were assessed with more instruments than previous literature in the area of victimization and social maladjustment: Peer Experiences Questionnaire, the Center for Epidemiological Studies-Depression (CES-D), UCLA Loneliness Scale, Harter's Self Perception Profile for Adolescents (SPAA), Diagnostic Interview Schedule for Children (DISC), and the Close Friend subscale. The study confirmed findings from prior studies that RA is correlated with girls' internalizing behavior and not externalizing behavior, and conversely that externalizing behavior was linked to overt aggression in both victims and perpetrators. Specifically, the relative contributions of RA on concurrent social-psychological adjustment for victims included: loneliness ($r = .40$), depressive symptoms ($r = .29$), and self-esteem ($r = -.31$). On the contrary, concurrent maladjustment findings were not evident in the aggressor population, aside from externalizing symptoms ($r = .39$). As predicted, the most severely maladjusted child suffered from dual forms of victimization and obtained higher levels on

all the previous variables mentioned. Finally, findings did indicate that having a supportive friend buffers children from experiencing high levels of distress ($r = -.49$, $p < .01$), specifically OCD/CD symptoms.

Therefore, Prinstein et al. (2001) claimed that it would be in an adolescent's best interest to seek out supportive friendships; however, this may not be feasible based on the internalized behaviors and patterns of victims in destructive relationships (Crick & Nelson, 2002). The fact that these investigators used dissimilar tools of assessment compared to the majority of research conducted for RA, may either provide new insight into ways of assessing, or may be inappropriate. Replications of the study in different populations may be able to determine the appropriateness and viability of the measures.

Implications for perpetrators. Social maladjustment research for perpetrators of RA has been limited to issues of non-normative gender role in aggression choice (i.e., boys exhibiting RA and girls exhibiting physical aggression). Recent literature has argued that engaging in non-normative forms of aggression may be just as likely, or even more likely to promote risky behavior and subsequent social-psychological problems. Aggression form, severity, and type of social-psychological maladjustment was examined across gender and type of aggression in a sample of 1,166, 3rd-6th graders in order to make reasonable conclusions to the nature of non-normative displays of aggression (Crick, 1997). Findings confirmed that overtly aggressive children display more externalizing behavior than their relationally aggressive and nonaggressive peers. However, contrary to other findings (Prinstein et al., 2001), relationally aggressive children also displayed significantly more externalizing behavior than their nonaggressive peers. As predicted, non-normative behavior is related to more significant

maladjustment problems, which may in part be due to the gender-stereotyped sanctions imposed on aggressors by peers, parents, and teachers.

One limitation of this study was reliance solely on teacher reports for identification of adjustment problems; oftentimes teachers may have biased accounts on what occurs in the classroom based on differential expectations and biases (Puig et al., 2003). In addition, parent reports may provide a different perspective than teacher reports because behaviors displayed at home may be different than those in school. This study has important implications for targeting children that are not usually the focus of intervention efforts because of the assumed relationship between gender and aggression type.

Concurrent social-psychological adjustment problems are not the only issues that victims and aggressors face. Based on the stable nature of aggression (Tomada & Schneider, 1997), future social adjustment also is a critical component in this topic area. According to Parker and Asher (1987), covert childhood aggression is one of the best predictive factors for adjustment problems later in life. However, very few studies have determined what the long-term effects are for relationally aggressive children. However, to date there has been one study that has addressed the role of overt and relational aggression as predictors of children's future adjustment in 3rd-6th graders (Crick, 1996). Based on teacher reports for boys, relational and overt aggression were positively correlated with future peer rejection ($r = .19$, $r = .25$, $p < .001$, respectively). The same results were found for girls, except that girls' correlations between aggression and future rejection were higher (overt, $r = .39$; relational, $r = .24$). The study found strong correlations and strong inter-rater measures of reliability between teacher and peer

assessments of behavior. However, the dependent variables being addressed (peer rejection) appear to be limited in nature. Peer rejection and peer acceptance do not provide an exhaustive portrayal of variables that account for future maladjustment issues. Therefore, it seems reasonable that future longitudinal studies should focus on other factors that can be addressed like depression, social anxiety, and phobias, which are also known to be linked to maladjustment problems in youth and adulthood.

Summary. Clearly, the effects on victims of relational and overt aggressive relationships are pronounced. Consistently victims, more so than aggressors, are found to have a fair number of concurrent social-psychological problems including depression and loneliness. It still is not clear whether there is a pathway link between antecedent psychological problems and likeliness of entering an aggressive relationship, or if the problems are a result of victimization. Research should be designed to assess this relationship, as it will directly affect when an intervention is implemented and what kinds of interventions are developed. Furthermore, the stable nature of aggression cultivates the necessity for design of future studies in evaluating future outcomes for victims and aggressors. The limited knowledge so far is an inhibiting factor in understanding the complete impact of RA, and deciphering the mediating link between victimization/aggression and desired prosocial behaviors.

Assessment Tools and RA

Understanding the profound effects of maladjustment on victims and perpetrators necessitates the ability to assess children appropriately that are at risk or are currently engaging in relationally aggressive behavior. As previously noted, indirect forms of aggression are more likely expressed in girls (e.g., Crick et al., 1997; Osterman et al.,

1998). However, studies pertaining to RA have demonstrated variable reports on prevalence rates. Two factors may account for the differences: different sample age groups and assessment tools utilized. As a result, extensive research has focused on trying to determine if boys or girls are in fact more relationally aggressive and to what extent this may be determined.

Overall, the main tools that have been used to assess relationally aggressive children are peer, teacher, and self-reports (e.g., Henington et al., 1998; Loudin et al., 2003; Crick, 1997). Findings by Crick and colleagues have provided extensive evidence that peer and teacher assessments of RA are correlated and have high reliability and validity. However, some investigators have suggested that peer reports are more reliable than teacher reports because they are based on multiple informants versus a single informant. In addition, children in elementary school have the cognitive capacity to inhibit their use of aggression in the presence of adults because of their desires to engage appropriately in that type of social context (Berk, 2003). Therefore, teachers may not be observing a true representation of a child's behavior. Crick et al. (1997) argued that the same may not hold true for children that are not as developed cognitively; for example, preschool-aged children. To date, few studies have examined the difference in rates of relational and overt aggression in preschool children compared to older children. However, this is an important group to evaluate because early detection of social problems may provide information on the necessity of early intervention and prevention programs.

Preschool-aged children. Crick et al. (1997) attempted to address this issue by trying to develop an age-appropriate, reliable instrument that could be used in younger

children. In addition, these researchers attempted to identify how gender predicted rates of RA. Participants included 65 (34 boys; 31 girls) predominantly Caucasian, preschool-aged children. The Preschool Social Behavior Scale-Teacher Form (PSBS-T) and the Preschool Social Behavior Form-Peer Form (PSBS-P), both adapted from previous research tools for elementary school children (Crick, 1996), were the basis of assessment for participants. The PSBS-T included questions that assessed overt, prosocial, and depressive behavior addressed on a Likert-type scale. Procedurally, teachers were asked to meet together to discuss each child before completing the questionnaire. The PSBS-P differed from the PSBS-T; the PSBS-P involved a picture-nomination procedure that attempted to address social behavior. Children were asked to point to peers that “pushed others” (physical aggression), and/or to peers that they would “like to play with the most or the least” (relational aggression). Cronbach’s alpha in both the PSBS-T and the PSBS-P was found to have a high reliability for relational and overt aggression (relational, $\alpha = .96, .94$; overt, $\alpha = .71, .77$, respectively). However, some items on each assessment measure were eliminated based on a significant cross-loading of items.

Findings from Crick et al. (1997) are the first to provide evidence that RA is expressed in children as young as three years of age. Additionally, teacher reports yielded a significant effect on gender differences for RA (girls, $M = 13.9$; boys, $M = 10.1$); however, peer reports yielded no difference. Teacher reports identified 0% of boys and 26% of girls as relationally aggressive. This finding validates previous research that show higher rates of RA in elementary-aged girls (see review of Crick, 1996). More importantly the outcomes of the research identified four distinct groups of behavior in children: nonaggressive, overtly aggressive, relationally aggressive, and both. Although

there was some overlap between the groups, the study provided enough evidence that the constructs are distinctive and should be studied as such.

One of the main limitations of Crick et al. (1997) is the viability of the procedural process used with teachers. Although the investigators intended for the group discussion to provide more reliable outcomes, one may argue that this could have potentially confounded the findings. Group-think or general persuasion strategies may have altered initial teachers' opinions about a child, and therefore scores may not realistically exemplify teachers' true beliefs. Although the inconsistent findings between teacher and student reports may not be a limitation, it would be interesting to assess why teacher and peer reports did not produce the same results and also attempt to understand if teachers or peers are more accurate in interpreting acts of RA.

McEvoy, Estrem, Rodriguez and Olson (2003) also addressed the inconsistent findings regarding prevalence rates of RA by focusing on physical aggression and RA in preschool children. The participants included 59 children across eight different preschool classrooms in a neighborhood that consisted mainly of children of White European descent. Their procedure differed from Crick et al. (1997) because direct observation measures were used in addition to the PSBS-T and PSBS-P. The main objective of the study was to address if the three assessment tools were equally useful in identifying RA and physical aggression among preschool boys and girls. In order to achieve this objective, analyses were conducted in three ways: 1) the researchers compared means and standard deviations of both forms of aggression for each assessment method; 2) they assessed the relationship between the three assessment methods; 3) they compared the differences in rank ordering of children for the three methods. A positive skew was

found on all three measures for both forms of aggression, indicating little or no aggression presentation in this sample. According to the authors, when comparisons were made between RA and physical aggression, all three methods showed the rates of girls' aggressive behavior to be higher than boys. However, the data is inconsistent with the authors' conclusions because the actual reported mean scores for aggressive behavior were actually higher for boys than for girls. The intermethod reliability findings between the three measures demonstrated small correlations between children's perception of rank order for RA and teacher ratings (RA, $r = .298$; physical aggression, $r = .390$). These results differ significantly from Crick et al. (1997), who determined that overall peer and teacher ratings obtained similar results. Correlations between direct observations and teacher ratings also did not show a significant overlap: RA showed little or no agreement between teacher and direct observation ($r = .070$), whereas physical aggression showed a slight agreement in rank order ($r = .360$). The same holds true for the correlation between direct observations and peer nominations.

Overall, McEvoy et al. (2003) found that physical aggression was more salient and therefore more readily identifiable than RA. As such, physical aggression has been studied more extensively than RA. However, the conclusions that could be made from the results of McEvoy et al. (2003) about gender engagement in RA should be made with precaution for three reasons. First, the mismatch of mean data and subsequent wrong explanations limit the validity of the final summary results. Secondly, the findings from the sample population may not be generalizable to a larger preschool population because few children were even identified as aggressive. Third, the sample was not designated as a highly aggressive population, and therefore the measures may have not been able to

detect children who exhibited a lower frequency of aggression. Previous studies disconfirm this notion that preschoolers are not aggressive.

The limitations of the McEvoy et al. (2003) study indicate a need for further research on assessment tools that yield reliable and valid results. Consistent with Crick et al. (1997), it appears that teachers may have a tendency to either over or under identify relationally aggressive children based on their own preconceived notions of the child and/or appropriate acts of behavior in the classroom. Evaluating teacher-based methods of assessment is a critical component in RA research because almost all tools utilized to address RA rely heavily on teacher reports as indicators of relationally aggressive children (Crick, 1996). However, in light of biases and expectations teachers may hold about particular behaviors or actions, evaluation of even using teachers as a source of information on RA must be addressed.

Elementary-aged children. Utilizing age to predict prevalence rates of RA may be a confounding variable because of the varying degrees of cognitive development associated with it. Therefore, it is imperative to look at a cross-section of studies that include other age groups to try to determine any inconsistencies in the data. Henington et al. (1998) evaluated an older population of students to determine the existence of gender differences and/or correlations between relational and overt aggression in early elementary school children. The measures in this study were different than previous studies mentioned; for example, the Child Behavior Checklist (CBCL) was used by teachers and the revised class play method was used for student reporting. In brief, the CBCL consists of 113 items that measure problem behavior via a Likert-type scale. The play method is known to have short-term stability, concurrent, and predictive validity. In

addition, the outcome scores have been related to high-risk status and observed maladaptive behaviors for aggressive children (Masten, Morison, & Pelligrini, 1985). It is one of the most appropriate tools for elementary school children because developmentally they are interested in imagination and play (Berk, 2003); therefore, it may be a more valid tool for this age group compared to the assessment tools defined by Crick (1996). Unlike the previous studies mentioned, this study had high statistical power because of the large sample size of 904 participants (boys, $n = 461$; girls, $n = 443$). The ethnic composition of the sample also differed compared to previous studies—it included African-American and Hispanic participants.

Results from Henington et al. (1998) were inconsistent with findings from Crick et al. (1996) because boys obtained higher peer ratings on both forms of aggression in a one way analysis of variance (overt, $M = .36$; relational, $M = .13$). Correlations among relational and overt aggression were moderate and statistically significant (boys, $r = .69$; girls, $r = .58$). The correlations are higher than what was found in Crick et al. (1996) and further suggest that peer and teacher conceptualizations of relational and physical aggression are two distinct constructs.

One major limitation of the study may be the validity in utilizing the CBCL to measure RA. Although it focuses on covert and overt forms of aggression, the measures were not designed specifically to address unique forms of covert aggression. Therefore, it may be possible that the findings were generalized to include RA, when in fact the construct was not assessed directly.

College-aged students. The study of RA has not been limited to primary and secondary aged children but has also been extended into adulthood. One such study

conducted by Loudin et al. (2003) assessed the contributions of empathy and social anxiety to RA in determining gender differences. Participants included 300 college students ranging in ages from 19-25 ($M = 21.25$). The self-report measure adopted from Werner and Crick's (1999) peer-nomination scale was utilized. Investigators believe that the self-report measure provides a reliable and valid way of assessing the actual rates and occurrences of RA; instead of relying on outsiders' views of a behavior that inherently are difficult to quantify, self-report measures are derived from the source of the behavior (Richardson & Green, 1999). Results suggest that boys are more relationally aggressive than girls (boys, $M = 7.39$; girls, $M = 5.82$). Specifically, males with low empathy were shown to be more relationally aggressive than females. Furthermore, across both genders, those individuals who feared more negative evaluation from peers were more likely to use RA as a means of retaliation and no difference in prevalence was found across gender (males, $M = 19.12$; females, $M = 20.88$). The findings in variability of gender differences must be tempered with methodological differences of self versus peer report and the age of participants. However, again this study highlights how prevalence rates of RA must be tempered with age and the kind of assessment tool utilized.

Loudin et al. (2003) did not clarify whether RA serves as an antecedent to fear of negative evaluation or if it perpetuates it. Further research may address this relationship as a means to determine where within a social cognitive framework people begin to implement aggressive strategies.

Summary. Participants' age and assessment tools used to identify overt and RA behaviors are factors that contribute to the variability in prevalence rates between males and females who exhibit aggression. However, even among studies that draw from same-

aged participants, drastically different findings in boys and girls levels of RA are evident. A macro view on the inconsistencies between gender and RA suggest that the tools used to assess RA are inconsistent and that there is a lack of agreement to which is the best and most appropriate tool. Therefore, instead of focusing on gender so heavily, it may be more imperative to shift further research efforts to more appropriate tools for assessment. As a result, the findings on gender prevalence should inherently be more statistically sound. Furthermore, if in fact researchers are not utilizing the best and most appropriate assessment tools, misidentification of children for treatment may result. Hence, the proven maladjustment trajectories may continue to develop and mental health options for children are hindered.

Prevention and Intervention Efforts

Although the argument has been made for developing or modifying valid assessment instruments, appropriate assessment is hardly enough. As delineated up to this point in the literature review, both boys and girls have been identified as engaging in, or being the victims of, RA. Consequently, we can infer that a majority of children may suffer from social maladjustment issues. Therefore, best practice indicates that the link between assessment and empirically-based intervention efforts is paramount in providing services to children in need (Goldstein & Keller, 1987).

Traditionally aggression has been treated in a myriad of ways including psychodynamic, cognitive-behavioral, person-centered, and behavioral approaches (Gontkovsky, 2002). The various approaches have informed prevention and intervention efforts that particularly focus on anger-management, conflict resolution, social problem-solving, and social skills programs (Nangle, Erdley, Carpenter, & Newman, 2002).

However, to date few programs have been empirically evaluated on outcome measures (Van Acker & Talbott, 1999) and fail to achieve a curriculum that is implementable across settings (August, Lee, Bloomquist, Realmuto, & Hektner, 2003) Therefore the following section serves to highlight prevention efforts (promoting constructive alternatives to aggression) that have been validated empirically (or are in the process of evaluation) on three levels: primary (universal programs to decrease or prevent aggression), secondary (selective programs which target at-risk populations), and tertiary (selective programs which target repeat offenders of aggression) (Young & Raffaele Mendez, 2003). This section is not intended to provide an all encompassing list of prevention programs available. Instead, it serves to highlight programs that are either at the forefront of research efforts or are well-recognized within the research community.

Primary prevention-General aggression. Dramatic efforts have been directed to the area of primary and early prevention in order to try to intervene when aggression is at its very mildest form—focused on curbing aggression prior to its increase in frequency and intensity in elementary and secondary schooling (Leff, Power, Manz, Costigan, & Nabors, 2001). Primary prevention programs have been shown to positively impact 75% to 85% of students with adjustment problems (Reid, 1993, as cited in Frey, Hirschstein, & Guzzo, 2000) Interestingly, most primary prevention efforts have been designed to be implemented within the school system. The rationale for this is potentially grounded in the fact that school is an institution where most children spend a sizeable amount of time during their developmental years. Also, aggression and violence occur on a daily basis on school grounds and therefore is a primary issue for school professionals (Batsche & Knoff, 1994).

Second Step is a well recognized program published in 1986, which intends to reduce development of behavioral, social, and emotional problems by addressing three areas of social-emotional competence within its curriculum: social problem solving, empathy, and anger management (http://www.cfchildren.org/program_ss.shtml). The program places responsibility of intervention delivery on teachers and counselors within the classroom for a wide range of students (preschool to middle school). The program's philosophies are emphasized through class discussion, role-plays, reinforcement, corrective feedback, and modeling. A total of 30 lessons are administered once or twice a week for intervals of approximately 35-45 minutes. In order to participate in an intervention effort, trainings of either one or three days are required and are provided by Second Step instructors. Additionally Second Step acknowledges the integral link between home and school as a means of cognitive and behavioral change and therefore video-based modules are available for parents.

Grossman and colleagues (1997) examined the Second Step program and the subsequent impact on social behavior and aggression on second and third graders across 49 classrooms ($N=790$). Outcome measures on the pre-/post-test control group design included teacher and parent rating measures and behavioral observation. The interobserver agreement for prosocial behavior was high ($k = .92$); however, agreement for physical and verbal negative behavior was not as significant ($k = .50$ and $.45$, respectively). The results from the measures indicated a slight positive effect for the intervention, indicated by a decrease in frequency of negative physical behavior observed (1.11 fewer acts/hour per student) and an increase in prosocial behavior (17.1 more

acts/hour/student). Teacher and parent reports showed no significant differences between the pre- and post- testing situations.

Past research on Second Step intervention effectiveness has been limited to predominantly urban and heterogeneous populations. Taub (2002) therefore investigated whether Second Step would impact a homogeneous sample of Caucasian children in grades three through five ($N=54$) from rural and low SES backgrounds. Implementation followed standard procedures as stated above. Outcome measures also were similar to Grossman et al. (1997) and included The School Social Behavior Scale (SSBS), which is a 65-item questionnaire that assesses social competence and anti-social behavior by teachers. Behavioral observations for each child were also conducted for fifteen, 90-second intervals by five trained observers. Across coding sessions, inter-rater reliability was fair to good ($k = .61 - .91$). Results indicate that children in the experimental group were rated as similar to children in the control group after intervention on the SSBS. This finding reveals that behavior in the experimental group improved because of the initial difference in competent behavior prior to the intervention. Behavioral observation results were not found to be significant: the decline of antisocial behavior was smaller in the intervention school than in the control school.

Compared to Grossman et al. (1997), the previous study was not a truly randomized controlled trial. Therefore, all the findings must be tempered with that in mind. That stated, it still appears that the findings for Taub (2002) seem to show weaker overall results for the effectiveness of this program on outcome measures in this population compared to Grossman et al. (1997). This may be due to lack of a formal and systematic approach to address treatment fidelity. Therefore, there was no guarantee that

Second Step was implemented with appropriate intensity and rigor. Additionally, each school is known to have its own set of cultural values and expectations and it is not clear whether Second Step was equipped for this population.

Although the Second Step program was not inherently designed to promote awareness of RA, researchers have tried to determine the relationship of the program to changes in attitudes about RA (Van Schoiack-Edstrom, Frey, & Beland, 2002). One such study examined Second Step in a middle school population of boys and girls from five schools in the United States and Canada ($N=714$). Teachers were the primary source of intervention and received a one-day training from experienced Second Step trainers prior to administering either a Level 1 or Level 2 module. The Level 1 module is intended for students in their first year of secondary school, and Level 2 for second year students. Students in year 2 were not exposed to the Level 1 module, which contributed to determining the effects of Level 2 without any confounding variable like pre-exposure to aggression curriculum. Two scales were used to determine outcome effectiveness: Endorsement of Aggression Scale and Perceived Social Difficulty Scale. The Endorsement of Aggression Scale is a 15-item survey that uses a Likert-type scale (1 = don't agree and 4 =, completely agree) to assess statements that either endorse or proscribe aggressive behavior. In order to address RA, questions about gossip and social exclusion were adapted from Crick and her colleagues work (1996) and rewritten as belief statements. The Perceived Social Difficulty Scale is an eight-item questionnaire that addresses difficulty of controlled anger and the ability to inhibit aggressive acts via a Likert-type scale (1 = EASY! and 4 = HARD!). Surveys were administered at the beginning and end of the intervention. Implementation integrity was thoroughly

examined via teacher evaluations, observations, rate of lesson completion, and exit interviews in order to make sound conclusions on outcome measures.

Results from the Year 1 curriculum suggested that between the pre- and posttest, students in the intervention group remained relatively constant with their endorsement of social exclusion ($t(72) = -2.50, p < .05$); whereas, the control groups increased their endorsement on this factor ($t(323) = 2.11, p < .05$). Endorsement of physical aggression did not change between pre- and posttest for boys ($t < 1.5$); however, girls' endorsement decreased between each survey ($t(111) = 3.04, p < .01$). The same pattern holds true for perception of performing proper social skills in difficult situations: boys show no change ($t < 1$) while girls do ($M = 19.92$, pre-test; $M = 18.53$, post-test). Year 2 curriculum showed more promising results for both genders. Both boys and girls obtained reductions in endorsement of physical, verbal derogation and social exclusion ($t(146) = 6.23, 8.12, 5.03$, respectively, all $p_s < .001$). Year 2 students also seemed to perceive social skills as easier to perform after the intervention ($M = 19.55$, pre-test; $M = 18.04$, post-test). An explanation for program differences across each year is not clear but may be due to variations in program content, lesson construction, or sample differences.

Although this study is one of the first to try to evaluate the effectiveness of a violence prevention program in changing attitudes and beliefs about RA, there are a few factors that may have confounded the findings. First and foremost, the measures used never specifically addressed RA as a unique construct. As shown in Crick and Nelson (2002), RA is its own distinct construct and must be addressed as such. Instead the current study circumvented RA by using the terminology verbal derogation and social exclusion. Although it may be true that these variables may constitute RA, there is no

clear evidence that these two factors contribute completely to the term RA. Additionally, adding such a few number of items to each scale to account for RA may not have adequately assessed the construct and children's beliefs behind it.

Secondary prevention-General aggression. Early Risers is another program that targets elementary school children (six through ten year olds) who are at high risk for development of aggressive and disruptive behaviors and conduct disorders (August, Realmuto, & Bloomquist, n.d.). The program was originally designed as a school-based intervention for primarily Caucasian children living in semi-rural communities; however, over a period of 10 years the program has transitioned into a community-based intervention delivered by community providers to more diverse families. Four competence areas are addressed by the Early Risers curriculum: children's academic and social competence, behavioral self-regulation, and parent's investment with their children. The responsibility of enhancing each competence areas primarily resides with the Family Advocate, who undergoes a five-day training course. The Family Advocate coordinates all intervention programs for approximately 25 to 30 child/family participants and is involved in both components of the intervention model: CORE and FLEX. The CORE (child-focused) involves regular visitation and consultation with teachers at the child's school, direct instruction of skills to make and sustain friends, and positive recognition of improvements and efforts on behalf of the child. The FLEX (parent/family-focused) portion of the intervention involves regular home visits, assessment of families' strengths and weakness, followed by goal setting and strategic planning.

Evidence-based prevention programs like Early Risers have become increasingly more important in urban and economically disadvantaged communities where mental health services are not as abundant as they are elsewhere. Therefore, it seemed necessary to replicate the findings from August, Realmuto, Hectner and Bloomquist (2001) in order to determine the effectiveness of both components of Early Risers in a natural service setting. August et al. (2003) addressed this issue by sampling children located in population dense neighborhoods in a large metropolitan city. Screening ensued for two different pools of kindergarten and first grade children; the Aggression Scale of the Child Behavior Checklist-Teacher Rating form was utilized to determine children that were at high risk for future problems ($t > 55$). A total sample of 371 participants were designated as high risk and were provided with treatment and 121 children served as the control group. Children were placed in three experimental conditions: no intervention, partial strength CORE-only, and full strength CORE+FLEX. Assessment of outcomes was based on a multimethod, multiinformant, and multitrait approach. See August et al. (2003) for exhaustive list of measures. Attrition rates over the two year period and program fidelity were accounted for and no significance was found in terms of confounding the outcome measures. Note that outcome analyses indicate that there was no significant difference between the CORE and CORE+FLEX components and therefore both intervention groups were collapsed together for statistical purposes and then compared to the control. Teacher reports indicated that all children who were designated as severely aggressive showed improvements in externalizing behavior over the first year of intervention; however, during the second year, children who received the intervention maintained gains while their control group counterparts showed an increase

in externalized behavior ratings. The effect size difference though was not significant ($p = .09$). Reports by teachers on social competence capacity indicated a significant increase in positive behavior for intervention groups between Year 1 and 2 ($Z = 2.64$, $p < .01$) and a gradual improvement in school adjustment (effect size = .31, $p < .01$).

A positive feature of this study is the adaptability of the curriculum and programming to meet the needs of the community in which it was implemented. It appears that this study was able to balance intervention fidelity with the needs of its consumers by holding meetings between program developers and supervisors of program implementation, instituting a comprehensive training program, and holding weekly supervision meetings. However, in a politically charged arena in which “cultural sensitivity” is such a central focus, it is absolutely necessary to deliver and advertise a program so that it meets the population in which it is supposed to serve. August et al. (2003) note that this was further facilitated by intervention agents that were, in fact, embedded within the community. The way this study adapted to its community serves as a paradigm for future research in terms of aiding in determination of how to modify pre-existing curriculum.

Tertiary prevention-General aggression. Tertiary prevention programs tend to be more intensive and long-term as compared to the other prevention efforts described so far. This is largely due in part to the display of life-long persistent patterns of maladaptive and destructive behavior. Students in need of tertiary prevention show a serious pattern of academic failure and repeated aggressions toward self and others (Carney, 1999). Prevention efforts are more direct and may include individualized anger management training, therapeutic day-treatment placement, and psychotherapy.

The Functional Family Approach (FFT) targets youth between the ages of 11 to 18 who are at risk or are demonstrating delinquency, conduct disorder, oppositional defiant disorder, substance abuse, or disruptive behavior disorder (*Youth Violence: A Report of the Surgeon General*, n.d.). The approach is a multi-step intervention that includes anywhere from 8 to 30 hours of direct services for both youth and their families. The phases include engagement (reduce risk for school dropout), motivation (change maladaptive behaviors and beliefs), assessment (clarify interpersonal relationships), behavior change (skills training), and generalization (ensure that new skills are applied to family needs). All of these services may be implemented in multiple settings by interventionists, mental health professionals, or any paraprofessional trained in the approach. In a meta-analysis, Lipsey and Wilson (1998) found that recidivism was lower among participants than controls. Additionally there appears to be a diffusion effect for siblings participating in the approach—court appearances decreased for two to three years after treatment.

The Multisystemic Therapy (MST) approach introduced in the early 1980s, as a “family-ecological systems approach” is founded upon the assumption that interconnected networks including the individual, peer, school, family, and neighborhood play a role in a child’s development (*Youth Violence: A Report of the Surgeon General*, n.d.). MST has been proven empirically to work as an evidence-based home treatment model for a variety of youth including drug abusers, juvenile justice delinquents, and children with mental health problems (Sheidow & Woodford, 2003). Services are provided for youth within their natural ecology and are intensive for a period of four months. Clinicians are available around the clock and take low caseloads in order to

establish an effective treatment alliance with the family. MST has been known to decrease rates of foster placements from 47% to 67% and is one of the most renowned practices for treating a myriad of mental health issues (Sheidow & Woodford, 2003)

Henggeler et al. (2003) designed a randomized clinical trial comparing MST with inpatient psychiatric hospitalization with follow-up usual services (US) to determine if MST could be adapted to treat youth with serious emotional and behavioral problems. A sample of 156 children and adolescents ($M = 12.9$ years old) were assigned to each condition and were assessed five times prior to, during, and after each treatment on a variety of factors. Factors included school attendance (Service Utilization Survey completed by school), family relations (Family Adaptability and Cohesion Scales-III), out-of-home placement (Service Utilization Survey completed by parent), and mental health symptoms (Global Severity Index, CBCL, Family Friends and Self Scale).

Outcome results of the study showed that youth in the MST condition spent more time in a regular school setting throughout the year than US youth; however, towards year-end the differences were almost unnoticeable ($r = -.28, -.12$, respectively). Caregiver reports indicated that children from the MST condition had a significant linear decrease in adaptability over time, whereas US reports indicated that children regressed to baseline over the given period of time. Additionally, cohesion data in the MST treatment group took the form of a “U-shape,” suggesting ultimately a rise in cohesion, while the US group had no significant change ($r = .49, -.25$, respectively). No statistical differences were found for out-of home placement or mental health symptoms.

The current study, unlike other published MST studies (Henggeler, Clingempeel, Brondino, & Pickrel, 2002) was unable to attain long-term outcomes for patients with

serious antisocial behavior. The researchers suggest that MST may not be appropriate for children that need sustained intervention over long periods of time. Additionally it has been noted that new intervention protocols should be designed to represent advances in the field of antisocial behavior and thus be integrated within the therapy model.

Prevention and relational aggression. Unlike the countless numbers of programs available to treat physical aggression and violence, research in the area of prevention for RA is much more restricted. It can be inferred that the historical background of RA research may partly contribute to the lack of program construction and implementation. However, the current zeitgeist in schools and social settings suggest the imminent need for programs specifically targeting RA for both boys and girls, on a primary and secondary prevention level. Unfortunately to date there are only three recognized programs in the United States specified as RA prevention programs. These programs have limited information available on empirical validation however. Therefore the next section of the literature review serves to primarily highlight key features of each program, and any statistical information available at this time. The necessity for further evaluation of these programs is evident.

Creating a Safe Social Climate In Our Schools (CASS). CASS was the first intervention program designed to treat RA. Development began in 1998 as a response to the experiences and emotions that girls expressed during council meetings at the Ophelia Project (*Creating a Safe Social Climate in Our Schools*, n.d.). The primary goal of CASS is to address peer covert aggression in schools by providing a set of normative behaviors for parents, students, and educators. CASS targets middle school communities placing a special emphasis on RA. Unlike other RA programs, CASS integrates the

entire school community including administration, students, and faculty over a period of at least one year to develop action plans that are appropriate for each individual school. The use of school mentors is a key component to the program and serves as an empowerment tool for the students.

CASS's one and a half day training has three major components: Awareness Building, Strategic Planning, and Mentor Training. "Awareness Building" is an examination of various ways that boys and girls use aggression. Aggression terminology and current research is addressed during this portion of the training. "Strategic Planning" involves both students and adults to develop socially normative and appropriate action plans. Finally "Mentor Training" occurs on the second full day of training, whereby high school students are taught how to deliver the curriculum, "How Kids Hurt Each Other/How Kids Help Each Other." The selected students facilitate hands-on activities, discuss personal stories, and engage in role-play and guided group discussions with the middle-school students. Key elements that result from participation of the training program include empowerment of students to make a change in their schools, motivation of parents to become an integral part of school safety, and creation of a task force to track short and long-term change in schools, etc.

Intervention studies began in 2001 to determine the effectiveness of the program. Results are currently unavailable.

Allies in Action. Information in the following section is based on an informal phone interview with the director of the Girls' Initiative Networks (GIN) and from the GIN website (*Allies in Action*, n.d.). The GIN copyrighted *Allies In Action*, a program intended to serve as both primary prevention and a direct intervention in 2003 after

piloting the program on over 1,000 girls over a five year period. The program looks at the full gamut of aggressive behaviors in women, asserting that women may be both relational and physical aggressors. However, GIN acknowledges that most likely RA is the impetus to further physical aggression, assuming the social acceptability of physical aggression within one's own culture. The curriculum is unique in that it was not only developed by professionals but by 300 Latina and African-American girls who also assisted to cultivate a "real-world" curriculum that would speak to girls of similar ethnic groups. The curriculum is based on trying to encourage girls in the program to analyze the root of conflict—be it race, class, gender, etc. From there, the program attempts to model appropriate conflict resolution techniques by group skill building exercises. "We want them to be their own cognitive-behavioral therapist and learn how to use their deconstructive faculties," said Annette Klinefelter, Director.

The essence of the program is based on the hope that girls will "pass on the wisdom" that they learn from Allies in Action. Therefore, the curriculum is set up as such: 20 girls are trained with a core group of adults that will act as co-facilitators. During the youth facilitator training, girls between the ages of 15-25 will attend an intensive six hour course. The objectives of the course include: development of facilitation and conflict resolution techniques, engagement of girls as "cultural healers" in their communities, and education on RA (where it comes from and how it affects others). The staff development training is wider in breadth including self-defense exercises, interactive brainstorming sessions on intervention techniques, and a discussion of risk and resolution strategies for RA. By the end of the workshop, adults should be able to

demonstrate tools for intervention to school administrators, serve as a resource for schools who need aide in RA challenges, and help school personnel learn to identify RA in girls.

Following training, it is intended that the group of 20 girls will then attend neighboring schools or other organizations and distribute the information they learned by conducting the curriculum with another set of 20 girls. This process is built on the intention that the program's goals will trickle down and permeate large groups of girls. The last piece of the curriculum involves parents and the community. As has been previously identified, parent and community involvement is a key success factor in implementation and efficacy. Therefore GIN holds workshops for parents and community members ranging from one and a half to three hours depending on expressed interest. The workshop offers concrete strategies to help their children cope with conflict, and also discusses various factors of RA.

A small study was run by the developers of the program. A pre-/post-test survey assessment was conducted measuring girls' self-efficacy to problem solve conflicts effectively. Based on the small sample size and lack of experimental design, the results were not statistically significant and no follow-up evaluation has been conducted.

Owning Up. The *Owning Up* curricula was developed over a number of years by The Empower Program and other experts and/or survivors of violence and aggression (*Owning Up*, n.d.). It is based upon the premise that pro-active intervention that addresses gender-specific violence predicts future trajectories of antisocial behavior. The assumption base of the program revolves around the concept of "boxes" of feminine and masculine behavior imposed upon by society; for example, boys are expected to be strong and athletic, whereas girls are expected to be demure. As a result, cultural scripts factor

into how boys and girls support and act out violent behavior. The program acknowledges research on non-normative gender aggressive behavior in addition to feminine forms of aggression such as social hierarchies and cliques.

Two separate curricula are administered for boys and girls. A combination of discussion, role-play, and lecture raises students' awareness of gendered violence. Special attention during all activities is given to avoid labeling girls or boys as victims or perpetrators when discussing violence. The goal of the first part of the curricula is to gain a basic understanding on how youth influence each other, in addition to developing an understanding on how isolation and gender stereotyping may increase risk of violent behavior. The second part of the curricula focuses on gender-specific intimate violence such as date rape, harassment, and domestic violence. At the completion of the program, students should be able a) to question their own assumptions and behavior regarding gender-based aggression and violence, b) to safely remove oneself from a violent situation c) to assist others in attaining help, and d) to recognize the signs of an abuser.

In September 2002, The Empower Program published a report that examined the effectiveness of the *Owning Up* curricula on a sample of 180 female students and 108 male students, ages 15 to 20 in Maryland and Washington DC. The evaluation served two primary purposes: 1) to examine Empower's assumptions on youth violence and 2) to evaluate the effectiveness of the curricula in achieving the stated goals. A pre-/post-test design was utilized to measure effectiveness in three areas: 1) immediate changes in knowledge about intimate violence, 2) immediate changes in attitudes on variables that have been known to be pre-cursors to both victimization and perpetration (e.g., gender stereotypes, low self-esteem). Five measures were used to assess program effectiveness:

Attitudes Toward Conflict Scale, Attitudes Toward Women Scale, Dating Violence Scale, General Perceived Self-Efficacy Scale, and Conflict Resolution Style. All measures show fairly stable internal consistency ranging from .62-.90.

Results relative to the first purpose of the evaluation (stated above) were measured using the Pearson Correlation; however no specific statistics are available. Findings indicate that the Gender Stereotyping Scale and the Conflict Resolution Style scale were positively correlated with Attitude Toward Conflict among girls. Scores on the Perceived Self-Efficacy measure were negatively correlated with Attitude Toward Conflict. In the male sample, scores on the Gender Stereotyping Scale were positively correlated with Attitude Toward Conflict. Three factors in both boy and girl populations were correlated with victimization of violence: scores from the Youth Dating Violence Scale, family violence, and previous perpetration of violence. Paired sample t-tests were used to determine whether changes were significant between pre- to post-test. According to the Change in Knowledge scale developed for this evaluation, females reported a knowledge increase about rape (54.8%), sexual assault (59.5%) and abusive relationships (69.8%); males reported similar results about knowledge increase about rape (60%), sexual assault (57.5), and abusive relationships (60%). Results from the Attitudes Toward Conflict and Attitudes Toward Women suggest that females and males had no significant change between pre- and post-test. Results on the Perceived Self-efficacy scale show no difference for females, however are statistically significant for males. Finally, female respondents reported a decline in their style of conflict resolution implying that they are able to deal with verbal and physical aggression impulses better. However, no statistically significant changes were found for boys.

The inconsistent results between boys and girls seem to be precarious based on the fact that they are receiving similar instruction. This is not to say that the curricula were ineffective, but instead that the treatment was differentially effective based on gender or other potential individual variables. Future research may want to address differences between each curriculum to determine if certain components which are not similar are actually producing a higher rate of positive change for either gender. Secondly, the lack of experimental design and treatment utility assessment may have contributed to the lack of consistency in results for both boys and girls. It may be possible that the interventionist was not adequately following the designated curriculum for each gender. Additionally, future research may pilot the curricula on younger students who may not have such deeply embedded aggressive attitudes and behaviors as do their older peers.

Limitations and barriers to prevention. Regardless of how efficacious and sound a program may be on paper, the reality is that numerous limitations across cultures, communities, and schools may limit how well a prevention program may actually play out in a naturalistic setting. Therefore, when implementing any evidence-based program, practitioners should be cognizant of a few key areas that may conjure resistance.

Culture. As highlighted previously in Taub (2002), cultural issues and expectations factored into the roll-out and efficacy of each intervention.

Additionally, culture within the home environment may also be a limiting factor in how well a child is able to integrate program skills; for example, in some household cultures, the use of aggression, intimidation and ridicule is taught at home as an acceptable method of self-defense (Vernberg & Gamm, 2003).

Therefore, practitioners should attempt to try to understand the youth they are working with and the backgrounds they are coming from and adapt curriculum to fit the cultural group.

Community. Buy-in from community members is seen as one of the primary limitations for prevention program implementation. In wealthier communities, convincing parents and teachers about the prevalence of aggression in their school system is extremely difficult. However, as research clearly states aggression runs the gamut of SES, ethnic, and racial groups. At the opposite end of the spectrum, poverty stricken communities face an enormous number of dilemmas and issues and may resist another program that requires more parental time and resources that are realistically unavailable (Vernberg & Gamm, 2003). Support from local government and business agencies may help to remedy both situations by either making parents aware of violence and crime rates or by providing funding.

School. School-based programs so far seem to be one of the best approaches to help students develop tolerance, settle differences and understand others (Gregg, 1998). However, schools oftentimes are at the forefront of resistance for new programming. Perceptions that teacher and school administrators may hold about the role of their job may limit what they see as an educational problem. Along the same lines, if a shared vision for school goals is not held equally by all staff, full implementation may not be possible or may be tainted by those that are not fully committed to decreasing school violence. Finally, the age old issues of limited time, money and resources may limit how

well a school responds to prevention efforts. A potential way to promote prevention efforts is to designate a “champion” that keeps school morale high for the program and follows through by evaluating and promoting the success of the program.

Summary. Prevention research clearly illustrates the link between effective programming and decreased rates of aggression and violence in our schools and communities. Assuming that a curriculum has the appropriate tools in place to combat limitations and barriers, there is a high probability that it will function to help increase prosocial behavior in youth. The lack of programs created for RA has indicated an area for further development and study.

Summary and Future Research

The adages, “boys will be boys” and “that’s just how girls are” are no longer justifiable excuses to condone physical and/or relational aggression. As evidenced in this literature review, aggression of any sort, affects youth in profound ways and cannot be dismissed as gender normative and appropriate behavior. In fact, aggression may impact the way a child is rejected or accepted by his/her peers, or it may alter one’s own mental health outcomes. Evidence shows that concurrent and future outcomes for aggressors and peers are not positive. Inevitably, limitations in the current body of research must be addressed in order for educators and mental health professionals to secure a comfortable place for children within the school environment, affording the children an opportunity to engage in healthy relationships that will help to develop their own social and personal skills. Based on the high prevalence rates of aggression exhibited within the school

system, lack of future research and limited links to practical strategies, will only hinder these possibilities of positive outcomes for aggressors and victims.

As shown throughout this paper, there are some serious gaps and limitations within the research in a few specific areas. First and foremost, the assessment tools to identify problematic behavior are wide and produce inconsistent results. For example, peer, self, and teacher reports result in different outcomes in some studies but not others. In order to correctly identify children at risk, we must have valid and reliable measures. It would behoove future researchers to try to assess the stability of current measures across different segments of child populations and determine which may be the most appropriate tool. In addition, development of new assessment tools dependent on developmental age groups may also be useful in gathering more consistent data.

Second, the majority of research tends to be focused on preschoolers and early childhood populations. Based on the cognitive changes within children as they develop, it is imperative that more studies are geared towards middle school and high school children to provide a complete picture of relational aggression. In fact, these children may exhibit completely different patterns of behavior and outcomes as a result of different values and interpretations that coincide with development.

Third, detection of the risk and resiliency factors associated with aggression and/or victimization are not evident. If indeed the outcomes are less than positive for victims, it is critical to determine if there are resiliency factors that may shield children from antisocial outcomes. Finally, one of the core limitations in this area of research—although not addressed directly in this analysis—is the extent to which various researchers operationalize and define indirect aggression. There is a great deal of overlap

between indirect, social, and relational constructs. Therefore, the inconsistent and/or overlapping findings between studies may be a result of how investigators consider similarities and differences between each form of aggression.

Fourth, an abundance of research has concluded that various programs serve as effective tools in fostering children's understanding and attitudes towards physical aggression and violence, in addition to producing behavioral changes. However, the same cannot be concluded about RA. A clear gap exists between the prevalence rates of RA expressed in girls and boys and the programs available to intervene. This creates a need to address efficacy of the existing programs with more rigor than has previously been accomplished, paving a pathway for effective programming in a principal area of aggression research that will have direct consequences for children that are in urgent need of support services and proper identification.

Current study. The current study attempted to do exactly this by evaluating the Empower program in terms of treatment utility and outcome effectiveness in modifying attitudes and behaviors related to RA and gender-stereotyped behavior. Although research has been conducted in one case for the Empower program, there are a few key limitations that the current study will address. First, the Empower program has not been designed in such a way that affords drawing the link between curricula and outcome effectiveness because of the weak pre-/post-test design. The current study addressed this using a control school, in addition to administering tests at the pre- and post-tests. Second, a clear "dosage-response" relationship (Olweus, 1997) was not established because the evaluation study did not account for variations in program implementation. This issue was addressed by measuring treatment integrity through observations,

interventionist checklists, consumer surveys, and exit surveys developed by the researcher. Third, the previous study only utilized self-reports to ascertain outcome effectiveness. Although self-reports provide valuable information, a multitrait, multiinformant approach has proven to be more effective. Consequently the current study utilized peer, teacher, and self-reports to provide a more comprehensive view of participants behavior. Fourth, the current study addressed attitudes towards RA with measures adapted from Crick (1996), as they were not addressed previously. Finally, the current study served to provide information on the generalizability of the program to populations other than the one described in the initial evaluation.

Chapter 3

Methods

Overview

The strategy for evaluating the *Owning Up* curricula had two major goals: 1) to determine if the project was implemented properly, and 2) to assess if participants experienced attitudinal and/or behavioral changes. In order to judge the integrity of the outcome data and make a reasonable conclusion on its validity, the project included an evaluation of the treatment integrity of the program. Therefore, the evaluation combined a mixed strategy of formative and summative evaluation. The formative component was designed to collect data throughout the term of the project and encourage facilitators to reflect on problems and/or progress with the implementation process. In essence, the formative component served to validate and/or ensure that the goals of instruction were being met. The summative component, on the other hand, served as a measure of the *Owning Up* curricula's efficacy and was quantitative in nature. This phase of the project used measures (delineated below) to determine what kinds of effects, if any, the curricula had on the participants compared to a control group.

The evaluation used a quasi-experimental design including two schools that received the intervention and one school that served as the control school. Pre- and post-test measures were collected at all schools.

Development of Study

It is important to note that the conceptualization of this study occurred in conjunction with an outside organization, The Ophelia Project-Tampa Bay (OP-TB). OP-TB is a non-profit organization founded in 1997 by parents who realized that action

was necessary to help counteract negative issues that girls were facing at school and in their communities. OP-TB initially trained some of their employees in the *Owning Up* program, and they piloted the program in girls' camps during the summer of 2003. Based on observations, the facilitators at OP-TB felt that the *Owning Up* program was a valid tool to help girls deal with issues of violence and aggression. However, they were interested in gaining more empirically validated evidence about the effectiveness of the program to determine if implementation in other schools and environments should be undertaken.

Participants and Setting

The participants in this study were drawn from three middle schools in Tampa, Florida. Each school was chosen based on relationships that were forged by OP-TB and school administrators. Overall, the students at each school differed in terms of demographic characteristics; however, the student groups that were chosen to participate shared similar demographics. The ethnic composition of students at School 1 was 35.2% Caucasian children, 42.8% Black children, 17.9% Hispanic children, 2.3% Asian children, and 1.9% Other children. Approximately 61.3% of students received free or reduced price lunch. School 1 has earned a rating of "A" for the past two years by the Florida Department of Education. The ethnic composition of students at School 2 was predominantly Caucasian children (52.6%), followed by Hispanic children (29.7%), Black children (11%), Asian children (3.4%), and Other children (3.3%). The percentage of children on free or reduced lunch was considerably lower (37.3%). This past year, School 2 received an "A" from the Florida Department of Education. School 3, the control school, received a "B" rating for the past two years. The exact ethnic

composition of School 3 was unavailable on the Florida DOE website and school administrators did not have statistics available; however, a total of 69% of students were designated as minority students. The percentage of children on free or reduced lunch was higher than in both Schools 1 and 2 (71%).

A select group of participants from a Hillsborough County school initiative, “Program to Empower Teens (PET),” were selected to participate in this study by school administrators prior to the researcher’s involvement with the evaluation. The main rationale for selection of students in PET was an identifiable need of the children to receive an intervention such as the *Owning Up* curricula. Children in the PET program indicate in some way that they need mentoring and/or encouragement beyond what is considered typical for school children. According to school administrators, youth in the PET program typically experience life challenges at home, demonstrate behavior problems in the classroom, lack social skills, have excessive absences, have poor grades, and are frequently tardy. Referral to the program may come from a teacher, parent, or administrator. Students in the program attend a standard of four core classes (main academic subjects), one elective, and one PE class—all of which are 50-minutes in length. Students in PET are pulled out of one of their electives once a week to attend a class led by a student intervention specialist (SIS). The SIS specializes in working with at-risk youth and provides tailored programming to help the students in PET succeed in school and in life. Programs may include lessons on social skills, study skills, or time management. Additionally the SIS is notified if any student in PET is involved in an incident at school and works with the child and others involved to determine the appropriate consequences and level of punishment.

Students in PET who participated in this intervention included seventh and eighth graders at all three sites. Due to the limited number of students in PET at the control school, sixth graders in PET were included in administration of the pre- and post-test measures. A total of 130 children ($n = 70$, boys; $n = 60$, girls) were initially expected to receive the intervention, and 60 children ($n = 30$, boys; $n = 30$, girls) were expected to serve as the control group. However, due to limited numbers of parents consenting, the actual number of students that participated was smaller. A total of 55 students ($n = 23$, boys, $n = 32$, girls) received the intervention, and 20 students ($n = 8$, boys; $n = 12$, girls) served as the control group. Notably, this number is even smaller when students are matched on pre- and post-test administrations. The total number of participants who completed both the pre- and post-test was 32 students ($n = 15$ boys; $n = 17$ girls) who received the intervention and 15 students ($n = 5$ boys; $n = 10$ girls) who served as the control group. Participants ranged between the ages of 10 and 15 years with the mean age being 13.02. In terms of ethnic composition of all the participants, the majority of students were African-American (53.4%) followed by Caucasian (24.87%), Latino students (20.7%), and then Asian students (.03%).

Procedure

Planning. Prior to implementation of the program, one staff member from the OP-TB worked with the three schools in order to oversee any logistical issues such as setting up exact days and times for the distribution of parent permission forms, pre-test measures, and the weekly intervention day and times. It is noted that the researcher was limited to interactions with the OP-TB and did not have the opportunity to interact with school personnel. Staff from of the OP-TB also worked with administrators to define

what role OP-TB staff and volunteers, SISs, the researcher, administrators, and students would have.

The researcher and two staff members from the OP-TB worked collaboratively to problem-solve issues that arose during the implementation of the project. Overarching goals as well as short-term goals were discussed and evaluated, allowing the researcher to make appropriate decisions on process and outcome evaluation designs. Regular emails and weekly meetings occurred in order for the researcher and the OP-TB to communicate about issues including budget, measurement tool decisions, and teacher and parent buy-in. The thesis chair also participated in three of the face-to-face meetings to guide and refine any decisions that had been made. Based on discussions during these meetings, specific measures were identified to evaluate outcomes. The researcher contacted various organizations and professionals to obtain the measures that were used in the current study.

Pre-data collection. This phase of the evaluation concentrated predominantly on amending or creating measurement tools. All of the measurement tools for the formative evaluation component were created by the researcher prior to the intervention (see “Measures” for further detailed information.) The Aggression Questionnaire (AQ) was the only tool used in this study that is fee-based and therefore was purchased during this phase. In order to receive a discount for the tool, a letter was sent to WPS, the publisher, stating the rationale for this project and the intent behind the use of the AQ. WPS agreed to provide a research discount of 20%.

Also during this phase, OP-TB facilitators and the researcher attended a conference in Washington D.C. to gain experience with the *Owning Up* curricula to either

a) lead youth groups or b) train others to lead groups. The purpose of this trip was to ensure that everyone who was to be involved in the implementation had been trained by the creators of the curricula, thus maintaining a level of consistency across facilitators. A total of five people attended the training including the researcher, a volunteer, and three paid OP-TB staff members. However, one staff member left the training early because she was uncomfortable with the way the curricula was being taught and therefore was no longer part of the implementation process.

It was decided that a total of three trainers who were either employed by the OP-TB full time or had volunteered to work as a facilitator leading this implementation effort were chosen to lead each group. Facilitator 1 was an African-American female in her mid-30s who received her Bachelor of Science from University of South Florida in Anthropology. She has been working in the non-profit industry since 1997. She has been trained in multiple youth programs including YMCA training modules, Foundations for Behavioral Healthcare, and Girls' Circle. Facilitator 2 was a Caucasian female in her late 40s who has a degree in business and marketing. She specializes in business marketing currently. She has been volunteering for the OP-TB for the past 3 years. She has led many groups for both boys and girls in private schools across Tampa Bay. Facilitator 3 was an African-American female in her late 20s who received her Bachelor of Science in Exercise Science and Sports Studies from The University of Tampa. She has been working with youth since 1999 in fitness-type settings. She has been part of the OP-TB staff since 2003 and directs programming for gender specific groups.

Prior to administering any of the outcome measures, the researcher obtained a roster from each SIS at the three schools. The roster included the names of all students

who received parental consent. The researcher then formatted the Peer Nomination Scales for Children to include each measure question at the top of each page and the names of each student under each question. This formatting was completed for each roster (at each school) in order to prepare copies for the pre- and post-test administrations.

The researcher also designed a detailed parent permission form for all students (see Appendix A). Parent permission forms required a signature and had to be received by the researcher prior to the pre-test administration in order for any of the students in PET to participate in the data collection. In order to try to establish a high rate of return for permission slips, the researcher intended to hold a parent meeting to answer any questions or review the parent form directly with parents. In this meeting, the researcher hoped to highlight any benefits of the program in addition to explaining the logistics of how the curricula would be taught. However, time constraints of the SISs and logistical issues in the school in terms of inviting parents after-school hours made this opportunity unavailable. Instead, parent permission forms were distributed to each student in PET. The initial return rate based on this distribution of permission forms to students was very low at all three schools, however. Therefore, the SIS at one of the schools made calls targeting specific parents and drafted a letter to the parents that was attached to a second round of parent consent forms. This decision on how to move forward to obtain more consent forms was made independently of the researcher. Students who did not receive permission from parents to participate in the study were not administered the pre- and post-test measures. However, all students who were part of the PET program still received the intervention during their regularly scheduled PET class time, based on the

SISs' recommendation. Although the researcher preferred the students to be separated out into participants and non-participants, this was not feasible for scheduling in the schools.

Once permission forms were received, the researcher and the facilitators had planned to attend each school to meet with the children to explain the goals of the program and the intervention structure. The main point of this meeting was not to highlight aspects of the curricula, which in fact could potentially taint the pre-test data. Instead, the meeting was intended to merely serve as an introduction of the researcher and facilitator, so that the students could become comfortable with each prior to implementation. In addition, the researcher wanted to address basic logistical issues about how long the program would last, the structure of the program, and pre-and post-test dates, as well as provide each student with a research number. For reference, the research number was derived by requiring each student to use the first letter of his or her first name and the first letter of his or her last name, and the number that is associated with the month of his or her birth (i.e., if a student was born in December, he would place a 12 behind his first and last initials). Because the purpose of this study was to make generalizations about the group as a whole, this number served to track each student and ensure that data was collected for each participant. Names were not associated with any students during data analysis.

Unfortunately, the researcher and the facilitators from OP-TB were unable to coordinate a time that was mutually convenient for both parties to attend each school. Therefore, members from OP-TB attended each school and held an icebreaker activity

during the first week of the implementation. Logistical issues, including distribution of research numbers and discussion of program structure, were not addressed.

Data collection. The researcher, an undergraduate student volunteer, a clinical psychologist supervisor, and a member of OP-TB were in charge of all data collection. Depending on schedules, each person was assigned to one of the schools to collect summative evaluation data. Each person or group of people were in charge of distributing the measures, proctoring the classroom, and collecting the data once it was completed, and finally returning it to the researcher if she was not present.

The three self-report measures were stapled in a packet and were randomized in order to eliminate any potential effects of completing one measure before another. For sake of confidentiality, the researcher had requested that children receive a barrier made of cardboard to protect students' answers. However, the SISs and OP-TB staff felt that student confidentiality would be maintained by placing students at separate tables in the school cafeteria. Upon distribution of the packet, the intention was for the data collector to follow standardized administration procedures as suggested by the manual for each measure in addition to instructions written by the researcher. However, some of the data collectors did not follow the standardized procedures and instead roughly followed the basic procedures without reading the script. The proctor also was supposed to ensure that each student's research number was on the top of each page of his or her testing packet. The proctor was instructed to walk around the room during testing and check children's packets to see if the number was on each page, as well as write a reminder on the board regarding the importance of writing down the research number. The proctors made the decision that writing the reminder on the board was unnecessary and this therefore was

omitted from the standardization instructions. Upon review of the data, the research numbers were not on each page of the measurement tools.

Once the participants completed the packet of questionnaires and turned them into the proctor, a roster of all participants was passed out to every student. The proctor read aloud instructions from the manual regarding the administration of the Peer Nomination Scales for Children, and the participants used the roster accordingly. Following the administration of this measure, all packets were handed into the proctor. Subsequent to data collection, the proctors placed all data in manila envelopes provided by the researcher and returned them to the OP-TB office. The researcher then collected the envelopes. In the event that a child was missing from the pre- or post-test, the researcher was asked to be notified so that rescheduling could take place. This did not occur during the study.

Evaluation data was maintained by the program facilitator at each school and was returned to the researcher via a pre-addressed and pre-paid manila envelopes sent through the mail. Each facilitator was asked to send their forms twice during the implementation: once halfway through the program (at four weeks) and once at the completion of the program (at eight weeks). This was to ensure that if there were any major issues halfway through the implementation, they would have had an opportunity to be addressed by the researcher. The researcher only received the forms once at the end of the implementation. Additionally, the researcher's phone number was given prior to the implementation to each facilitator for the purpose of helping address any major issues that needed it. The researcher was called by one facilitator during the program who expressed her concerns on methodological issues of the current study. I advised this

facilitator to either contact the chair of my committee or the research coordinator about her concerns if she had any further questions, after our discussion. This facilitator contacted OP-TB and was thanked for her concern but was asked to please allow the research portion to be conducted by the researcher and those specifically involved in the research methods of the project. Data collection of observational data was conducted by three different licensed psychologists and after each observation, data was sent to the researcher.

Intervention. During this phase, each student was expected to receive a total of eight, 50-minute sessions with an *Owning Up* facilitator. Unfortunately, this was not the case at School 1. There were many reasons for this gap in instruction including lack of communication between the SIS and the OP-TB facilitator, inadequate facility space for teaching the students, SIS cancellation of classes, and limited follow-through from the SIS to encourage students to attend the *Owning Up* class.

Participants were either pulled out of PE class or an elective class once a week to attend the intervention. Students in PET were separated into gender groups for the implementation, as there are separate *Owning Up* curricula for boys and girls. The facilitator conducted four groups on the same day each week: two for girls and two for boys. During the class period, the SIS who usually leads programming was asked not to take part in activities or be present in the room during instruction. The intention behind this was to allow the children to build rapport with the facilitator and have a clean slate during this program. The SIS at both schools played a role during implementation, however. At School 1, the SIS periodically moved in and out of the classroom and participated during discussions. At the mid-point of the intervention, the SIS wanted to

change facilitators—with her being the primary facilitator. After discussions with the facilitator and the OP-TB team leader, this did not occur. The SIS at School 2 also participated quite often in discussions and helped to manage behavior in the classroom. At the end of each group session (hence a total of four in one day), the facilitator was asked to fill out an implementation checklist related to what she just taught in the lesson, as well as a reflection on the effectiveness of the facilitation. The facilitator at School 2 completed all the forms after each session. The facilitator at School 1 only completed 10 forms because she felt that by the middle of the implementation, the curricula was not being taught effectively and there were too few students in attendance.

It is noted that modifications were made to the initial curricula, which consists of a total of 16 lessons. Based on time constraints of the class period and the fact that it was near the end of the academic year, it was decided that the program would be run for a total of eight weeks. The modifications and choice of lessons taught are listed below:

Class 1: Getting started: Understanding social and cultural definitions of femininity/masculinity (Omit HS role play; Omit HW assignment)

Class 2: Groups (Omit HS role play)

Class 3: Friendship is not a joke (No changes)

Class 4: Knowing your culture (No changes)

Class 5: Knowing your image (No changes)

Class 6: What's up with guys/girls (No changes)

Class 7: Reputations & the double standard (No changes)

Class 8: Power of apologies & self acceptance (Letter of apology exercise. Ask girls to get in pairs and share their letters if they feel comfortable; Handout already created for the “I am” exercise; Students can write “Dear Self” letter in their journals; Give them rectangle shaped card stock paper to create their mission. Post missions on a large poster board)

Pre-Survey Day (Open with “ The Last Word” class and then administer post-test measures)

Adjustments in time also included: Objectives and set up (5 min.), Talk about it (10 min.), Do it (8 min.), Act it out-RP (15 min.), Write about it-(8 min.), and Wrap up (4 min.).

Please see “Measures” for a formal outline of the program contents in their entirety.

At the same time that the students in PET at the School 1 and 2 received the intervention, the lessons for the students in PET at School 3 were monitored. This data was considered important to collect in order to determine if the control school students received instruction that resembled the experimental schools in the intervention. The SIS at the control school submitted a total of six lesson plans. Lessons included discussions on academic goals, grades, peer pressure, attendance, and motivation after the FCAT. The SIS did note that one class was taught that specifically targeted bullying and conflict resolution. She noted however, that bullying is an on-going problem with females in PET, and that she has been working on this all year with three girls specifically.

However, only one of three girls that have been targeted by the SIS participated in this study.

At two points during the implementation process (beginning and end of the intervention), three Ph.D. level psychologists were to randomly observe the facilitators leading each group. The main rationale for this was to validate that the facilitators were leading the program in a manner that is consistent with best practices in psychology and education. Based on the late start of the program, time constraints for one of the psychologists, and lack of consistent programming at School 1, only one observation was completed by each of the three psychologists.

At the end of the intervention, the post-test measures were distributed. As highlighted previously in the pre-test design, the same structure and measures were utilized for the post-test design. Again, if the child was absent for the post-test, the SIS and the researcher were expected to work together to administer the measures to the child at another time. However, since the post-test was administered close to the last week of school, this was not feasible.

Post-data collection. Once all the data was submitted to the researcher, she and the undergraduate assistant matched the pre- and post-test packets by matching research numbers. The researcher also reviewed each packet to ensure that a majority of the questionnaire was filled out and that no pages were skipped. This also was supposed to be done by the facilitators during each administration of the measures. The researcher also did a preliminary review of the formative measures to ensure that they had been filled out correctly and completely. Exit interviews with the facilitators also were completed at this point to determine the consistency with which the facilitator believed

that delivery of the curricula was in line with the goals of the program. Once all the data had been thoroughly reviewed, the researcher and the undergraduate volunteer entered the data into a database for future analyses. To see a timeline of the expected process, see Appendix J. Note that this timeline was not followed however, as each site followed their own timeline that changed each week.

Measures

Measures for the summative portion of this study are empirically validated for use with a wide variety of youth. Based on the nature of this curricula, it seemed suitable to address aggression as a general construct and then specifically as a relational construct. Based on the limited programming available to treat RA, it seemed more than appropriate to ascertain if the *Owning Up* curricula was able to effectively alter both affect and behavior as it relates to RA. Furthermore, because RA is so subjective in nature and requires a relationship between people, it seemed that a self-report measure alone would not be sufficient in determining if there was a significant change in outcome measures. Therefore, the researcher decided to employ a multi-informant method using both peer and self-reports.

To determine the level of fluidity and utility of the program, it was not only important to evaluate how the facilitator felt about the implementation of the program but also what students took away from it. It was possible that the two schools had very different interpretations of the program. Therefore, formative measures designed by the researcher were developed for both the facilitators and the students to gain an understanding of each person's perception on the implementation. Furthermore, if the outcome data showed inconsistent or unusual results, the data from the formative piece

could highlight any areas that were not implemented properly and/or any other issues that may have impacted outcomes.

Dependent Measures

The Aggression Questionnaire (AQ). The Aggression Questionnaire (AQ; Buss & Warren, 2000) is a self-report inventory designed to evaluate an individual's aggressive responses and the ability to channel those responses in a safe and constructive way (see Appendix B for a copy of the measure). It is a revised version of the Buss-Durkee Hostility Inventory (BDHI; Buss & Durkee, 1957), which is a well-established measure to assess aggression (Lambert, 2001). The major difference between the two scales is that the most recent one does not include defensive or socially acceptable (untrue) answers from the respondents because of the Inconsistent Responding Index (INC) that is built into the scoring procedure. The AQ is a 34-item questionnaire composed of five scales: physical aggression (PHY), verbal aggression (VER), anger (ANG), hostility (HOS), and indirect aggression (IND). The respondent rates items that are written at a third-grade level along an intensity scale ranging from one to five. The two poles are: 1 = not at all like me to 5 = completely like me. Scores range from 34 to 170 depending on the severity of the self-reported behaviors, with higher scores indicating a more elevated level of aggression. The measure assesses either a treatment need or treatment outcomes and can be used in a variety of settings including school, clinical, military, and business.

Standardization of the scale was based on a sample of 2,138 individuals ranging in age from 9 to 88 in the United States. There are three age group norms: 9 to 18, 19 to 39, and 40 to 88. Of the individuals sampled, 1,252 were women, and 880 were men. The racial background of the norming sample was predominantly composed of Caucasian

participants (72%), followed by Black (15%), Latino (8%), Asian (1%), and Other (3%).

A clear limitation is that the sample suffers from under-representation of minorities.

Although the sample that was used in the current study was heavily skewed towards minority students, this measure still seems appropriate based on its comprehensive assessment of overall aggression and its various forms.

Buss and Warren (2000) reported good to moderate reliability: PHY ($r = .88$), VER ($r = .76$), ANG ($r = .78$), HOS ($r = .82$), IND ($r = .71$), total scale ($r = .94$). The concurrent validity was assessed by comparing the AQ with other scores from similar scales. The correlations between the AQ and the Novaco Anger Scale and the Provocation Inventory were moderate to good ($r = .74, .59$, respectively). However, the same is not true for correlations between the AQ and the Attitudes Towards Guns and Violence Questionnaire and The Children's Inventory of Anger ($r = .38, .37$, respectively). Researchers note that further studies must be conducted to assess the validity of this tool but that overall it is a valid measure for assessing children's aggressive tendencies (Brown & Kelley, 2000).

From a practical assessment standpoint, for a large group of children, this measure may be administered either via a paper-and-pencil form (which has a built in autoscore answer form that may be analyzed after completion) or a computer form that takes approximately 10 to 20 minutes to complete. The researcher chose to use the paper-and-pencil form because it appeared to be easier to administer than trying to facilitate a process that would allow participants to enter their data into a PC. Because the scale is brief and easy to read, it may be used by virtually anyone, including respondents like the participants in this study who may be below average in reading ability.

The Normative Beliefs about Aggression Scale (NOBAGS). The Normative Beliefs about Aggression Scale (NOBAGS: Huessman, Guerra, Miller, & Zelli, 1989) is a 20-item self-report questionnaire designed to measure the perception of acceptability of behaving aggressively in a variety of conditions including provocation or no specified conditions (see Appendix C). Seven derived subscales with some overlapping items have been identified and include total approval of aggression (TAM), general approval of aggression (GAM), approval of retaliation (APR), approval of retaliation (weak and strong provocation; APW, APS), and approval of retaliation (against males and against females; ARM; ARF). Items 1-8 are brief scenarios where child A is the perpetrator of verbal aggression on child B. The scenarios vary by gender of the children involved. For example the first scenario is, "Suppose a boy says something bad to another boy, John." Item 1 addresses if it is "ok" for John to scream back and item 2 assesses if it is "ok" to hit him back. The remaining items follow the same scenario for a girl-girl interaction and a girl-boy interaction. In items 9-12, the child is presented with the same scenario (child A hits child B); however, the scenarios vary by the gender of the children involved. The last eight items do not specify gender or a scenario, but instead describe a provocative action and ask the respondent to determine how wrong or right it would be to aggress.

All items are scored on a Likert-type scale: perfectly ok = 4, sort of ok = 3, sort of wrong = 2, and really wrong = 1. The ten items whose wording includes "it is wrong" are reverse-scored in the measure itself. The scale does not inherently focus on RA; however, it includes two items that are associated with RA: "saying mean things to others" and "insulting others." For the purposes of this study, two additional statements were added to the questionnaire to assess the RA construct. These statements included,

“It is wrong to leave someone out from an activity, just to make them feel bad” and “In general, it is OK to spread rumors about someone to make other kids not like them.”

These questions were derived from Crick’s (1996) definition of RA. This scale is appropriate for the current study because it not only addresses aggression in a general sense but also includes gender-based violence items which are addressed specifically in the *Owning Up* curricula.

The norming sample of 1,150 first-grade ($n = 784$) and fourth grade students ($n = 766$) was taken from one mid-sized city and one large city approximately 40 miles apart in the Midwest region of the United States. These children were predominantly African American (38.3%), Hispanic (36.6%), and White (18.1%). However, ultimately the scale is designed to test subjects in the age range of 6 to 30 in a variety of countries and ethnic groups. High internal reliabilities for the overall score and for the subscales across gender and ethnicity were identified (see Table 2).

Table 2

Reliabilities (Cronbach’s Alphas) of the NOBAGS and Subscales (Huessman et al., 1998)

| Scale | Overall (N=1,550) | Female (N=806) | Male (N=744) | A.American (N=602) | White (N=280) | Hispanic (N=567) |
|-------|----------------------|-------------------|-----------------|-----------------------|------------------|---------------------|
| TAM | .86 | .86 | .86 | .84 | .90 | .86 |
| GAM | .80 | .79 | .80 | .80 | .83 | .79 |
| APR | .82 | .83 | .80 | .80 | .85 | .81 |
| APW | .75 | .76 | .74 | .71 | .79 | .75 |
| APS | .71 | .75 | .68 | .72 | .76 | .64 |
| ARM | .70 | .73 | .67 | .66 | .74 | .70 |
| ARF | .69 | .70 | .68 | .67 | .76 | .68 |

Note. TAM refers to total approval of aggression, GAM refers to general approval of aggression, APR refers to approval of retaliation, APW and APS refer to approval of

retaliation, weak and strong provocation, and ARM and ARF refer to approval of retaliation, against males and females.

Validity was examined on a sub-sample of the original norming sample ($N = 1,070$) by conducting an analysis comparing peer and teacher nomination scales of aggression and the NOBAGS. All scales on the NOBAGS correlated significantly with both teacher and peer reports (Williams $t(995) > 18, ps < .001$). However, the correlation between peer reports and the NOBAGS was consistently lower for girls than boys ($Z = 2.32, p < .01$, one-tailed). The stability of the NOBAGS also was addressed in this sub-sample ($N = 846$). The stabilities overall were not considered high for each subscale ($r = .20-.55, p < .001$); however, they were statistically significant. Notably, scores of the younger children (i.e., first-graders) were consistently less stable over the one-year time period.

The administration of this measure is brief and simple, taking about 10 minutes to complete. Scaled scores are the means of items on the different subscales. Therefore, each score represents the child's normative beliefs about aggression. The subscales are not all orthogonal.

Behavioral Scale-Relational Aggression.. MacDonald, D'Amico, & O'Laughlin, (2000) designed a 16-item self-report measure with two aggression subscales: RA and physical aggression. The development of a new measure was warranted based on the fairly limited measures available and the inconsistent findings in research concerning students beyond sixth grade. Participants are asked during the past two weeks how often they were involved in a relationally aggressive incident such as spreading rumors that were not true. This deviates from the typical measure that asks students about the past

six months. The scale is based on an intensity scale ranging from 1-5 (never, once, twice, 3-4 times, 5+ times). Items also include such topics as “tried to exclude someone from group activities” and “gotten into an argument with schoolmates.”

The normative sample included 6th, 7th, and 8th graders from a middle school in a moderate sized midwest urban middle school. The sample was 60.1% Caucasian, 27.4% African-American, 5.8% Hispanic, 4.3% Native American, and 2.4% Multiracial/Other. The mean age of all participants was 13.62 years old, which is similar to the mean age group of the population used in the current study.

Examination of reliability indicated a moderate to high correlation for overt, relational, and relational victimization scales ($\alpha = .81, .70, .78$, respectively). The researchers noted that the alphas were lower in the latter two subscales but acknowledge this may have been due to the small number of items that comprise the scale. In terms of validity, MacDonald et al. (2000) intended to replicate findings found by Crick and colleagues regarding the correlation between RA and overt aggression. However, their findings were inconsistent with Crick’s findings of correlational relationships between .54 and .77. MacDonald et al. (2000) found the relationship between the two variables to be considerably lower at .31, $p < .001$. The correlation of RA and RA victimization ($r = .54, p < .001$) was within the range of previous studies ($r = .45-.58, p < .001$). Additionally, the researchers found that there was a significant relationship between RA and RA victimization and negative orientation towards school, depression, and loneliness. Results also confirmed that there were no significant differences related to gender, age, or ethnic group. This tool again seems applicable to the current research because it is specifically designed for the age group that will be assessed. Also, the fact that there are

no differences across ethnic groups may make this a satisfactory assessment tool for this project, whose participants are primarily African American.

The Peer-Nomination Scales for Children. The Peer-Nomination Scales for Children is a 26-item scale used to assess a child's social behavior from the viewpoint of a child's peers (see Appendix E). Therefore, an implicit assumption in administering this test is that children chosen to rate each other should know each other well. This assumption should hold true in the current study—each student in PET knows each other based on sharing class sessions together. Additionally, the structure of middle school is such that students share similar courses together outside of the PET program. The measure contains seven scales including aggressive behavior, prosocial behavior, popularity, rejection by peers, victimization, hyperactivity, and avoidance of aggression (Huessman, 2002). For the purposes of this study, hyperactivity is not a construct that is addressed in the curricula and therefore any items pertaining to it will be eliminated prior to administration. Also, two RA questions have been added to the original measure to assess this construct. These additional questions included “Who has not been invited to parties or out-of-school activities?” and “Who has spread rumors to ruin someone's reputation?” These questions again were based on Crick's (1996) definition and theoretical model of examples of RA.

Each item on the scale takes the form of “Who does...?” Examples of questions include “Who says mean things?” and “Who pushes or shoves other kids?” Each child was prompted to think of these questions relative to the past two weeks. This was an additional instruction the researcher added to the standardized procedure. The purpose of this additional direction was to prompt students to think about each student relative to the

present time and not to a child's history of behavior. If, in fact, we wanted to ascertain a student's attitudinal or behavioral change, we needed to prompt students to refer to a specific time period. The score that results from the completed survey represents the proportion of items the child was nominated by his/her peers out of the total number of times the child could possibly have been nominated. Thus, the total number of times a child can be nominated is always the total number of peers multiplied by the number of items on the scale. Therefore, a resultant score always ranges from 0 to 1. A DOS program is available to determine the final scores for each student.

For the purposes of analysis for this study, three categories were derived based on the questions on the Peer Nomination Form. These included Positive Perpetrator (any questions related to someone who did positive things to another person), Negative Perpetrator (any questions related to a negative behavior emitted by another), and Victim (questions that elicit one to point to a victim). Each question was categorized into one of these groups, and analysis was then conducted on total sum scores in each category.

It is noted that the researcher modified the standardized instructions, as that external reinforcement to complete the tasks was not stated during the reading of the instructions for the survey because it was not necessary for this population.

According to Huessman (2002), the Peer Nomination Scales for Children has high internal consistency and high test-retest reliability over a one-month period in a variety of children in different countries and age groups. For example, a study in five countries composed of 186 to 748 children, the coefficient alpha ranged from .96 to .97 for the aggression scale (Huessman & Eron, 1986, cited in Huessman, 2002). Additionally, a large sample of inner city children exhibited equal reliability across ethnic groups. This

is promising data that supports use of the tool in the current study. Table 3 shows the internal consistencies for this scale across gender and ethnic group.

Table 3

Internal consistencies in an inner city US sample (N = 1130). Huesmann (2002).

| Sub-scale | Overall | Females | Males | Af-Amer | White | Hispanic |
|-----------|---------|---------|-------|---------|-------|----------|
| Agg | .97 | .97 | .98 | .97 | .98 | .97 |
| Pro | .92 | .91 | .91 | .91 | .93 | .91 |
| Pop | .85 | .85 | .86 | .84 | .87 | .85 |
| Rej | .92 | .92 | .93 | .93 | .96 | .89 |
| Vic | .78 | .77 | .80 | .73 | .86 | .78 |
| Hyp | .89 | .88 | .90 | .89 | .89 | .90 |

Note. Agg refers to aggression, Pro refers to prosocial, Pop refers to popularity, Rej refers to rejection, Vic refers to victim, and hyp refers to hyperactivity.

Validity studies have been conducted on the Peer Nomination Scales for Children in over 25 countries with subjects ranging from 6 to 19 years of age. Results indicate that boys tend to receive higher scores than girls on the overt aggression measures.

Additionally there is a moderate correlation between scores on this measure and reports on the CBCL-T ($r = .67$).

Consumer Satisfaction Survey-Student. The consumer satisfaction survey was used to cross-validate outcome data, interventionist checklists, and exit interviews. Assuming that the outcome data exhibited peculiar or unusual results, the survey could be a tool to qualitatively explain the results (see Appendix F). For example, students' depictions of sessions they felt were valuable or not valuable may help to explain why a certain outcome was not attained (e.g., physical aggression had no significant change). Basically, the tool served as a check and balance system to help further explain any inconsistencies in the data. By the same token, the survey intended to balance out any

bias that the facilitator may have had in her reporting. For example, a facilitator may have described the implementation of all curricula as favorable; however, students may have reported opinions very differently. Based on the aim of the program to affect students positively, this tool helped to determine consistency in perceptions of how well the intervention was implemented.

The survey consists of a total of six questions that address participants' perceptions of the program. Each question was developed based upon the interests of members at the OP-TB. The questions in the survey tried to address specific areas within the curricula that seemed to impact the children in a positive manner. A mixture of quantitative items (Likert-type scale) and qualitative items (open-ended questions) were used. The survey was designed specifically for this study and therefore there are no predetermined norms or standards for this survey.

Implementation checklist. In order to determine if the facilitators were following the intended plan for implementation, the research designed a checklist was used to measure how well the program was being implemented and if there were any inconsistencies between implementation across the two schools and/or between lessons (see Appendix G).

This checklist includes statements of activities that should have been accomplished according to the curricula in every lesson including encouraging students to participate, creating an open and honest atmosphere, etc. Additionally, the checklist prompted facilitators to reflect on any issues that arose during the 50-minute period that may have compromised the integrity of the lesson implementation. The reflection portion of the checklist also served as an avenue for the facilitator to express any concern

with the students or with her own performance. Again, the checklist was designed specifically for this study, and therefore norms and standards have not been established.

A similar checklist was used by outside observers to evaluate the quality of the facilitation (see Appendix I). This checklist was intended to estimate the level of quality with which the program was being implemented. Specifically, observers examined the facilitators' ability to lead the group effectively in behavior management, role-play, and sensitive topic discussions.

As a final safeguard to ensure that the implementation process proceeded in a smooth manner, an exit interview was conducted by the researcher individually with each facilitator (see Appendix H). The interview served as another measure to validate the overall effectiveness of implementation and to suggest areas for improvement in further implementation efforts. The interviews lasted no longer than 30 minutes each and covered a range of questions regarding process and logistical issues, perception of student commitment and involvement, reflection on one's own performance and the curricula as a whole.

Independent Measure

Instructional Process. The *Owning Up* curricula, designed by the Empower Program in Washington D.C., served as the independent measure. The role of the facilitator in this curricula is that of a guide: the facilitator must create a culture within the classroom that affords students an open arena to discuss feelings, thoughts, experiences and problems in a safe and confidential atmosphere. Facilitators are expected to get to know their students on a personal basis. This is accomplished by listening to them without interrupting their thought processes, respecting students'

varying perspectives and feelings, acknowledging silence in the classroom when difficult issues are raised, avoiding trivializing student issues, and recognizing diversity based on class, race, ethnicity, and sexuality. Additionally, building relationships with students involves allowing them to learn about the facilitator. Therefore, the facilitator provides students with his/her impressions, shares his/her own experiences, and connects with participants' feelings at their developmental level. Finally, facilitators encourage students to reach a new level of personal intimacy by avoiding teasing and personal attacks as well as bestowing ownership of what they say and do to other children.

Facilitators are expected to accomplish four things during each session:

- Write it down: Encourage students to keep a journal of thoughts and feelings during the session and after
- Set it up: Discuss topics and issues to be covered during the session and make any connections possible to previous lessons
- Wrap it up: Make time at the end of the session to discuss issues that youth experienced
- Think about it: Help students think about issues that were not resolved during class and allow them to be addressed at the conclusion of the session

These four goals are facilitated by hands-on activities including role-play, writing, question and answer sessions, and discussions. Each lesson provides for and delineates opportunities for hands-on activities. Facilitators also have a choice to hand out information on specific subject areas like sexual harassment if they find it to be appropriate.

Instructional content. The boys' curriculum is based on the assumption that the only way to help boys begin to address violence and aggression is to address it directly. The Empower Program believes that most young men are actually bystanders to violence and often struggle with speaking up against violent perpetrators. Therefore, the program examines definitions of masculinity and the paradigm that is associated with those definitions. The following outline serves to highlight key topic areas covered in each session of the boy's curriculum:

- **Are You Being Played? :** Introduces Empower and its philosophies; lays the foundation for understanding the impact of violence and gender based ideals
- **Do the Clothes Make the Man?:** Explores the relationship between clothing and images of masculinity and its effect on friendships
- **The Man: Power and Privilege:** Addresses the impact of male privilege and the costs to individual autonomy; discusses cultural power and privilege relative to racism and other forms of bigotry
- **Friendships & Boundaries:** Defines characteristics of boys' friendships, criteria boys use in choosing friends, and standards for friendships
- **Bullying:** Examines bullying by dissection of boys' personal social structures; identifies leaders, bystanders, and potential targets who facilitate the bullying cycle; discusses situations and reasons that perpetuate the cycle of vulnerability
- **What's Up with Girls?:** Establishes boys' right to be interested in the opposite sex; discusses social norms about sexual attention; addresses gender-based ideals for women's behavior towards men

- **Words of Destruction:** Explores why boys use demeaning words to put down other boys; explains how debasing language impacts women
- **Sexual Harassment:** Defines and identifies three main components of sexual harassment; intends to help students develop an understanding about some differences between male and female perceptions of sexual harassment
- **Knowing When to Date and When to Dump:** Addresses respectable ways of dealing with intimate and crush-like relationships
- **Rejection:** Teaches coping skills when one is rejected in a dating or social situation
- **Getting Intimate: Boundaries and Respect:** Challenges boys to integrate previously learned material in relation to respecting a woman's right to say "no"
- **When Relationships Go Bad:** Highlights emotional and physical manifestations of domestic violence; acknowledges the cycle of violence and the personality traits of potentially abusive people.
- **Getting Involved?:** Explains the nature of abusive relationships; discusses action plans for prevention and/or intervention
- **Physical Safety:** Provides tools for protection against physical attacks
- **Power of Apologies:** Investigates the connection between masculinity and difficulty in apologizing and forgiveness of others
- **Self Acceptance:** Provides time to reflect on personal standards, values and behavior relative to friends, intimate partners and the self

The girls' curriculum is based on the premise that the value a girl places on herself, her feelings, and her skills are key attributes that aide in empowering girls to

either remove themselves from or defend themselves in a violent situation. Therefore, the girls' program focuses more on boundary-setting, addresses more violent topics, and examines social and relational forms of aggression that are typically exhibited by girls.

An outline of the content of the girls' curriculum is as follows:

- **Getting Started & Friendships:** Similar to the boys' curricula; establishes guidelines for students; discusses friendships between girls
- **Cliques:** Examines exclusionary tactics used by girls; discusses cliques, popularity, and bystander effect
- **Reputations:** Explains the association between gossip and girls' sense of isolation and internalizing behaviors; examines how girls make concessions in their own values to maintain their social status
- **Setting Boundaries with Friends:** Challenges girls to understand the nature of confrontations; teaches effective strategies to maintain boundaries
- **We're All Beauty Queens:** Explores the societal emphasis on beauty and models of femininity; engages girls in conversation that helps to move "mental models" of beauty beyond physical appearance
- **What's Up with Guys?:** Same as the boys' curriculum, but adapted for gender appropriateness
- **You Oughta Know...About Sexual Harassment:** Same as the boys' curriculum, but adapted for gender; additional discussion on verbal and non-verbal cues and use of sexual harassment in female coercion in the workplace/school
- **Knowing How to Date and When to Dump:** Same as boys' curriculum, but adapted for gender appropriateness

- **Getting Intimate:** Continuation of the previous class; establishes importance of boundaries; empowers women to say “no” to unwanted sexual encounters
- **When Relationships Go Bad:** Same as the boys’ curriculum, but adapted for gender appropriateness
- **Getting Involved?** Same as the boys’ curriculum, but adapted for gender appropriateness
- **Your Physical Safety:** Similar to the boys’ curriculum in that it focuses on teaching skills for self defense; majority of the discussion deals with violent encounters often dealt with by girls including muggings and assaults
- **Support - Giving and Getting:** Discusses the benefits of giving and receiving healthy support and advice
- **Power of Apologies:** Same as the boys’ curriculum, but adapted for gender appropriateness
- **Self Acceptance:** Same as boys’ curriculum, but adapted for gender
- **The Last Word:** Addresses any unresolved issues from class sections; allows a time for reflection and thought

Data Analysis

Four main types of analysis were used in this study including the Pearson Product Moment Correlation, descriptions of central tendency, factorial analysis of variance, and determination of effect size for the summative dependent measures. A frequency distribution and a qualitative assessment were used for the formative evaluation measures.

Research question 1, which relates to implementation integrity, was addressed by interventionist checklists, random observations, and exit interviews. Each of these qualitative measures was reviewed for any convergent or divergent major trends. Notable findings are highlighted and discussed in Chapter 4.

Research question 2, which looked at the level of consistency across measures, was evaluated using the Pearson Product Moment Correlation Coefficient (PPMCC). Each individual's score in the pre- and post-test was used to establish a correlation coefficient. Findings are illustrated both quantitatively and pictorially with a bivariate scatterplot.

Note that the Pearson Product Moment Correlation Coefficient is one of the most common methods to ascertain a relationship between variables. It serves as an index of the relationship between variables whereby 0.00 = no relationship, to -1.00 = perfect negative correlation, to 1.00 = perfect positive correlation. The conditions for using the PPMCC are three-fold: 1) data must either interval or ratio and must be from paired observations; 2) a linear relationship should exist between variables and; 3) there are no extreme values.

Research question 3 focused on the changes between pre- and post-test scores and thus a factorial analysis of variance (ANOVA) was utilized. An ANOVA is an assessment of the interaction between two or more independent variables. It is important in factorial ANOVAs that the sample size in all cells should be the same to guarantee orthogonality. Additionally, one is unable to utilize this statistic if there are less than two people per cell. Three assumptions must be met in order to validate statistical findings: 1) the scores for each group must be normally distributed; 2) the population variances

must be equal for groups (homogeneity of variance); and 3) there must be independence of individual scores. The robust nature of ANOVAs accounts for some violation in assumptions 1 and 2. However, if assumption 3 is violated, the probabilities for the obtaining the same results again are unknown and power values are imprecise.

In order to assess if there was a significant change between pre-and post-test scores, the current study employed a 3 (School 1, 2, and 3) x 2 (male versus female) x 2 (pre-test versus post-test) factorial design. Both a within-subject and between-subject design was used to analyze the interaction. The rationale for using the within-subject design is based on the notion that participants will take repeated measurements (i.e. pre-and post-tests). A between-subject design is also warranted because participants differ at each level of the independent variable. The main strengths of the within-subject design are power and reduction in error variance linked with individual differences. The fundamental weakness of the within-subjects' design is "carryover effects." This suggests that participation in one condition may affect the outcome in the other condition. The two basic types of carryover effects are practice and fatigue (Hall, 2004). The advantage of the between-subjects design actually accounts for this limitation because each score is independent of the other scores because each participant is only measured once. The disadvantage however is that there are potential confounding variables that alter participant characteristics including environmental and individual differences (*Between Subject Designs*, 2004). Assuming significance, one will see positive changes in the treatment schools, but not the control school. Note that to further understand the results of the ANOVAs, post-hoc testing (i.e., Dunn's Test) was conducted.

Effect sizes were also computed on the significant findings from the ANOVA results. Effect size was used in the current study to quantify the true difference between pre- and post-test scores. The effect size is simply the standardized mean difference between the two groups divided by the standard deviation. An advantage of using effect size quotients is that they can be easily interpreted (Coe, 2000). One standard way of interpretation described by Cohen (1969) is as follows: effect sizes that are small =.2, medium =.5, and large =.8.

Finally in order to answer the last research question, all of the aforementioned measures were used to compare the effectiveness of the program with qualitative interpretations from facilitators and students. However, an additional measure was used in this evaluation, in which an item-by-item analysis was conducted for the Consumer Satisfaction Survey-Student. This is highlighted by both a percentage and total number frequency count.

Note that in addition to specifically answering each research question with designated analyses, the researcher also conducted measures of central tendency to understand the nature of the data. Measures that are representative of a sample or population are known as measures of central tendency. They permit objectivity when drawing conclusions or making inferences. These measures typically characterize the distribution and also identify the middle of a set of values.

The researcher in the current study obtained data on the mean, median, standard deviation, and skewness/kurtosis of individual scores on the pre- and post-test across the four summative measures. This is represented quantitatively and pictorially with box and whisker plots.

Chapter 4

Results

Overview

In this section, the researcher reports the results of both the summative and formative measures. The format of this section will include an introductory section, “Treatment of the Data,” which discusses data entry and verification techniques, as well as descriptive information in detail (i.e., central tendencies) for each specific measure. Following this preliminary section, each research question posed in Chapter 1 will be addressed based on the results found during this study. For the purposes of this chapter, please note the following: 1) “all participants” refers to the total number of students that participated in the study, 2) “complete data subset” refers to students that completed both the pre- and post-test measures, 3) “NbTotal” refers to the total score on the NOBAGS measure, 4) “RbTotal” refers to the total score on the measure that addresses RA and overt aggression, and 5) Peer Nomination Scores are separated into three categories: PNVIC+ (represents positive perpetration); PNVIC- (represents negative perpetration); PNVIC (represents victim).

Treatment of the Data

The data were entered into an excel spreadsheet by both the researcher and an undergraduate psychology student after the pre-test administration and also following the post-test administration. Each score was entered for every participant on each subscale and/or individual question so that analysis could be done per question and for the measure as a whole. Missing data were coded as a blank space in the excel document. After all the data were entered, a separate database was created for matched participants,

to ensure that the appropriate sources were pulled when analyzing all participants in the study and just those who completed the pre- and post-tests. Both the researcher and the undergraduate assistant checked the data by randomly picking various subjects' ID numbers and subsequently matching the data in the database to the paper-and-pen data filled in previously by the randomly chosen participant. Additionally, extreme values were checked across each participant for each item to ensure that the data were either a) entered correctly or b) the case was exceptional.

Measures of central tendency. Upon review of the pre-test data, it appears that the subscales that compose the AQ have roughly the same means and standard deviations (see Table 4). The scores, relative to the norming sample indicate that this population is slightly higher than the average student on various forms of aggressive behaviors. The mean for Nbttotal ($M = 42.85$) is considerably higher than Rbttotal ($M = 29.97$), which may indicate that more overt forms of aggression are present in this group. In terms of the Peer Nomination Scales, the results show that the mean for PNPR+ was significantly higher than PNPR- and PNVIC. This may indicate that the students overall viewed their peers as relatively positive in terms of how they perpetrate or victimize other students in this specific peer group. The standard deviation however for PNPR+, is larger than the other two categories, indicating a wider variability of scoring.

Table 4

Descriptive Statistics of all participants on the pre- and post-test

| Scale | <i>Pre-test</i> | | | <i>Post-test</i> | | |
|----------|-----------------|-------|-------|------------------|-------|-------|
| | N | M | SD | N | M | SD |
| AQ-Total | 64 | 52.52 | 10.11 | 45 | 53.87 | 10.97 |
| AQ-Phy | 63 | 55.00 | 11.09 | 31 | 54.48 | 11.53 |
| AQ-Ver | 64 | 50.75 | 10.71 | 45 | 54.44 | 10.46 |
| AQ-Ang | 64 | 54.06 | 9.70 | 45 | 55.18 | 8.75 |
| AQ-Hos | 64 | 51.61 | 10.25 | 45 | 51.20 | 9.74 |
| AQ-Ind | 64 | 47.59 | 10.05 | 45 | 51.84 | 11.04 |
| Nbtotal | 74 | 42.85 | 13.50 | 45 | 30.10 | 23.26 |
| RbTotal | 75 | 29.97 | 10.93 | 45 | 19.55 | 16.56 |
| PNPR+ | 75 | 42.99 | 36.44 | 45 | 47.28 | 70.97 |
| PNPR- | 75 | 26.69 | 22.58 | 45 | 29.21 | 41.96 |
| PNVIC | 75 | 16.52 | 16.47 | 45 | 16.37 | 33.35 |

The means and standard deviations for the complete data subset group basically resembles the results for all participants that took the measure (see Table 5). Thus, we may infer that the complete data subset adequately represents the entire group of participants in the study for the pre-test.

The complete data subset group post-test scores differ slightly from the all participant group data. On all measures administered (excluding Rbtotal), mean scores increased over time. However, it should be noted that the increase in mean scores ranges from very minor (.20) to significant (20). The most notable changes in mean scores found were for PNPR+ and PNPR-, and AQ-verbal. Also interesting, are the standard deviations differences between pre- and post-test for this specific group. On all measures of the Peer Nomination Scale, the standard deviations had a wide range of variability ($SD = 35-77$) as compared to the pre-test ($SD = 17-34$).

Table 5

Descriptive Statistics for complete data subset on the pre- and post-test

| Scale | <i>Pre-test</i> | | | <i>Post-test</i> | | |
|----------|-----------------|----------|-----------|------------------|----------|-----------|
| | <i>N</i> | <i>M</i> | <i>SD</i> | <i>N</i> | <i>M</i> | <i>SD</i> |
| AQ-Total | 44 | 50.02 | 11.88 | 44 | 53.57 | 10.91 |
| AQ-Phy | 43 | 53.67 | 9.57 | 43 | 54.23 | 11.64 |
| AQ-Ver | 44 | 48.84 | 10.95 | 44 | 53.98 | 10.10 |
| AQ-Ang | 44 | 52.61 | 9.30 | 44 | 54.91 | 8.66 |
| AQ-Hos | 44 | 55.09 | 9.65 | 44 | 50.84 | 9.55 |
| AQ-Ind | 44 | 46.84 | 10.53 | 44 | 51.89 | 11.17 |
| Nbtotal | 47 | 41.8 | 13.68 | 44 | 43.43 | 13.94 |
| RbTotal | 47 | 30.66 | 10.88 | 44 | 28.26 | 12.32 |
| PNPR+ | 47 | 43.30 | 34.85 | 44 | 68.09 | 77.43 |
| PNPR- | 47 | 30.11 | 22.70 | 44 | 42.34 | 45.34 |
| PNVIC | 47 | 17.43 | 16.63 | 44 | 22.00 | 35.81 |

Since the curriculum is supposed to be different, based on gender, it would appear that the data might indicate that males and females exhibit different levels of aggression and may be affected by the curriculum in different ways. Therefore, an analysis was done by gender and also by school to see if the intervention schools differed from the control schools (see Table 6). Comparing females across schools, the females in School 1 exhibited the highest means on all measures (excluding RbTotal) for the pre-test. Relatively high means for all females were found for AQ-Verbal, AQ-Physical, AQ-Anger, and PNPR+. Note that the standard deviation for PNPR+ was extremely large, ranging from 30.93-44.89. Comparing males across schools, males at School 2 showed the highest means on all the AQ subscales. Males at the control school exhibited the second highest means across the board on the AQ. Interestingly though, males at the control school exhibited the lowest means on the other summative measures. High means for School 1 and 2 were found for PNPR+ and Nbtotal. Looking at both males and females on the pre-test measures, females at School 1 scored higher on the majority of

AQ measures. However, both males and females scored relatively the same means on the NbTotal and RbTotal. Also, females tended to exhibit higher means on PNVIC; whereas males overall higher means for PNPR-.

The post-test results indicate that, like the pretest, females at School 1 exhibited the highest means compared to females and males. Overall, males at School 2 had the highest levels of aggression on the AQ compared to other males. The control school females exhibited the highest mean on NbTotal and RbTotal compared to all participants.

Table 6

Distribution of scores for complete data subset on pre- and post-test measures

Males

| Scales | <i>Pre-test</i> | | | | | | <i>Post-test</i> | | | | | |
|---------|-----------------|-----------|----------|-----------|----------|-----------|------------------|-----------|----------|-----------|----------|-----------|
| | School 1 | | 2 | | 3 | | School 1 | | 2 | | 3 | |
| | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> |
| AQTotal | 50.50 | 9.74 | 47.00 | 4.64 | 55.50 | 7.13 | 54.75 | 14.06 | 43.40 | 7.30 | 60.63 | 4.34 |
| RbTotal | 44.50 | 17.71 | 21.86 | 4.67 | 35.00 | 10.23 | 37.50 | 18.14 | 18.29 | 8.38 | 34.88 | 10.86 |
| NbTotal | 40.50 | 16.03 | 41.86 | 8.23 | 42.38 | 11.92 | 56.50 | 22.31 | 36.57 | 18.10 | 48.63 | 12.86 |
| PNPR+ | 41.00 | 46.27 | 31.00 | 14.25 | 50.88 | 51.04 | 53.00 | 58.66 | 53.00 | 40.39 | 123.38 | 125.06 |
| PNPR- | 17.50 | 12.15 | 28.57 | 18.97 | 33.50 | 20.87 | 29.75 | 10.81 | 46.14 | 62.93 | 62.63 | 48.27 |
| PNVIC | 14.50 | 12.07 | 10.00 | 3.16 | 13.50 | 15.05 | 20.50 | 24.78 | 23.00 | 40.60 | 47.38 | 68.26 |

Females

| Scales | <i>Pre-test</i> | | | | | | <i>Post-test</i> | | | | | |
|---------|-----------------|-----------|----------|-----------|----------|-----------|------------------|-----------|----------|-----------|----------|-----------|
| | School 1 | | 2 | | 3 | | School 1 | | 2 | | 3 | |
| | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> |
| AQTotal | 40.44 | 15.65 | 54.60 | 12.54 | 50.40 | 11.74 | 46.67 | 11.57 | 58.80 | 10.71 | 55.20 | 9.98 |
| RbTotal | 31.20 | 7.58 | 27.18 | 7.15 | 32.00 | 14.06 | 31.80 | 13.72 | 25.55 | 9.80 | 25.00 | 8.88 |
| NbTotal | 38.70 | 15.25 | 43.20 | 18.72 | 43.83 | 12.58 | 39.50 | 10.49 | 46.10 | 9.49 | 38.33 | 12.69 |
| PNPR+ | 30.40 | 14.01 | 65.82 | 35.05 | 33.83 | 34.25 | 18.40 | 12.45 | 115.09 | 72.33 | 24.33 | 27.78 |
| PNPR- | 20.10 | 9.89 | 51.09 | 27.80 | 16.00 | 17.37 | 17.40 | 13.38 | 70.91 | 51.18 | 10.33 | 12.94 |
| PNVIC | 16.70 | 9.88 | 32.73 | 23.87 | 7.83 | 8.01 | 6.50 | 5.21 | 27.36 | 21.60 | 6.17 | 5.70 |

Due to the attrition rate between pre- and post-test, the extent to which the means are biased between the total number of participants and the complete data subset was also computed by determining the effect sizes, by gender for both the pre- and post-tests. The

post-test scores for males exhibit large effect sizes for Nbtotal, Rbtotal, PNPR+, and PNPR-. Overall though, the effect sizes are relatively small. The effect sizes (see Table 7) and stem and leaf plots are presented (Figure 1).

Table 7

Effect sizes for female and male participants

| Scale | <i>Females</i> | | <i>Males</i> | |
|----------|----------------|-----------|--------------|-----------|
| | Pre-test | Post-test | Pre-test | Post-test |
| AQ-Total | .24 | 0.05 | 0 | 0 |
| AQ-Phy | .19 | 0.03 | 0.03 | 0 |
| AQ-Ver | .18 | 0.08 | 0.20 | 0 |
| AQ-Ang | .25 | 0.05 | 0 | 0 |
| AQ-Hos | .30 | 0.06 | -0.04 | 0 |
| AQ-Ind | .15 | 0 | -0.02 | 0 |
| Nbtotal | 0 | -0.50 | 0.19 | -0.67 |
| RbTotal | .05 | -0.45 | -0.19 | -0.62 |
| PNPR+ | .14 | -0.28 | -0.20 | -0.34 |
| PNPR- | .02 | -0.28 | -0.07 | -0.37 |
| PNVIC | -0.05 | -0.28 | -0.01 | -0.19 |

Figure 1. Stem and leaf representation of effect sizes by gender across time.

| Pre-test, females | | Post-test, females | |
|-------------------|--------------------|--------------------|-------------|
| Frequency | Stem & Leaf | Frequency | Stem & Leaf |
| 2.00 | -0 .05 | 1.00 | -.05 |
| 4.00 | 0.13489 | 1.00 | -0.45 |
| 5.00 | 0 .25 | 1.00 | -.03 |
| | | 3.00 | -.2788 |
| | | 1.00 | 0 |
| | | 5.00 | 0 .024457 |
| Pretest, males | | Post-test, males | |
| Frequency | Stem & Leaf | Frequency | Stem & Leaf |
| 2.00 | Extremes (= <-.19) | 2.00 | -0 .66 |
| 1.00 | -0.7 | 2.00 | -0.33 |
| 3.00 | -0.124 | 7.00 | -0.0000001 |
| 3.00 | 0.003 | | |
| 2.00 | Extremes (>=.19) | | |

In line with describing the data, the researcher also evaluated the normality of each of the scores on the subscales and/or by total score per measure. Skewness and kurtosis were evaluated to determine any degree of asymmetry in the data and to ascertain if there were any major peaks or tails in the distribution of the data (refer to Tables 8 and 9). The pre-test data for all participants indicate that there are three items that are slightly negatively skewed (AQ-Total, AQ-Ver, and AQ-Ind). Additionally, there is a slight positive skew for all the categories on the Peer Nomination Scale. In terms of kurtosis, AQ-Phy, PNPR+, and PNVIC would be considered to be moderately leptokurtic. Also, AQ-Hos and AQ-Ind would be considered to be platykurtic with less outlying values than a normal distribution.

Similar to the pre-test for all participants, all the categories for the Peer Nomination Scale had a slight positive skew. However, in the post-test there are more items that could be considered negatively skewed including AQ-Total, AQ-Ver, AQ-Hos, and Nbttotal. The kurtosis value of PNVIC in the post-test is extremely notable (17.422) in addition to PNPR+ (6.468). Additionally all of the subscales of the AQ, Nbttotal, and Rbttotal have negative kurtosis values indicating again that there are less outlying values than the norm. The matched group participants follow a similar pattern of distribution between pre- and post-tests; however, the main difference is that during the pre-test all of the scores on the subscales of the AQ had a negative skew.

Table 8

Distribution of scores for all participants on pre- and post-test measures

| Scale | <i>Pre-test</i> | | | | <i>Post-test</i> | | | |
|----------|-----------------|------------|----------|------------|------------------|------------|----------|------------|
| | Skew | | Kurtosis | | Skew | | Kurtosis | |
| | <i>S</i> | <i>SEM</i> | <i>S</i> | <i>SEM</i> | <i>S</i> | <i>SEM</i> | <i>S</i> | <i>SEM</i> |
| AQ-Total | -.51 | .30 | .41 | .59 | -.08 | .35 | -.72 | .70 |
| AQ-Phy | .77 | .30 | 2.93 | .60 | .10 | .42 | -.74 | .82 |
| AQ-Ver | -.57 | .30 | .33 | .59 | -.44 | .35 | -.07 | .70 |
| AQ-Ang | .07 | .30 | .20 | .59 | -.04 | .35 | -.09 | .70 |
| AQ-Hos | .08 | .30 | -.31 | .59 | -.26 | .35 | -.37 | .70 |
| AQ-Ind | -.43 | .30 | -.52 | .59 | .02 | .35 | -.72 | .70 |
| Nbtotal | .13 | .28 | .49 | .55 | -.17 | .29 | -1.24 | .56 |
| RbTotal | .56 | .28 | .67 | .55 | .32 | .29 | -.60 | .56 |
| PNPR+ | 1.45 | .28 | 1.80 | .55 | 2.24 | .29 | 6.47 | .56 |
| PNPR- | 1.15 | .28 | .81 | .55 | 1.59 | .29 | 1.74 | .56 |
| PNVIC | 2.23 | .28 | 7.52 | .55 | 3.78 | .29 | 17.42 | .56 |

Table 9

Distribution of scores for complete data subset on pre- and post-test measures

| Scale | <i>Pre-test</i> | | | | <i>Post-test</i> | | | |
|----------|-----------------|------------|----------|------------|------------------|------------|----------|------------|
| | Skew | | Kurtosis | | Skew | | Kurtosis | |
| | <i>S</i> | <i>SEM</i> | <i>S</i> | <i>SEM</i> | <i>S</i> | <i>SEM</i> | <i>S</i> | <i>SEM</i> |
| AQ-Total | -1.25 | .36 | 2.33 | .70 | -.04 | .36 | -.67 | .70 |
| AQ-Phy | -.15 | .36 | -.13 | .71 | .16 | .43 | -.74 | .83 |
| AQ-Ver | -.70 | .36 | -.18 | .70 | -.56 | .36 | -.09 | .70 |
| AQ-Ang | -.09 | .36 | .42 | .70 | -.01 | .36 | .02 | .70 |
| AQ-Hos | .06 | .36 | .01 | .70 | -.28 | .36 | -.32 | .70 |
| AQ-Ind | -.28 | .36 | -.59 | .70 | .01 | .36 | -.77 | .70 |
| Nbtotal | .35 | .40 | 1.07 | .69 | -.21 | .35 | 1.81 | .68 |
| RbTotal | .86 | .40 | .45 | .68 | .57 | .35 | .99 | .68 |
| PNPR+ | 1.21 | .40 | .92 | .68 | 1.89 | .35 | 4.76 | .68 |
| PNPR- | 1.04 | .40 | .49 | .68 | 1.13 | .35 | .36 | .68 |
| PNVIC | 2.62 | .40 | 11.00 | .68 | 3.67 | .35 | 16.75 | .68 |

The researcher also wanted to pictorially represent these data using box and whisker plots. By doing this, the reader is able to quickly see how the distribution is arranged via quartiles, as well as making handy comparisons between the data. Each box

and whisker plot below represents each individual score relative to each of the four measures. Note that for the purposes of this representation, all participants will be used to depict the overall distribution of pre- and post-test measures. Scores on the AQ pre-test and post-test are fairly homogenous as indicated by the median line in the same place for the most part across scales. In the pre-test there are a total of four outlier points and none exist in the post-test. For the MacDonald et al. (2000) measure, the pretest exhibits more outlier scores. Additionally, the median is higher on the pre-test than the post-test. The post-test, contrary to the pre-test only has one whisker that extends upwards. Just like the MacDonald et al. (2000) measure, the NOBAGS measure has a higher median and a number of outliers on the pre-test; whereas, the post-test consists of zero outliers, a smaller median, and an extended whisker. The categories of the Peer Nomination Scale have similar distributions. No major differences exist for the Peer Nomination Scales except that during the post-test, there are more outliers and extremes for PNVIC. Figures 2-9 provide specific details of each measure.

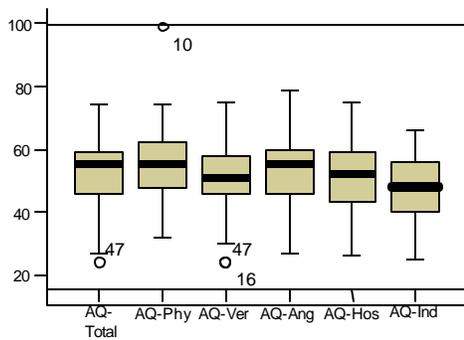


Figure 2. Box and whisker plot for pre-test AQ scales.

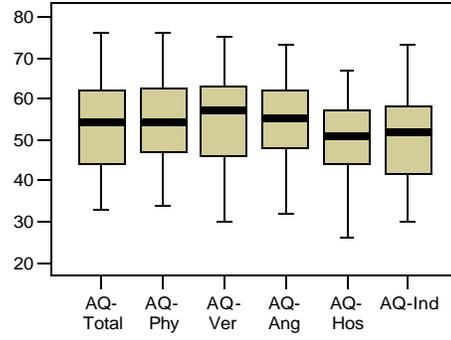


Figure 3. Box and whisker plot for post-test AQ scales.

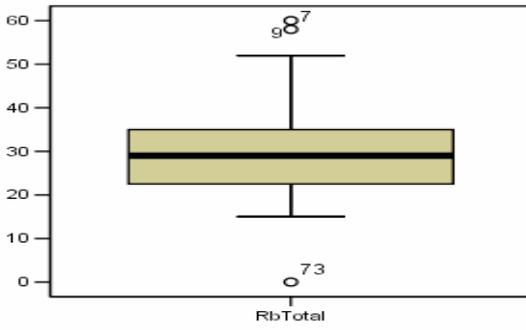


Figure 4. Box and whisker plot for pre-test MacDonald et al., 2000.

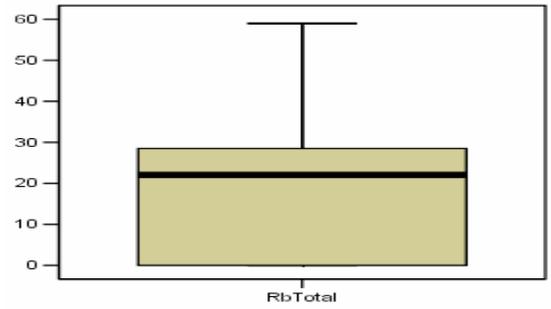


Figure 5. Box and whisker plot for post-test MacDonald et al., 2000.

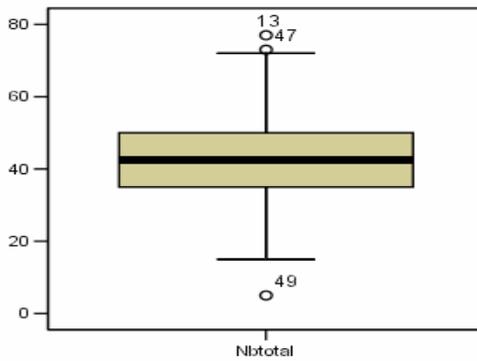


Figure 6. Box and whisker plot for pre-test NOBAGS scales.

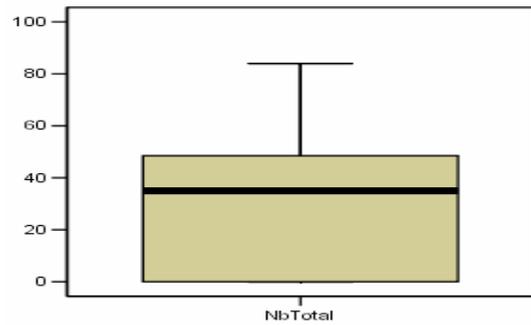


Figure 7. Box and whisker plot for post-test NOBAGS scales.

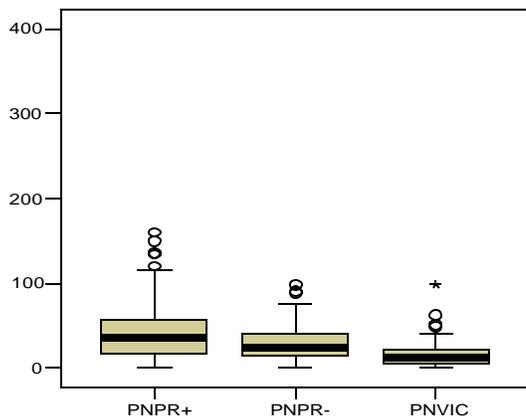


Figure 8. Box and whisker plots for pre-test Peer Nomination Scale.

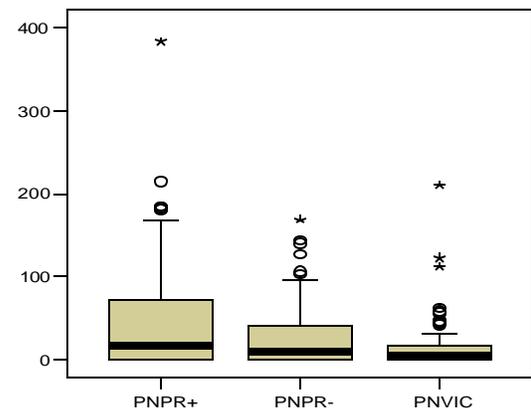


Figure 9. Box and whisker plots for post-test Peer Nomination Scale- Post-test.

Data Quality Checks-Cronbach's alpha. In order to determine if there was any evidence of items/subscales not measuring the same underlying construct, the researcher assessed the reliability of items using Cronbach's alpha. More specifically, it is very important to know whether the same set of items would elicit the same responses assuming that the same questions were re-administered to the same respondents. Variables derived from test instruments are declared to be reliable only when they provide stable and reliable responses over a repeated administration of the test. Overall, the Cronbach's alpha for each measure was high for both the pre-test and post-test (NOBAGS: $a = .912$, $a = .918$; Peer Nomination: $a = .747$, $a = .848$; AQ-Total: $a = .918$, $a = .938$, RelBeh: $a = .838$, $a = .832$). According to the analyses run, there were only two items that did not represent a high reliability. These items included RelBeh12 ($r = .068$) and RelBeh13 ($r = .100$). Both of these items represent extreme violence and therefore may have accounted for the low reliability. For example, RelBeh12 asks the student to circle the number of times he/she has "been in a fight with members of a gang" and RelBeh13 states, "Carried a weapon, like a knife or gun." For the purposes of future analyses, these items were dropped from the study. Please refer to Table 10 for specific information on item-total correlations for the pre- and post-test measures.

Table 10

Item-Total Statistics

MacDonald et al, 2000

| <i>Pretest</i> | | | | | | <i>Post-test</i> | | | | | |
|----------------|----------|-----------|---------------------------------------|-----------------|--|------------------|----------|-----------|---------------------------------------|-----------------|--|
| <u>Item</u> | <u>M</u> | <u>SD</u> | <u>Corrected</u> <u>Item Total</u> | <u>R-Square</u> | <u>Alpha</u> <u>if Item</u> <u>Deleted</u> | <u>Item</u> | <u>M</u> | <u>SD</u> | <u>Corrected</u> <u>Item Total</u> | <u>R-Square</u> | <u>Alpha</u> <u>if Item</u> <u>Deleted</u> |
| 1 | 1.55 | 1.03 | .30 | .46 | .82 | 1 | 1.47 | .93 | .43 | .63 | .82 |
| 2 | 1.83 | 1.13 | .42 | .35 | .81 | 2 | 1.83 | 1.20 | .36 | .53 | .83 |
| 3 | 1.79 | 1.26 | .52 | .54 | .80 | 3 | 1.47 | .80 | .50 | .55 | .82 |
| 4 | 1.25 | .72 | .33 | .44 | .82 | 4 | 1.36 | .61 | .33 | .59 | .83 |
| 5 | 2.39 | 1.50 | .36 | .38 | .80 | 5 | 2.28 | 1.46 | .42 | .68 | .82 |
| 6 | 2.44 | 1.40 | .57 | .52 | .80 | 6 | 2.51 | 1.21 | .30 | .35 | .83 |
| 7 | 1.75 | 1.25 | .59 | .66 | .81 | 7 | 1.51 | .91 | .55 | .70 | .82 |
| 8 | 2.19 | 1.41 | .47 | .63 | .81 | 8 | 1.89 | 1.05 | .61 | .62 | .82 |
| 9 | 1.69 | 1.23 | .48 | .51 | .81 | 9 | 1.26 | .64 | .39 | .39 | .83 |
| 10 | 2.05 | 1.26 | .47 | .66 | .81 | 10 | 1.81 | 1.28 | .75 | .85 | .81 |
| 11 | 2.37 | 1.44 | .40 | .87 | .81 | 11 | 2.52 | 1.50 | .67 | .93 | .81 |
| 12 | 1.72 | .96 | .07 | .23 | .83 | 12 | 1.55 | 1.08 | .33 | .48 | .83 |
| 13 | 2.04 | 1.05 | .10 | .23 | .83 | 13 | 1.17 | .53 | .33 | .48 | .83 |
| 14 | 2.28 | 1.35 | .45 | .84 | .81 | 14 | 2.36 | 1.39 | .55 | .90 | .82 |
| 15 | 1.47 | 1.46 | .31 | .28 | .82 | 15 | 1.40 | .77 | .56 | .77 | .82 |
| 16 | 3.01 | 1.53 | .51 | .44 | .80 | 16 | 3.21 | 1.37 | .64 | .67 | .81 |
| 17 | 2.03 | 1.46 | .54 | .62 | .80 | 17 | 2.70 | 3.05 | .41 | .66 | .86 |

NOBAGS

| <i>Pretest</i> | | | | | | <i>Post-test</i> | | | | | |
|----------------|----------|-----------|---------------------------------------|-----------------|--|------------------|----------|-----------|---------------------------------------|-----------------|--|
| <u>Item</u> | <u>M</u> | <u>SD</u> | <u>Corrected</u> <u>Item Total</u> | <u>R-Square</u> | <u>Alpha</u> <u>if Item</u> <u>Deleted</u> | <u>Item</u> | <u>M</u> | <u>SD</u> | <u>Corrected</u> <u>Item Total</u> | <u>R-Square</u> | <u>Alpha</u> <u>if Item</u> <u>Deleted</u> |
| 1 | 2.40 | 1.00 | .58 | .75 | .91 | 1 | 2.44 | .99 | .61 | .77 | .91 |
| 2 | 1.81 | .97 | .82 | .87 | .90 | 2 | 1.75 | .89 | .61 | .77 | .91 |
| 3 | 2.87 | 1.10 | .55 | .75 | .91 | 3 | 2.88 | .98 | .61 | .71 | .91 |
| 4 | 1.96 | 1.07 | .46 | .68 | .91 | 4 | 2.17 | 1.14 | .60 | .82 | .91 |
| 5 | 2.51 | .96 | .69 | .67 | .91 | 5 | 2.42 | .92 | .72 | .87 | .91 |
| 6 | 1.84 | .96 | .85 | .91 | .90 | 6 | 1.79 | .87 | .68 | .84 | .91 |
| 7 | 1.93 | 1.04 | .66 | .58 | .91 | 7 | 1.81 | .87 | .35 | .75 | .92 |
| 8 | 1.20 | .64 | .47 | .58 | .91 | 8 | 1.21 | .58 | .40 | .72 | .92 |
| 9 | 3.01 | 1.15 | .53 | .89 | .91 | 9 | 3.00 | .99 | .48 | .77 | .92 |
| 10 | 3.04 | 1.21 | .46 | .70 | .91 | 10 | 3.15 | 1.15 | .47 | .76 | .92 |
| 11 | 2.97 | 1.12 | .61 | .87 | .91 | 11 | 2.92 | 1.13 | .59 | .85 | .91 |
| 12 | 1.74 | 1.05 | .41 | .41 | .91 | 12 | 1.35 | .84 | .26 | .57 | .92 |

| | | | | | | | | | | | |
|----|------|------|-----|-----|-----|----|------|------|-----|-----|-----|
| 13 | 2.07 | .96 | .41 | .39 | .91 | 13 | 2.02 | .91 | .58 | .70 | .91 |
| 14 | 1.69 | .88 | .45 | .52 | .91 | 14 | 1.83 | .91 | .55 | .70 | .92 |
| 15 | 1.76 | .95 | .54 | .63 | .91 | 15 | 1.71 | .77 | .73 | .89 | .91 |
| 16 | 1.46 | .80 | .58 | .69 | .91 | 16 | 1.54 | .68 | .70 | .83 | .91 |
| 17 | 1.59 | .84 | .53 | .65 | .91 | 17 | 1.69 | .80 | .65 | .70 | .91 |
| 18 | 1.53 | .85 | .38 | .61 | .91 | 18 | 1.79 | .87 | .52 | .65 | .92 |
| 19 | 2.12 | 1.04 | .65 | .72 | .91 | 19 | 2.17 | 1.02 | .66 | .69 | .91 |
| 20 | 1.55 | .91 | .55 | .72 | .91 | 20 | 1.81 | 1.05 | .49 | .64 | .92 |
| 21 | 1.59 | .86 | .47 | .63 | .91 | 21 | 1.56 | .80 | .69 | .75 | .91 |
| 22 | 1.22 | .57 | .22 | .43 | .91 | 22 | 1.52 | .80 | .46 | .80 | .92 |

Peer Nomination Inventory-PNPR+

| <i>Pretest</i> | | | | | | <i>Post-test</i> | | | | | |
|----------------|-------|-------|-------------------------|----------|-----------------------------|------------------|-------|-------|-------------------------|----------|-----------------------------|
| Item | M | SD | Corrected Item Total | R-Square | Alpha if Item Deleted | Item | M | SD | Corrected Item Total | R-Square | Alpha if Item Deleted |
| 1 | 7.93 | 6.82 | .50 | .33 | .66 | 1 | 12.70 | 12.87 | .69 | .49 | .75 |
| 4 | 10.45 | 10.03 | .58 | .43 | .61 | 4 | 15.84 | 15.67 | .60 | .40 | .79 |
| 10 | 6.74 | 8.95 | .48 | .30 | .67 | 10 | 9.58 | 12.14 | .60 | .39 | .79 |
| 18 | 4.96 | 5.18 | .54 | .37 | .67 | 18 | 11.98 | 16.29 | .68 | .50 | .75 |

Peer Nomination Inventory-PNPR-

| <i>Pretest</i> | | | | | | <i>Post-test</i> | | | | | |
|----------------|------|------|-------------------------|----------|-----------------------------|------------------|-------|-------|-------------------------|----------|-----------------------------|
| Item | M | SD | Corrected Item Total | R-Square | Alpha if Item Deleted | Item | M | SD | Corrected Item Total | R-Square | Alpha if Item Deleted |
| 2 | 8.47 | 9.25 | .75 | .74 | .82 | 2 | 10.84 | 12.90 | .73 | .81 | .85 |
| 3 | 2.35 | 3.18 | .15 | .33 | .86 | 3 | 2.16 | 5.84 | .61 | .77 | .87 |
| 7 | 6.59 | 6.38 | .87 | .89 | .81 | 7 | 10.74 | 13.92 | .55 | .83 | .87 |
| 8 | 2.74 | 4.75 | .74 | .66 | .82 | 8 | 5.65 | 11.57 | .51 | .77 | .87 |
| 9 | 5.92 | 7.20 | .60 | .70 | .83 | 9 | 8.65 | 10.85 | .79 | .84 | .85 |
| 13 | 1.93 | 2.57 | .50 | .55 | .85 | 13 | 3.33 | 6.27 | .77 | .81 | .86 |
| 14 | 3.51 | 4.76 | .57 | .66 | .84 | 14 | 6.91 | 12.21 | .39 | .36 | .88 |
| 15 | 4.03 | 4.48 | .34 | .55 | .85 | 15 | 6.65 | 8.34 | .71 | .72 | .86 |
| 16 | 6.31 | 9.64 | .83 | .80 | .82 | 16 | 12.49 | 15.84 | .64 | .85 | .87 |
| 17 | 1.71 | 5.07 | .04 | .58 | .86 | 17 | 3.88 | 9.05 | .47 | .68 | .87 |
| 22 | 2.78 | 3.74 | .59 | .67 | .84 | 22 | 4.91 | 7.92 | .66 | .70 | .86 |

Peer Nomination Inventory-PNVIC

| <i>Pretest</i> | | | | | | <i>Post-test</i> | | | | | |
|----------------|------|------|-------------------------|----------|-----------------------------|------------------|------|-------|-------------------------|----------|-----------------------------|
| Item | M | SD | Corrected Item Total | R-Square | Alpha if Item Deleted | Item | M | SD | Corrected Item Total | R-Square | Alpha if Item Deleted |
| 5 | 2.73 | 6.31 | .43 | .29 | .27 | 5 | 2.07 | 5.36 | .72 | .79 | .70 |
| 6 | 5.66 | 8.27 | .16 | .07 | .51 | 6 | 8.74 | 13.19 | .35 | .34 | .77 |
| 1 | 2.04 | 2.79 | .44 | .29 | .38 | 1 | 2.60 | 6.33 | .57 | .67 | .71 |
| 1 | 1.17 | 2.01 | .08 | .11 | .48 | 1 | 3.70 | 8.43 | .44 | .35 | .73 |
| 1 | 2.49 | 3.17 | .29 | .15 | .41 | 1 | 3.81 | 9.85 | .58 | .71 | .69 |
| 2 | 2.42 | 4.09 | .22 | .13 | .43 | 2 | 6.09 | 13.23 | .60 | .62 | .69 |

AQ-Phy

| <i>Pretest</i> | | | | | | <i>Post-test</i> | | | | | |
|----------------|------|------|-------------------------|----------|-----------------------------|------------------|------|------|-------------------------|----------|-----------------------------|
| Item | M | SD | Corrected Item Total | R-Square | Alpha if Item Deleted | Item | M | SD | Corrected Item Total | R-Square | Alpha if Item Deleted |
| 8 | 2.63 | 1.53 | .64 | .45 | .82 | 8 | 2.63 | 1.55 | .87 | .89 | .90 |
| 10 | 2.41 | 1.48 | .52 | .39 | .83 | 10 | 2.68 | 1.56 | .73 | .70 | .91 |
| 11 | 3.39 | 1.64 | .61 | .42 | .82 | 11 | 3.08 | 1.68 | .80 | .78 | .90 |
| 17 | 2.75 | 1.57 | .56 | .36 | .83 | 17 | 2.68 | 1.69 | .62 | .47 | .92 |
| 23 | 2.44 | 1.53 | .66 | .58 | .81 | 23 | 2.63 | 1.70 | .79 | .71 | .90 |
| 24 | 1.89 | 1.34 | .58 | .49 | .82 | 24 | 2.18 | 1.49 | .76 | .64 | .91 |
| 25 | 4.16 | 1.40 | .49 | .27 | .84 | 25 | 4.13 | 1.44 | .61 | .60 | .92 |
| 27 | 3.41 | 1.52 | .54 | .34 | .83 | 27 | 3.21 | 1.53 | .67 | .62 | .91 |

AQ-Ver

| <i>Pretest</i> | | | | | | <i>Post-test</i> | | | | | |
|----------------|------|------|-------------------------|----------|-----------------------------|------------------|------|------|-------------------------|----------|-----------------------------|
| Item | M | SD | Corrected Item Total | R-Square | Alpha if Item Deleted | Item | M | SD | Corrected Item Total | R-Square | Alpha if Item Deleted |
| 1 | 2.17 | 1.14 | .52 | .29 | .69 | 1 | 2.53 | 1.40 | .61 | .43 | .78 |
| 4 | 2.70 | 1.26 | .53 | .40 | .68 | 4 | 3.05 | 1.31 | .59 | .46 | .79 |
| 6 | 2.41 | 1.35 | .48 | .30 | .71 | 6 | 2.87 | 1.38 | .68 | .47 | .76 |
| 20 | 2.97 | 1.44 | .51 | .36 | .69 | 20 | 3.21 | 1.51 | .49 | .36 | .82 |
| 26 | 3.42 | 1.42 | .50 | .40 | .69 | 26 | 3.63 | 1.32 | .70 | .51 | .76 |

AQ-Ang

Pretest

Post-test

| Item | M | SD | Corrected Item Total | R-Square | Alpha if Item Deleted | Item | M | SD | Corrected Item Total | R-Square | Alpha if Item Deleted |
|------|------|------|-------------------------|----------|-----------------------------|------|------|------|-------------------------|----------|-----------------------------|
| 3 | 3.13 | 1.55 | .18 | .08 | .73 | 3 | 3.42 | 1.29 | .18 | .21 | .72 |
| 7 | 2.41 | 1.56 | .54 | .38 | .63 | 7 | 2.79 | 1.60 | .59 | .43 | .62 |
| 12 | 3.31 | 1.61 | .53 | .33 | .63 | 12 | 3.03 | 1.67 | .70 | .60 | .58 |
| 19 | 3.34 | 1.46 | .36 | .25 | .68 | 19 | 3.26 | 1.47 | 0 | .20 | .76 |
| 22 | 2.34 | 1.41 | .41 | .21 | .67 | 22 | 2.32 | 1.34 | .45 | .38 | .66 |
| 29 | 3.06 | 1.53 | .36 | .28 | .68 | 29 | 3.03 | 1.65 | .59 | .45 | .62 |
| 32 | 2.06 | 1.37 | .53 | .32 | .64 | 32 | 2.18 | 1.43 | .41 | .48 | .67 |

AQ-Hos

Pretest

Post-test

| Item | M | SD | Corrected Item Total | R-Square | Alpha if Item Deleted | Item | M | SD | Corrected Item Total | R-Square | Alpha if Item Deleted |
|------|------|------|-------------------------|----------|-----------------------------|------|------|------|-------------------------|----------|-----------------------------|
| 2 | 2.22 | 1.25 | .46 | .26 | .73 | 2 | 2.32 | 1.44 | .52 | .33 | .77 |
| 5 | 2.02 | 1.11 | .47 | .40 | .73 | 5 | 2.47 | 1.35 | .62 | .43 | .75 |
| 9 | 2.69 | 1.46 | .53 | .38 | .71 | 9 | 2.87 | 1.46 | .59 | .43 | .76 |
| 16 | 3.17 | 1.43 | .43 | .31 | .73 | 16 | 2.79 | 1.42 | .57 | .46 | .76 |
| 21 | 2.98 | 1.60 | .54 | .45 | .71 | 21 | 2.63 | 1.63 | .56 | .39 | .76 |
| 28 | 3.52 | 1.62 | .47 | .28 | .73 | 28 | 3.82 | 1.57 | .29 | .25 | .80 |
| 32 | 2.72 | 1.63 | .53 | .41 | .71 | 32 | 2.58 | 1.24 | .52 | .38 | .77 |
| 33 | 1.73 | 1.23 | .20 | .15 | .77 | 33 | 2.21 | 1.49 | .37 | .38 | .79 |

AQ-Ind

Pretest

Post-test

| Item | M | SD | Corrected Item Total | R-Square | Alpha if Item Deleted | Item | M | SD | Corrected Item Total | R-Square | Alpha if Item Deleted |
|------|------|------|-------------------------|----------|-----------------------------|------|------|------|-------------------------|----------|-----------------------------|
| 13 | 1.42 | .92 | .29 | .10 | .54 | 13 | 1.92 | 1.48 | .55 | .40 | .63 |
| 14 | 3.08 | 1.49 | .43 | .22 | .47 | 14 | 3.18 | 1.60 | .55 | .34 | .63 |
| 15 | 2.63 | 1.60 | .41 | .22 | .47 | 15 | 3.16 | 1.55 | .14 | .14 | .76 |
| 18 | 1.52 | .93 | .32 | .11 | .54 | 18 | 2.13 | 1.49 | .35 | .16 | .69 |
| 30 | 2.67 | 1.60 | .24 | .07 | .57 | 30 | 3.03 | 1.35 | .44 | .30 | .66 |
| 34 | 2.75 | 1.43 | .22 | .06 | .57 | 34 | 2.95 | 1.56 | .63 | .45 | .60 |

Question 1: How well was the Owing Up curricula implemented? A) Were there any major issues that could have impacted outcomes at pre- and post-test measure administrations? B) Did implementation differ across schools? Across facilitators?

While this was not technically designated as a pilot study, it is important to remember that this was the first time that OP-TB implemented this program in the public school system. Therefore, while there were many attempts to standardize and streamline the implementation process, there were still inconsistencies with how the process transpired at each school. It may be speculated that implementation did not occur with a high degree of integrity for various reasons. First, this implementation was the first time that OP-TB worked with the public school sector to facilitate a program. Therefore, OP-TB may have not completely understood how to interact with the public schools. Also, relationships forged with the public schools may have not been as strong as in the private sector. Second, one of the facilitators was considered a novice in relation to the number of times she led the program, and therefore she may have encountered difficulties that affected integrity. Finally, lack of integrity may have been accounted for by a potential misunderstanding on the expectations held by the researcher and OP-TB.

There were a few major issues that could have impacted the outcome findings during the pre- and post-test. The first major issue is the level of confidentiality that the students really felt while filling out the Peer Nomination Scales. Although they had their own tables and were instructed to keep their eyes to themselves, students tried to whisper to each other and continually moved their chairs to see what their friends were doing. This happened at both sites and occurred during both the pre- and post-test administrations.

Another major issue that could have impacted the students during the post-test at School 2 was their attention levels on completing the measures. According to one of the exit interviews, a facilitator noted that the individual boys groups were combined into one large group for the post-test. This was a decision made by administration. She noted that the boys kept talking to each other because they were excited to be in a large group and continually spoke about the end of the school year, regardless of the proctor asking them to stop.

School 1 also experienced some issues that may have impacted how students performed on the post-test. The first major issue according to the attendance data was that student attendance rates were highly variable. Sometimes only three students attended class and at other times up to eight students were in class. Additionally, students that did attend class were not consistently the same students. The facilitator noted that some of the students that participated in the pre-test and post-test only received three of the intervention classes. The second major issue is that the students that did attend the classes only received a total of six classes due to logistical and communicative problems onsite.

Finally, according to all of the observers, it was noted that the students had difficulty with some of the language and abstract concepts that the curricula tried to teach. For example, when the words “passive” and “associate” were used, students started to get frustrated and call out and say, “These words don’t make sense to us.” They noted that the interest of all of the groups waxed and waned throughout each of the classes they observed. Also, the scores specified by the observers during random observation for the facilitators differed substantially. While the facilitator at School 2

received a mean score of 4.875¹, the facilitator at School 1 received an overall observation score of 3.45. It is unclear whether these scores reflect a) if the observers had different standards in scoring, b) if the populations were more or less difficult to teach, or c) facilitator reactivity.

Clearly the major difference between implementation across schools was the level of commitment and consistency by the staff on site. While School 1 had fairly inadequate structures in place to promote the students to attend (i.e., disorganization of classrooms, lack of communication between SIS and OP-TB, etc.), School 2 had a 92% attendance rate of its students. Attendance data are not available for School 1 since the data were not collected. The second major difference was the activity and participation level of the students in the class. In School 2, the participation and interest level of the students averaged 4.16 out of 5.00. Although data are not available for School 1 because of lack of collection, during the exit interview it was noted that the girls participated more than the boys and seemed to take the lessons more seriously. Additionally, it was noted by both the observers and the facilitator that many students came in half-way through or very late, and missed crucial parts of the lesson plan.

Another major difference between the two implementation sites was the time it took to break down the barriers between the facilitators and the students. One facilitator noted that there was a level of difficulty for her to relate to the students and vice-versa because of ethnic and cultural differences. She felt that this impacted how open the students were to what she had to say about *Owning Up*. On the other hand, the other facilitator did not note any issues related to problems with race and/or identity.

¹ Note that these scores are derived from observational data by which raters scored each facilitator on a scale from 1-5, whereby 5 is the highest score possible.

The final difference between how the program was implemented was the approach in which each facilitator taught the curricula. In School 1, the facilitator noted that she approached the students from a very nurturing standpoint. She also noted that the curricula and some members of the Ophelia team preferred utilizing a more direct approach when working with the students. Based on the sensitive nature of the topics, it is unclear which approach may work better; however, clearly, relationships forged and the depth of discussion is partially dependent upon how comfortable the students felt with the facilitator's teaching style.

Question 2: What level of consistency does the multi-informant approach provide for the evaluation (i.e., what is the relationship between measures)?

In order to assess the relationship between measures, the researcher analyzed the data using the Pearson Product Moment Correlation. Significant and moderate correlations for the pre-test were found for AQ-Total and Nbttotal ($r = .263$), AQ-Total and Rbttotal ($r = .395$), AQ-Total and PNPR+ ($r = .400$), Nbttotal and Rbttotal ($r = .292$). Stronger correlations were found for PNPR+ and PNPR- ($r = .604$) and PNPR+ and PNVIC ($r = .642$).

For the post-test, weak to moderate correlations were found for AQ-Total and PNPR+ ($r = .317$), Nbttotal and PNPR+ ($r = .471$), Nbttotal and PNPR- ($r = .387$), Nbttotal and PNVIC ($r = .235$), and RbTotal and PNPR+ ($r = .340$). Moderate to strong correlations were found for AQ-Total and Nbttotal ($r = .557$), Nbttotal and Rbttotal ($r = .789$), PNPR+ and PNPR- ($r = .778$), PNPR+ and PNVIC ($r = .787$), and PNVIC and PNPR- ($r = .690$). Please refer to Table 11 for more specific information on pre- and

post-test correlations. Please note that scatterplots are also provided in Appendix M, pictorially representing the relationship between measures.

Table 11

Correlations between measures for pre- and post-test administrations

| | AQTotal | Nbtotal | RbTotal | PNPR+ | PNPR- | PNVIC |
|---------|-------------|--------------|-------------|-------------|-------------|-------------|
| AQTotal | ----- | .56 (45) | .24 (45) | .32 (45) | .20 (45) | .19 (45) |
| Nbtotal | .26 (63) | ----- | .79 (45) | .47 (45) | .39 (45) | .24 (43) |
| RbTotal | .40 (64) | .30 (74) | ----- | .34 (45) | .29 (43) | .20 (43) |
| PNPR+ | .40 (64) | .19 (74) | .22 (75) | ----- | .78 (43) | .79 (43) |
| PNPR- | .23 (64) | .09 (74) | .08 (75) | .61 (75) | ----- | .69 (43) |
| PNVIC | .23 (64) | -.14 (74) | .12 (75) | .64 (75) | .44 (75) | ----- |

Note. Sample size are in parentheses. Pre-test correlations are below the diagonal and post-test correlations are above.

Question 3: Is there a significant change between pre- and post-test scores for the intervention group? 3a. Assuming the change is significant, what is the magnitude of the change?

A 3x2x2 factorial analysis of variance (ANOVA) was conducted for each outcome measure to evaluate the effect of the intervention relative to gender, time, and school. Note that for this analysis SAS was used instead of SPSS because of limitations of this program to run this analysis. Please refer back to Table 6 for all pre-test and post-test means for the outcome measures.

AQTotal. Note that the means for males in School 1 and 3 increased over time, while the means for males in School 2 slightly decreased. On the other hand, females in all three schools evidenced higher means at the post-test administration. The results of the ANOVA are presented in Table 12. A statistically significant main effect for time

was revealed on this measure, $F(1, 35) = 6.12; p < .05$. This effect suggests that the overall post-test mean for all schools and both genders was significantly larger than the overall pre-test means on reported total aggression on the AQ. Additionally, an interaction effect was found for school x gender, $F(2, 35) = 3.87; p < .05$. This interaction effect suggests that the gender differences on AQTotal differed across the three schools. However, the lack of a significant school x gender x time interaction effect indicates that these gender differences were not consistent across pre- and post-test. Overall the results of this ANOVA suggest that while AQTotal scores increased over time between pre- and post-tests, the amount of change did not vary by gender and school.

Table 12

ANOVA Summary Table for AQTotal

| <i>Source</i> | <i>Df</i> | <i>Type III SS</i> | <i>Mean Square</i> | <i>F</i> | <i>p</i> |
|------------------|-----------|--------------------|--------------------|----------|----------|
| Between Subjects | | | | | |
| School (S) | 2 | 645.47 | 322.73 | 1.66 | .21 |
| Gender (G) | 1 | 16.27 | 16.27 | .08 | .77 |
| S x G | 2 | 1505.73 | 752.86 | 3.87 | .03 |
| Error | 35 | 6814.66 | 194.70 | | |
| Within subjects | | | | | |
| time (T) | 1 | 223.55 | 223.55 | 6.12 | .02 |
| T x S | 2 | 97.72 | 48.86 | 1.34 | .28 |
| T x G | 1 | 45.25 | 45.25 | 1.24 | .27 |
| T x S x G | 2 | 56.41 | 28.20 | .77 | .47 |
| Error (time) | 35 | 1278.39 | 36.53 | | |

Rbtotal. The means between pre- and post-test for both males and females in all schools were fairly stable over time. Decreases in means for males and females were evident at School 2 and 3 across time. What is noteworthy is that all the standard

deviations were more variable for the post-test outcome measure (excluding females at School 3). The results from the ANOVA may be found in Table 13. A statistically significant main effect for school was found $F(2, 40) = 9.01; p < .05$. This effect suggests that overall participants' scoring varied as a function of which school they attended. School 1 overall exhibited the highest means, followed by School 3, and finally School 2. More so, individuals at each school also differed as a function of gender, such that there was a statistically significant interaction effect for school x gender, $F(2, 40) = 3.72; p < .05$. Evidence of this significance is represented by females in School 1 reporting the highest means followed by females at School 3. No main or interaction effects were found for any time factor for this measure. Overall the results of this ANOVA suggest that there were no statistically significant changes in scores across pre- and post-test for each of the factors.

Table 13

ANOVA Summary Table for Rbtotat

| <i>Source</i> | <i>Df</i> | <i>Type III SS</i> | <i>Mean Square</i> | <i>F</i> | <i>p</i> |
|------------------|-----------|--------------------|--------------------|----------|----------|
| Between Subjects | | | | | |
| School (S) | 2 | 2527.19 | 1263.59 | 9.01 | .00 |
| Gender (G) | 1 | 212.54 | 212.54 | 1.52 | .23 |
| S x G | 2 | 1044.48 | 522.24 | 3.72 | .03 |
| Error | 40 | 5607.96 | 140.20 | | |
| Within Subjects | | | | | |
| time (T) | 1 | 200.43 | 200.43 | 2.29 | .14 |
| T x S | 2 | 3.62 | 1.81 | .02 | .98 |
| T x G | 1 | 4.04 | 4.04 | .05 | .83 |
| T x S x G | 2 | 169.60 | 84.80 | .97 | .39 |
| Error (time) | 40 | 3494.77 | 87.37 | | |

Nbtotal. Aside from all participants at School 2, all other gender means increased as a function of time. Noteworthy changes in means are evidenced for males at School 1 because these means increased by over 16 points. This increase in mean score indicates more reported aggression by participants. However, in this specific case, the standard deviation was also extremely high ($SD = 22.31$). The results from the ANOVA may be found in Table 14. The only statistically significant interaction effect for this outcome measure was for time x school x gender, $F(2, 39) = 3.32$; $p < .05$. This effect suggests that mean score differences were not the same over time across all schools and participants on *Nbtotal*. Specifically, overall scores for females increased in School 2, whereas scores for males decreased. In both School 1 and School 3, reported levels of aggression for males increased while female scores remained relatively the same. See Figures 10-12 for a pictorial representation of the ordinal interactions.

Table 14

ANOVA Summary Table for Nbtotal

| <i>Source</i> | <i>Df</i> | <i>Type III SS</i> | <i>Mean Square</i> | <i>F</i> | <i>p</i> |
|------------------|-----------|--------------------|--------------------|----------|----------|
| Between Subjects | | | | | |
| School (S) | 2 | 53.58 | 26.79 | .09 | .92 |
| Gender (G) | 1 | 158.82 | 158.82 | .52 | .48 |
| S x G | 2 | 808.02 | 404.01 | 1.31 | .28 |
| Error | 39 | 11998.12 | 307.64 | | |
| Within subjects | | | | | |
| time (T) | 1 | 129.99 | 129.99 | 1.48 | .23 |
| T x S | 2 | 336.08 | 168.04 | 1.91 | .16 |
| T x G | 1 | 199.03 | 199.03 | 2.26 | .14 |
| T x S x G | 2 | 584.50 | 292.25 | 3.32 | .05 |
| Error (time) | 35 | 3433.46 | 88.03 | | |

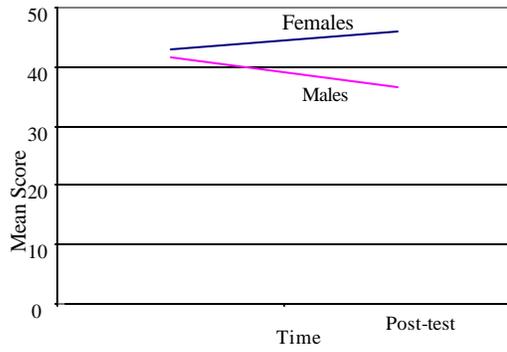


Figure 10. Ordinal interaction for School 2 on Nbttotal.

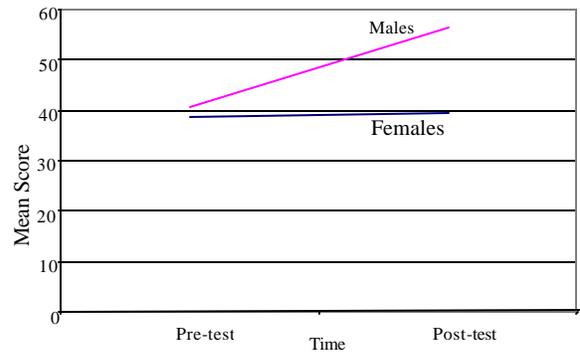


Figure 11. Ordinal interaction for School 1 on Nbttotal.

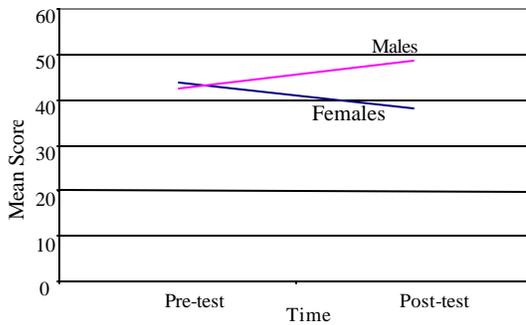


Figure 12. Ordinal interaction for School 3 on Nbttotal.

PNPR+. The mean scores over time for females at School 1 and males at School 3 rose dramatically (by 50 and 73, respectively). Additionally, increases in means over time were found for males at School 1 and 2. Overall, it appears that females exhibited declines in their means. Note that *PNPR+* represents a positive construct, by which participants rated certain students for the positive and/or helping interactions they had with others. Therefore, these results suggest that many of the participants perceived more support and kindness from students at the post-test administration. The results of the ANOVA are presented in Table 15. An interaction effect was found for school x gender, $F(2, 40) = 5.78; p < .05$. This suggests that the gender differences on *PNPR+* differed across the three schools. For example, females in School 2 initially exhibited higher

scores (more positive reports of prosocial behavior) at pre- and post-test compared to females in Schools 1 and 3. Additionally, females in Schools 1 and 3 actually reported lower scores at post-test suggesting less prosocial behavior in the population they were scoring. On the other hand, males at all three schools exhibited higher mean scores at post-test.

Table 15

ANOVA Summary Table for PNPR+

| <i>Source</i> | <i>Df</i> | <i>Type III SS</i> | <i>Mean Square</i> | <i>F</i> | <i>p</i> |
|------------------|-----------|--------------------|--------------------|----------|----------|
| Between Subjects | | | | | |
| School(S) | 2 | 13082.70 | 6541.35 | 1.66 | .20 |
| Gender (G) | 1 | 2366.85 | 2366.85 | .60 | .44 |
| S x G | 2 | 45506.33 | 22753.16 | 5.78 | .01 |
| Error | 40 | 157565.42 | 3939.14 | | |
| Within subjects | | | | | |
| time (T) | 1 | 10297.29 | 10297.28 | 5.46 | .02 |
| T x S | 2 | 4826.54 | 2413.27 | 1.28 | .29 |
| T x G | 1 | 3539.95 | 3539.95 | 1.88 | .18 |
| T x S x G | 2 | 11369.92 | 5684.96 | 3.02 | .06 |
| Error (time) | 40 | 75390.84 | 1884.77 | | |

PNPR-. While some mean scores decreased slightly over time (i.e., females at School 3 and females at School 1) the vast majority of scores increased anywhere between 12 and 30 points. While PNPR+ is a positive construct, PNPR- is not; therefore, evaluation of these mean scores suggests that overall negative perpetration across gender and school increased over time. The results of the ANOVA are presented in Table 16. A statistically significant main effect for school, $F(2, 40) = 4.32, p < .05$, and for school x gender, $F(2, 40) = 4.81; p < .05$. This suggests that on a more macro-level, the schools differed in their means scores; however the interaction effect also highlights that females

and males within each of these schools also exhibited differences. Females in Schools 1 and 3 reported a decline in negative perpetration across time; whereas, females in School 3 reported an increase in negative perpetration. In contrast, males in all three schools reported more perpetration across time. No other effects were rendered significant. Overall the results of this ANOVA suggests that while PNPR- scores differed across gender within each school whereby the majority of females scored higher on this subscale, the time component was not a contributing factor. Thus, it may be inferred that the intervention had no effect on PNPR- across pre- and post-test measures.

Table 16

ANOVA Summary Table for PNPR-

| <i>Source</i> | <i>Df</i> | <i>Type III SS</i> | <i>Mean Square</i> | <i>F</i> | <i>p</i> |
|------------------|-----------|--------------------|--------------------|----------|----------|
| Between Subjects | | | | | |
| School (S) | 2 | 11739.39 | 5869.70 | 4.32 | .02 |
| Gender (G) | 1 | 594.25 | 594.25 | .44 | .51 |
| S x G | 2 | 13069.15 | 6534.58 | 4.81 | .01 |
| Error | 40 | 54319.44 | 1357.99 | | |
| Within subjects | | | | | |
| time (T) | 1 | 2830.53 | 2830.53 | 3.71 | .06 |
| T x S | 2 | 672.80 | 336.40 | .44 | .65 |
| T x G | 1 | 1288.37 | 1288.37 | 1.69 | .20 |
| T x S x G | 2 | 1305.98 | 652.99 | .86 | .43 |
| Error (time) | 40 | 30515.20 | 762.88 | | |

PNVIC. Unlike most of the other outcome measures described, the mean score differences on the PNVIC were much more variable. Males in all three schools exhibited an increase in mean scores over time. On the contrary, females in all three schools exhibited declines in mean scores. It is interesting to note that the findings on PNVIC

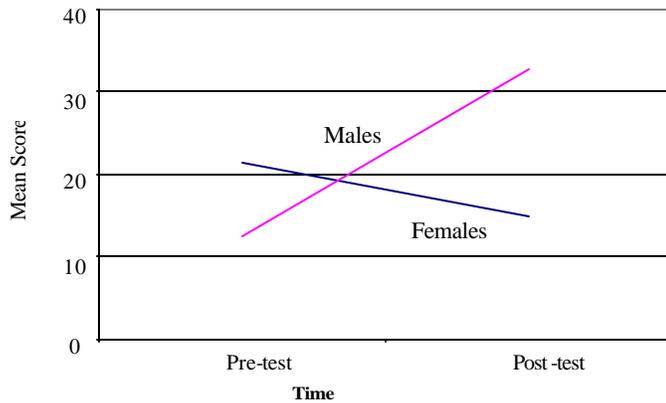
seem to be contradictory to the findings from PNPR-, such that if a majority of scores increased for aggressive perpetration so would scores on the victimization subscale. However, because the students in PET also interact with students outside the program, it may be that the victims were not necessarily part of the participant group, but that the aggression was extended to other students in each respective school. The results of the ANOVA may be found in Table 17. The only interaction effect that was found was for time x gender, $F(1, 40) = 4.79; p < .05$. This suggests that between the pre- and posttest administrations, scores differed as a function of gender. Males exhibited a sharp increase in reported victimization over time. In contrast, females exhibited a decline in the number of persons deemed as victims. Note Figure 13 for a pictorial representation of the ordinal interaction.

Table 17

ANOVA Summary Table for PNVIC

| <i>Source</i> | <i>Df</i> | <i>Type III SS</i> | <i>Mean Square</i> | <i>F</i> | <i>p</i> |
|------------------|-----------|--------------------|--------------------|----------|----------|
| Between Subjects | | | | | |
| School (S) | 2 | 1063.55 | 531.77 | .61 | .55 |
| Gender (G) | 1 | 569.75 | 569.75 | .65 | .42 |
| S x G | 2 | 5243.45 | 2621.73 | 3.00 | .06 |
| Error | 40 | 34956.69 | 873.92 | | |
| Within subjects | | | | | |
| time (T) | 1 | 725.67 | 725.67 | 1.24 | .27 |
| T x S | 2 | 1114.83 | 557.41 | .95 | .40 |
| T x G | 1 | 2807.04 | 2807.04 | 4.79 | .03 |
| T x S x G | 2 | 381.02 | 190.51 | .32 | .72 |
| Error (time) | 40 | 23448.18 | 586.20 | | |

Figure 13. Ordinal interaction on PNVIC by gender and time.



The magnitudes of the interactions found were computed by determining the effect sizes between mean scores on the pre- and post-test measures. A large effect size on Nbttotal was found for the interaction between time x gender x school, specifically for males at School 1 (.99). Moderate effect sizes were found for males at School 2 (.64) and 3 (.52) and females at School 3 (.44). Small effect sizes were found for females at School 1 (.05) and 2 (.15). In order to validate that the effect sizes were meaningful and based upon truly statistically differences, posthoc testing was conducted.

The Dunn's test was used as a posthoc testing approach to control the familywise error rate, which is the probability that the series of comparisons contained at least one Type I error. Based on comparing the mean scores between males and females on the PNVIC post-test, it was confirmed that the null hypothesis be rejected because the obtained t-value was greater than the critical t ($2.80 > 2.62$), therefore suggesting that the groups are not the same. In order to determine if the null hypothesis could be rejected for the Nbttotal measure, the researcher used school-specific mean gender differences based on pre- and post-test scores. For example, the researcher compared the mean of males and females in School 1 on the pre-test and post-test separately. After these were determined they were factored into the numerator of Dunn's equation, subtracting the

post-test mean gender difference from the pre-test mean gender difference. The researcher was able to reject the null hypothesis that there were differences in School 1 ($3.27 > 2.78$); however the researcher was unable to reject the null hypothesis in Schools 2 ($1.77 < 2.78$) and 3 ($1.73 < 2.78$) by time, gender, and school. Therefore, these results suggest that the gender difference in school 1 increased across time but stayed the same in schools 2 and 3.

Question 4: What is the relationship between how each facilitator viewed the effectiveness of the program and the results from the outcome measures? 4a. How did students view the program relative to how the facilitator viewed the utility of the program? 4b. Is the outcome data consistent with facilitator and student reports?

The consumer satisfaction survey provided positive information on how the students viewed the programming. Overall, both males and females who participated in the study scored *Owning Up* higher in terms of the content being “interesting” ($M = 4.09$; $M = 3.88$, respectively), “learning new things” ($M = 4.75$; $M = 4.55$, respectively), and recommending it to other students ($M = 4.38$; $M = 4.18$).² Histograms for each category listed previously are shown below in Figure 14. Additionally, frequency charts are noted next to each histogram, beginning with Table 18. A score of three on “Interest in the program,” five for “New Learnings,” and 20 for “Recommending to others” were the scores with the highest frequency in each category. While these scores are high and indicate feelings of positive satisfaction towards the curriculum, verbatim information provided by the students does not necessarily indicate such feelings. For example, one student rated the “New Learnings” a “4,” yet indicated “I didn’t learn much. I knew all

² Note that the majority of students either did not write the correct research number on their paper, or failed to write a research number at all. Therefore, results are combined for both schools.

this anyways.” However, for the majority of cases qualitative information provided by the students was comparable with how they scored. Some specific remarks regarding students’ interest in the program included, ““Most of the time I was interested but you guys need to get someone that can relate to what is going on in our teenaged lives like (slang) etc., (4)” “cause it wasn’t very long and than we only talked about the same thing through the program (3),” “In this program I was somewhat interested but most of the things we talked about didn’t really appeal to me (2),” “I love it (5).” Remarks about new things the student’s learned included, “You always learn something new no matter how good you know bout it (5),” “It taught me how to hold back (4),” “I learned that some people feeling get hurt easily (3),” “Somewhat, I think I learned a little most of it I already knew. Nothing that was said was really new to me (2),” and “We didn’t really learn about violence and aggression (1).” Remarks regarding recommending the program to other students included, “Because most students these days are violent and very aggressive (5),” “Yes if they need it (4),” “I don’t now cause most kids want to have fun and this class is not that fun (3),” “Yes but then again no cause it boring (2),” and “Most kids my age don’t care (1).” When prompted to tell us what students would keep the same about the program the following is a representation of the students’ opinions: “The people cause there are some nice people in the program.”; “The way they teach it needs a little more excitement.”; “I would keep Mrs. D. the same because I like the way she teaches the program and she also can relate to what we be saying.”; “I would say that it would have to be the amount of people cause if there is to many then you would not get to hear everybody.”; “It would be the way that they tell things.” On the converse side, students’ also were prompted about ways that the program could be changed. Some of

their responses included: “That when you talk about something for her not to bring you down.”; “More time in the group.”; “I think they should put the boys and girls in the same room at the same time.”; “Nothing really.”; “I think we should have more activities, such as games to play while learning to make it more interesting.” The lessons that students noted as most valuable included “the bullying lesson,” “when we gave our feelings about girls,” “when you don’t control yourself the environment does,” “nothing was valuable,” “Own up to things you did wrong, own up to some things not all things,” “Relationships,” “I found that everything was really valuable.” Please note that it was intended for the students to actually name a lesson that they found to be most valuable; however, based on their responses it appears that the students referred to anything in the program that they remembered or had an impact on them.

Table 18

Frequency distribution, consumer survey, Interest in program item

| | Frequency | Percent | Valid Percent | Cumulative Percent |
|--------|-----------|---------|---------------|--------------------|
| Valid | 2 | 4 | 7.3 | 10.0 |
| | 3 | 15 | 27.3 | 47.5 |
| | 4 | 10 | 18.2 | 72.5 |
| | 5 | 11 | 20.0 | 100.0 |
| Total | 40 | 72.7 | 100.0 | |
| Missin | 15 | 27.3 | | |
| Total | 55 | 100.0 | | |

Figure 14. Histogram of consumer satisfaction survey-Interest in program item.

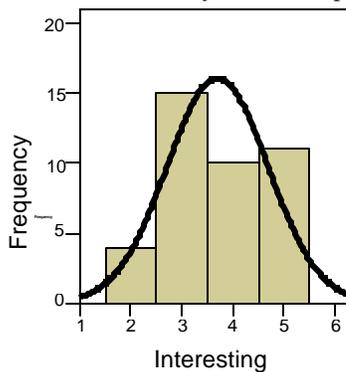


Figure 15. Histogram of consumer satisfaction survey-New Learning item.

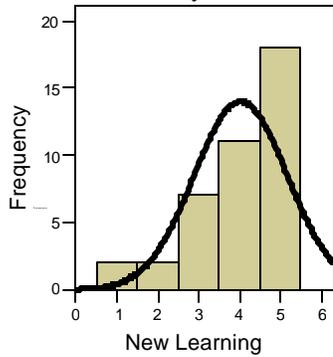


Table 19
Frequency distribution, consumer survey, New Learning item

| | Frequency | Percent | Valid Percent | Cumulative Percent |
|---------|-----------|---------|---------------|--------------------|
| Valid | 1 | 2 | 3.6 | 5.0 |
| | 2 | 2 | 3.6 | 10.0 |
| | 3 | 7 | 12.7 | 27.5 |
| | 4 | 11 | 20.0 | 55.0 |
| | 5 | 18 | 32.7 | 100.0 |
| | Total | 40 | 72.7 | |
| Missing | 15 | 27.3 | | |
| Total | 55 | 100.0 | | |

Figure 16. Histogram of consumer survey--Recommendation of program item.

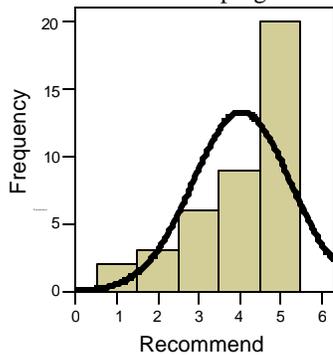


Table 20
Frequency distribution, consumer survey, recommendation of program item

| | Frequency | Percent | Valid Percent | Cumulative Percent |
|---------|-----------|---------|---------------|--------------------|
| Valid | 1 | 2 | 3.6 | 5.0 |
| | 2 | 3 | 5.5 | 12.5 |
| | 3 | 6 | 10.9 | 27.5 |
| | 4 | 9 | 16.4 | 50.0 |
| | 5 | 20 | 36.4 | 100.0 |
| | Total | 40 | 72.7 | |
| Missing | 15 | 27.3 | | |
| Total | 55 | 100.0 | | |

Facilitator commentary about the program differed quite drastically from the students. This may be indicative of each facilitator's ability to express herself better than the students and/or familiarity with the curricula. Regardless of the rationale, each facilitator provided valuable information to propel the curricula forward, if warranted, in future implementations. Changes for the future of the program differed by facilitator. Whereas two facilitators felt the curricula was appropriate for all children and served as a

great tool to teach people about violence and aggression, another facilitator felt that it merely changed the labels and categories that students use to talk about each other. For example, one facilitator who is a proponent of the program noted her feelings about the appropriateness of the program for diverse groups of children:

“Yes. I had a really really diverse group, not just color wide. Some low to middle income, all SES, and kids coming from different life experiences, which is always good. No one could say mine [my situation] is more horrible than yours. Some kids were from foster settings and other from two parent-family groups. [*Owning Up* is] good for the kids to see that they are all going through the same stuff globally and it is not just specific to their demographic.”

The facilitator that felt that the program was partially adequate noted:

“It provided the opportunity for the students to recognize and acknowledge negative or aggressive behavior patterns; opportunity to do self-examination; not sure that *Owning Up* provided more accomplishments than any other violence/aggression programs. I am not crediting the program but the opportunity that it provides.”

Therefore, there were mixed reports regarding the actual utility of this specific program. The commitment of the SIS was also something that was indicated by some of the facilitators regarding what they would keep the same (“I felt free to contact her with concerns [SIS]. Administration involvement--thought it was excellent.”) It appeared that commitment by each school either helped the program run effectively or not.

All facilitators expressed that they would have liked to have more time to teach the students the curricula, either in the length of weeks or in the class time. As noted by

one facilitator, “50 minutes were more like 40 minutes because of stragglers coming to class and then packing up their stuff before class ended.” Another facilitator also commented about time constraints: “Time would work if the schools would allow someone to do the program for a longer session (week-wise). Class length is fine though.”

Additionally one facilitator noted that she would like to include male facilitators in the future because they might be able to relate to the male students better. Other points that facilitators touched upon relative to changing the curricula/implementation process included teaching the curricula closer to the beginning of the year instead of the last nine weeks of school and to add more diverse representations of media clips (“Add something more than tough guys on the media side of things. The kids liked knowing about their culture and the music piece of it, but I think the same thing should happen with other forms of media like magazines, movies, T.V. programs--that could be fleshed out”) and examples that pertain to a wider demographic group.

Strengths of the curricula as indicated by the facilitators included discussion because “it breaks down barriers”, role-plays because “students like to act out things”, and journaling because some students are unable to get up in front of other students; yet would still like to express themselves. Weaknesses included too much material packed into every class and that *Owning Up* is a difficult program to teach if the facilitator has not had a lot of experience leading groups. One facilitator cited a serious weakness in the program: “the curriculum teaches the students to categorize, label, and address each other based on the label they have been assigned” which is actually what she wanted the students not to do. Additionally, “the PDRA which intends to teach you to stop, explain,

ask, and lock in actually reinforced students to stop, attack, and walk away.” Clearly the viewpoints on what the program actually advocates to students are quite different, as reported by each facilitator.

It would be expected based on the varying degrees of satisfaction with the program by both students and facilitators that the results between the pre- and post-tests may also have exhibited some variability. Based on the results from the ANOVAs, it appears that there are a limited number of truly significant interactions in terms of change over time (i.e., between pre- and post-test). It is difficult however to delineate the reason why there was not a significant change. It is precarious however, that overall reports from students were relatively high yet change was not exhibited in behavioral or cognitive terms in presentation on the quantitative measures. Chapter 5 will examine these issues further and address potential explanations for the results.

Chapter 5

Discussion

Summary of Study

The present study was conducted to explore the effects of a gender-based violence prevention program on a group of middle school children. This study was novel in nature because it was one of the first studies to empirically test a curriculum that deals with both physical and relational aggression. A mixed methods design was used in order to gain insight on both significant changes over time among the participants and treatment integrity of the intervention. Findings from this study suggest that treatment integrity differed across facilitators and across sites. Participants also exhibited minimal gains between pre- and post-test administrations on both behavioral and cognitive measures of various forms of aggression. This chapter will summarize the results from Chapter 4, discuss implications of the results, examine limitations, and finally suggest directions for future research.

Examination of Results

Prior to examining the highlights from Chapter 4, it is important for the reader to put the findings from this study in context. Based on the fact that the curricula were not implemented according to plan, it is very difficult to determine if the relative lack of change from pre- to post-test means that the program was not efficacious. The program may in fact be efficacious for certain populations if implemented with integrity. As suggested previously, the lack of implementation integrity may have been due to the novel nature of OP-TB working in the public school sector, the experience of one of the

facilitators, and the degree to which communication was clear about the researchers intent to keep the implementation streamlined.

Notable findings from the aggression measures. The results from the measures administered in this study produced some interesting findings. Females overall tended to be more aggressive than males on physical, verbal, and relational forms of aggression. While these results are consistent with McEvoy et al. (2003), overall this finding contradicts many other studies that suggest that males tend to be more physically aggressive than females (e.g., Crane-Ross, Tisak, & Tisak, 1998; Osterman et al., 1998.) Additionally, the level of relational aggression found in this group of students overall was significantly lower than their reported levels of physical aggression. This may be a function of the demographics of the population sampled. It has been suggested that physical aggression may be more functional for African-American children from low-income communities as compared to middle-to-high SES youth (Rodkin, Farmer, Pearl, & Van Acker, 2000).

Additionally, results showed a high degree of variability in scores. There seemed to be very little stability between the scoring of items for both males and females on physical and relational aggression items. This again is inconsistent with other researchers that suggest that overall females tend to have higher rates of relational aggressive tendencies and tend to report other forms of aggression to a lesser degree. (e.g., Crick et al, 1997.)

It also is noteworthy to mention that the results on the Peer Nomination Scales for Children may have been compromised by the lack of partition use. Students may not

have felt that their confidentiality was being maintained and therefore they may have not responded as honestly.

Comparing female means for the pre-test suggest that the treatment group and the control group exhibited similar scores. The highest scores for females were found on their normative beliefs about perpetration and victimization and various aggressive behaviors including verbal, indirect, physical, and social. Overall, males in School 3 (i.e., the control school), exhibited the highest level of aggression, particularly for positive perpetration (i.e., helping others) and use of various forms of aggression including physical, verbal, etc. Interestingly, the scores for males on the relational aggression measure were higher than for females, again providing evidence that this sample does not fit the standard paradigm of females exhibiting more relational aggression.

The pre-test examination of reliabilities for each item exhibited important results. All items for each measure used in this study exhibited high inter-item correlations; however, two items on the MacDonald et al. (2000) measure did not. These two items related to fighting with gang members and bringing weapons like guns and knives onto school grounds. As a result they were excluded during the analysis of the data. It may be hypothesized that either a) the population sampled was still too young to have gang affiliation or b) the level of aggression and violence was not substantial enough in this population to elicit use of guns and knives. Note that because of the low incidence of these kinds of behaviors, a low reliability is also yielded. This is one disadvantage of utilizing a survey to examine low incidence aggressive behaviors because important data about individuals is lost.

Additionally, a range of correlations from weak to moderate were found across measures. The strongest correlations were found between the various categories on the Peer Nomination Scales for Children. This is important because the researcher determined that the kinds of categories that were dichotomized in this study, had never been examined in the past in terms of reliability. Fairly low correlations were found for total scores on NOBAGS and MacDonald et al. (2000). This seems problematic since some of the questions on the NOBAGS pertain to relational and social aggression constructs. Further research may attend to this and address why the correlations tended to be low. It may be hypothesized though, that students were unable to see the similarities of the various forms of aggression because they had not received the intervention. This relates to the fact that the post-test correlations (see Results section) exhibited a strong relationship.

Implementation integrity. While the design of the study intended for the implementation to be consistent and stable across treatment, results suggest that the implementation of the *Owning Up* program suffered from many deviations from the original plan. Therefore, results of this study have to be considered in light of the fact that the intervention was not conducted under the conditions that it was designed to. Specifically, the number of lessons offered at each site differed because of communication and scheduling difficulties. Students who were supposed to be in attendance were not compliant and there were no repercussions for not attending the intervention. Furthermore, facilitators had different views on what the curricula meant and how it was supposed to be taught. Thus, each participant group did not get the same type of training.

Regardless of the fact that there were many problems with implementation, as well as limited significance in the results, the consumer satisfaction survey did elicit some positive feedback about the program. What is most interesting about the results is that the majority of the students (45%) ranked the question regarding learning new things a “5.” Clearly, if the students believed that they had learned from this program, it would be expected that their learning would generalize to the various measures presented at post-test. However, the relationship between pre- and post-test scores does not exhibit this kind of trend. It would be beneficial to try to ascertain what the students learned and if the assessments administered actually measured this.³ Additionally, it may be hypothesized that the students did not have enough time to internalize and integrate their new learning into their behavior prior to taking the post-test and thus timing of the post-test could have been a confound in this study. Finally, the majority of students also noted that they would highly recommend this program to other students. This again is noteworthy because it indicates that the students did enjoy the program and found it valuable, regardless of the fact that no change was evident on the post-test measures.

In light of the results of both the summative and formative measures, it is important for the reader to take these results with a grain of salt. It is unclear whether the demographics of the participants, the lack of implementation integrity and consistency across settings, or the familiarity with relational forms of aggression confounded the findings.

Effectiveness of the intervention. For the majority of the scores for both males and females in all three schools, post-test results indicated higher levels of self- and peer-

³ Note that students were asked to specifically note what they learned, but overall minimal written feedback was provided.

reported aggression. Specifically, males in School 1 espoused more positive perceptions of aggressive behavior and males in all three schools exhibited a large increase in the number of victims reported. The reported high scores for males is consistent with one facilitator's viewpoint that the program reinforces poor behavior. Interestingly, female gains in mean difference scores were not as extreme as males. Females tended to show just slight increases across all three schools on the outcome measures. It may be suggested that the role of a male facilitator would change these outcomes. To date, the responsiveness to intervention, based on the gender of the facilitator has not been examined and may have important implications.

An interesting finding on the Peer Nomination Scale is worthy of discussion. Among the complete data subset group, the mean for positive perpetration increased by over 30 points, the mean for negative perpetration increased by 12 points, and the mean for victim increased by 5 points. Although there is no way to determine why this occurred, it may be hypothesized that students who were involved in the program became more aware of persons in their environment who typified the role of victim, perpetrator, etc. However, the standard deviations were so large that one may have to question the validity of the increase in means and how truly representative they were of the entire sample. Again, this relates to the aforementioned point that the ratings between individuals were extremely variable.

Interestingly, unlike the pre-test, the post-test measures actually exhibited stronger relationships overall. Again, the strongest relationships were between the Peer Nomination categories; however, the total score on NOBAGS and the total score on MacDonald et al. (2000) had a very strong relationship as well. The fact that the

measures had higher correlations at the post-test may be indicative of the effect that the curricula had on the students. Specifically, the students may have been able to comprehend the questions better because the curricula provided opportunities to discuss various forms of aggression and violence that the students may have not been aware of prior to the program.

While main effects were found in this study (including gender and gender x school), the main focus of this study was concerned with addressing significant interactions over time. This is because the main point of the curricula was to impact students in a positive way, thus decreasing their overall levels of aggression after the program. According to the analyses, the total score on NOBAGS exhibited a very powerful interaction by time x school x gender suggesting that mean score differences were not the same over time across all schools and participants. However, in School 1, males' and females' mean scores increased, in School 2, males' scores increased and females' scores decreased in School 3 more analysis was needed to make a conclusion regarding this significance. Post hoc testing found that gender differences in School 1 did actually change; however, no differences were truly found in School 2 and 3. Based on this information, it appears that the intervention may have not actually had the results that it intended. The results at School 1 and at the control school were very similar, which is indicative of the fact that the intervention may have not been the primary factor in changing the mean scores of the participants. Additionally, the fact that male means increased so significantly in School 1 actually had the opposite effect of what was intended.

Finally, the other interaction that was significant was on the PNVIC for time x gender. Males again increased their scores significantly while females declined. One would hope that the level of victimization would drop as a function of this curriculum; however, according to the results, this is not the case. However, alternative explanations do exist. First, males may have become more aware of victim and perpetrator roles, thus accounting for the increased levels of victimization. For example, the curricula may have been able to make roles and categories that certain students fall under more concrete. Therefore, males may have been able to categorize certain students based on new definitions and understanding that they acquired about aggression. Second, based on the cyclical nature of victims and perpetrators (Rodkin & Hodges, 2003), the time of testing may have assessed a period whereby role reversal was occurring. For example, more students may have attained a victim role during the post-test than during the pre-test.

Implications of the Results

Oftentimes in both profit and non-profit settings like schools, issues with program intervention and evaluation stem from systemic troubles rather than from the material presented in an intervention. While the researcher is not advocating in this case, that the curricula does not have faults and may in fact be attributed to the minimal change among students, the researcher is merely trying to shed light on the fact that many of the implementation issues may have evolved from a larger, systems level problem.

In the case of this evaluation, there were three separate systems functioning to implement and evaluate Owing Up: the OP-TB, the researcher, and the schools. In a perfect world, these three systems would work collaboratively in order to problem-solve the best way to implement this program from start to finish. However, the fact that these

systems each had their own agenda, culture, and way of functioning limited the degree to which they could work conjointly. Therefore, a discussion of key organizational factors that were not present in this study will help the reader to understand the complexity of the interactions between systems.

Supportive and shared leadership. The first issue that relates to each of the systems is the degree to which supportive and shared leadership was established. It is suggested that leaders during a change process participate without dominating, share authority, and are able to facilitate a positive working environment (Hall & Hord, 2001). In this situation, a member at OP-TB restricted many aspects of the intervention. First, the researcher was unable to attend to logistical issues and was only allowed limited levels of communication with the school system. This is an inherent problem according to the organizational literature because the researcher was unable to truly understand the climate and the culture within the schools. Additionally, the researcher was unable to build rapport with the administrators on site. The high attrition rates for the post-test may have been circumvented if the researcher was able to build relationships with each school and maintain them over time.

Roles and responsibilities. The second major issue is the degree to which the three systems understood their own and each member's roles and responsibilities (Fleming & Monda-Amaya, 2001). Throughout the process, persons in each system were vying for power and control as a function of not being clear on who was in charge of certain responsibilities. For example, a member at OP-TB thought that she was in charge of making certain decisions regarding the research component of the program, while myself and another member thought that each one individually was supposed to decide

upon research issues. While initial meetings among the three systems attempted to define roles and responsibilities, they were continually shifting throughout the process, leading to confusion and miscommunication.

Commitment of key personnel. The gatekeeper, also known as the person that has authority and decision-making power, is critical for an intervention to be sanctioned by members of an organization. While the SISs at each school noted that they were excited and committed to the intervention programming, the behavior of one of the SISs did not reflect this. This SIS behaved in ways that indicated that her level of buy-in was less than would be ideal as evidenced by her trying to lead the group herself, not letting OP-TB know that she was canceling the intervention on certain days, not holding her students accountable for attending class, and providing inconsistent information to various OP-TB staff. Clearly, if the cooperation is not evident on behalf of one of the systems involved, it makes it very difficult for a curriculum to be taught in the way it was intended.

The second component related to key personnel is involvement of all stakeholders (Curtis & Stollar, 2002). Although all adults were consulted about the program process, student stakeholders were not. For example, students were never explicitly asked what they wanted to gain out of the program nor were their opinions solicited prior to the intervention. Although it may have been difficult to include students in the planning and implementation process, buy-in may have been higher if the students felt like they had provided their opinions and thoughts about the new program. Additionally, a feeling of belonging to the process and investment may have afforded better attendance rates throughout the intervention.

In order for these systems issues to be dealt with proactively in further implementation efforts, it may be suggested that future teams be trained in the problem-solving process: a systematic way to identify and analyze problems as they arise, as well as developing solutions to the problems (Tilly, 2002). Furthermore, prior to implementing an intervention, each system may attempt to work on specific problem-solving in the areas of team structure and process, specifically, addressing, explicit goal generation, roles and expectations, group communication and listening skills (Iverson, 2002).

The next major implication of the study relates to the “goodness-to-fit” between the program and the cultural background of students with whom it was being used. In order for an intervention to be effective, it must be sensitive to the needs of the audience it is intended for (Kratochwill & Stoiber, 2002). Traditionally, intervention research has been dominated by the assumption that interventions have a universal effect on all participants. However, current research suggests that cultural issues in intervention programs should not only be reviewed but should be a key aspect in intervention development (Roosa & Gonzales, 2000). It appears based on the qualitative data component analyzed in this study that the fit between participant and intervention was not congruent. There are two facets of the intervention that may have not been appropriate for the students in this study. The first aspect has to do with language. Based on consumer satisfaction surveys, outside interviews, and facilitator checklists, it is evident that students struggled with much of the wording of the curriculum. For example, students often said things like, “I don’t understand your language.” “You don’t make no sense.” “These words are too big.” This is problematic for two reasons. First, it isolates the

students from both the facilitator and the curricula, making it uncomfortable for students to learn.

Secondly, students may not have comprehended what the curricula was actually trying to achieve. Thus, the “non-findings” in this study may inherently be a result of the lack of cultural relevance to this population (Mattis, 2002). Therefore, future implementation of this program may try to adopt the cultural specificity hypothesis (culture is a specific contextual variable that plays a role in treatment effectiveness) during the planning phases (Wampold, 2002). Specific ways that this theory may be applied is by making taking care to devise linguistically appropriate meanings of typically Euro-centric words.

As it was suggested previously, the lessons taught to the students may not really address the underlying aggressive tendencies of this ethnic population. While the research indicates that all students are targets of relational aggression, it is still unclear, based on the limited data, if African-American students perceive RA in the same way as Caucasian students. Therefore, prior to implementing this program again, it may be suggested that a focus group be conducted in order to ascertain the level by which most African-American students experience aggression in their school and community. The focus group could assess life circumstances of the students, their socialization patterns at home, and their perceptions of aggression. In this way, facilitators would be able to fully comprehend what their students need and then may adapt the curriculum as needed. If this intervention does not meet the needs of the students, it either must be adapted, or not be used with this specific population (Yu & Seligman, 2002).

Limitations

The first limitation of this study is the method by which participants and setting were chosen. Random selection and assignment were not possible in the current study because expressed interest in the program from administration was the key determinant in selection. Therefore, the study cannot claim a true experimental design. Additionally, because the participants are an at-risk population, they may be predisposed to more violent and aggressive behaviors and affect. This may confound the level of effectiveness that the curricula may have provided for this population. Future research may want to test the curricula in children who are not affected by daily violence and aggression. Also, because of the demographics of this population being at risk and predominately African-American, the results may not generalize to other groups of students.

The second limitation was that this was the initial piloting of the program in the Hillsborough County School District. Therefore, it was hypothesized that many issues would arise with the implementation and this hypothesis was validated in the current study. Thus, findings from this study must be validated again in future studies.

The third limitation was the degree to which attendance remained stable in School 1. Based on qualitative information provided by the facilitator, class size diminished over time and ranged from three to five students. Additionally, student attendance was not consistent and the facilitator may have only seen the student once or twice throughout the entire implementation.

The fourth limitation is the degree to which the researcher was able to manage the process of implementation from start to finish. Although logistically it makes it simpler

for school personnel to notify one contact person to communicate with during the implementation, challenges arose as a result of designating one key contact person. Miscommunication between the schools, OP-TB, and the researcher may have compromised attendance rates at pre- and post-test administrations. Additionally, rapport was not built between the researcher and the schools, thereby potentially limited the buy-in from the schools to participate in the research aspect of the project.

The fifth limitation may have been test burden placed on the students. Although testing only occurred for approximately 30 minutes, it is unclear if this student population had prevalence of any disorders that may have attributed to attentional difficulties during testing.

The sixth and final limitation indicates that the students may have not received the same kind of training in School 1 and 2 based on facilitator motivation, style, and ethnic similarity of the facilitator to the participant group. Therefore, generalizability of the findings must be couched in terms of the different ways the intervention was handled between School 1 and 2. More importantly, at this point, based on all of the confounding variables of the intervention in the classrooms it would be difficult to decipher which if any factor contributed to a more beneficial experience for the students.

Suggestions for Future Research

The most basic suggestion for future research is for the curricula to be implemented in their entirety. It is unclear if the change from pre- to post-test would have been more significant if the students received the entire intervention from start to finish. Along the same lines, the full curricula should be expanded to other ethnic populations which may contain both moderately at-risk and general education students.

By expanding the research efforts across several domains, questions regarding the “goodness to fit” model may highlight for which ethnic group or kind of student this curricula is most appropriate.

While the measures utilized in this evaluation covered the gamut of types of aggression, it may be that the *Owning Up* program was targeting different underlying constructs that were not sensitive to the tools used in the present study. Therefore, future research may attempt to use other instruments related more to gender-based aggression.

This study clearly took place over a fairly limited time span and thus may have not been sensitive to long-term changes. Future research may attempt to do follow-up longitudinal studies on participants to understand the level to which students’ behavior and attitudes remain positive and stable over time following program participation.

Addressing systems level issues also remains a predominant issue. Currently, the curricula is heavily focused on the student-facilitator relationship. This makes it very difficult for teachers, parents, and administrators to adapt their own behavior and understanding of students if they are not familiar with what is being taught during a specialized program. Therefore, it is suggested that the curricula be adapted to integrate various persons that may actually help to make generalize the learning from the *Owning Up* program more prevalent to various real-life situations.

A prime example of how the curricula may be adapted is the Stop and Think Social Skills Program (Knoff, 2001). This program involves all aspects of the school ecology to promote the messages of social skills training. For example, signs are placed around schools to serve as primers for children to remember what they have been taught, all faculty from the front-desk secretaries to the principals are well-versed in the program,

and parents are notified of what is occurring in the school environment. Consequently, students are saturated with direct teaching, modeling, and consistent programming in all aspects of their daily school routine. This is a strong benefit of adapting a program like *Owning Up* into a more systems-based approach.

Finally, a critical component in working within a system is being able to effectively problem-solve various situations and/or issues. Problem-solving involves a collaborative effort of a team to work through systematic steps in resolving a problem. Using a problem-solving approach ensures that possible interventions that are derived are based on evidence and collective opinion. Therefore, future research may attend to training public school intervention teams on how to use the problem-solving process and correlate problem-solving with implementation integrity.

Final Thoughts

The evaluation of the *Owning Up* curricula provided the researcher with a novel opportunity to understand not only the effect that the intervention had on students but the ramifications of variables that impacted the overall evaluation including culture, process, and systems. While the results did not suggest any major impact on the students' behaviors or attitudes, the findings still remain significant for multiple reasons. First, undertaking an evaluation like this promotes the use of evidence-based interventions (EBI's) in schools by acknowledging that just because an intervention looks appealing at a surface level, it must be empirically validated. Secondly, this study highlighted the importance of acknowledging multiculturalism and diversity in our schools. Prior to the data collection, culture was not inherently a focus of this study; however, it became evident that it was an important variable that must be considered in future

implementation efforts. Finally, results of this study strongly suggest that it is important for systems to work together, using documented problem-solving strategies, for an intervention such as this to be successful in the public school system.

References

- Archer, J., & Haigh, A. (1994). Social representations of aggression in children. *Aggressive Behavior, 20*, 101-114.
- August, G. J., Hektner, J. M., Egan, E. A., Realmuto, G. M., & Bloomquist, M. L. (in press). The Early Risers longitudinal prevention trial: Examination of three year outcomes in aggressive children with intent-to-treat and as-intended analyses. *Psychology of Addictive Behaviors*.
- August, G. J., Lee, S. S., Bloomquist, M. L., Realmuto, G. M., & Hektner, J. M. (2003). Dissemination of an evidence-based prevention innovation for aggressive children living in culturally diverse, urban neighborhoods: the Early Risers effectiveness study. *Prevention Science, 4*(4), 271-286.
- Bandura, A., Ross, D., & Ross, S. A. (1961). Transmission of aggression through imitation of aggressive models. *Journal of Abnormal and Social Psychology, 63*, 575-582.
- Batsche, G. M., & Knoff, H. M. (1994). Bullies and their victims: Understanding a pervasive problem in the schools. *School Psychology Review, 23*(165-174).
- Berk, L. E. (2003). Self and social understanding. In *Child Development* (pp.435-471). Boston: Pearson Education, Inc.
- Berkowitz, L. (1989). The frustration-aggression hypothesis: Examination and reformulation. *Psychological Bulletin, 106*, 59-73.
- Between Subject Designs*. (2004). Retrieved 2/1/2004, from

http://www.mnstate.edu/chalikia/psy%20232/Overheads/betwen_subject_designs.htm.

- Brown, J. A., & Kelley, M.-L. (2000). *Aggression Questionnaire*. Retrieved 1/13/2004, from <http://80web5.silverplatter.com.proxy.usf.edu/webspirs/previewFullRecordPopupContent>.
- Campbell, J. J., & Frabutt, J. M. (1999). *Familial antecedents of children's overt and relational aggression*. Paper presented at the Biennial Meeting of The Society for Research in Child Development, Albuquerque, NM.
- Carney, P. (1999). *School violence: The school setting is only part of the picture*, from <http://www.cpa.ca/Psynopsis/Violence-99.html>.
- Coe, R. (2000). *What is an 'Effect Size'?* Unpublished manuscript.
- Coie, J. D., & Dodge, K. A. (1983). Continuities and changes in children's social status : a five-year longitudinal study. *Merrill-Palmer Quarterly*, 29, 261-282.
- Coie, J. D., & Dodge, K. A. (1998). The development of aggression and antisocial behavior. In W. V. Damon (Editor-in-Chief) & Nancy Eisenberg (Volume Editor), *Handbook of Child Psychology*, Fifth Edition, Vol. 3: Social, emotional, and personality development (pp. 779-861). New York: J. Wiley.
- Connor, D. F. (2002). *Aggression and Antisocial Behavior in Children and Adolescents: Research and Treatment*. New York: Guilford Press.
- Connor, D. F., Steingard, R. J., Anderson, J. J., & Melloni, R. H. (2003). Gender differences in reactive and proactive aggression. *Child Psychiatry and Human Development*, 33(4), 279-294.

- Crane-Ross, D., Tisak, M. S., & Tisak, J. (1998). Aggression and conventional rule violation among adolescents: social-reasoning predictors of social behavior. *Aggressive Behavior, 24*, 347-365.
- Crick, N. R. (1996). The role of overt aggression, relational aggression, and prosocial behavior in the prediction of children's future social adjustment. *Child Development, 67*, 2317-2327.
- Crick, N. R. (1997). Engagement in gender normative versus nonnormative forms of aggression: links to social-psychological adjustment. *Developmental Psychology, 33*(4), 610-617.
- Crick, N. R., Casas, J. F., & Ku, H.-C. (1999). Relational and physical forms of peer victimization in preschool. *Developmental Psychology, 35*(2), 376-385.
- Crick, N. R., Casas, J. F., & Mosher, M. (1997). Relational and overt aggression in preschool. *Developmental Psychology, 33*(4), 579-588.
- Crick, N. R., Nelson, D., Morales, J. R., Cullerton-Sen, C., Casas, J. F., & Hickman, S. E. (2001). Relational victimization in childhood and adolescence: I hurt you through the grapevine. In J. Juvonen & S. Graham (Eds.), *Peer Harassment in School: The Plight of the Vulnerable and Victimized* (pp. 196-214). New York: Guilford Press.
- Crick, N. R., & Nelson, D. A. (2002). Relational and physical victimization within friendships: nobody told me there'd be friends like these. *Journal of Abnormal Child Psychology, 30*(6), 599-607.
- Curtis, M.J. & Stollar, S.A. (2002). Best Practices in System-Level Change. In A. Thomas and J. Grimes (Eds.), Best Practices in School Psychology, IV. Bethesda, MD: NASP Publications.

- Domitrovich, C. E., & Bierman, K. L. (2001). Parenting practices and child social adjustment: multiple pathways of influence. *Merrill-Palmer Quarterly*, 47(2), 235-263.
- Empower. (2002). *An evaluation of the Owing Up Curricula: Educational intervention into the cycle of violence*. Washington DC: The Empower Program.
- Fleming, J.L., & Monda-Amaya, L.E. (2001). Process variables critical for team effectiveness: A Delphi study of wraparound team members. *Remedial and Special Education*, 22(3), 158-171.
- Frey, K. S., Hirschstein, M. K., & Guzzo, B. A. (2000). Second Step: Preventing aggression by promoting social competence. *Journal of Emotional and Behavioral Disorders*, 8(2), 102-120.
- Galen, B. R., & Underwood, M. K. (1997). A developmental investigation of social aggression among children. *Developmental Psychology*, 33(4), 589-600.
- Goldstein, A. P., & Keller, H. (1987). *Aggressive Behavior: Assessment and Intervention*. New York: Pergamon Press.
- Gontkovsky, S. T. (2002). The psychology of aggression: a critique of correctional approaches to treating violent behavior. *Administration and Policy in Mental Health*, 29(6), 525-528.
- Gregg, S. (1998). *School-based progress to promote safety and civility*. Washington DC: Appalachia Educational Lab.
- Grossman, D. C., Neckerman, H. J., Koepsell, T. D., Liu, P. Y., Asher, K. N., Beland, K.,

- et al. (1997). Effectiveness of a violence prevention curriculum among children in elementary school: A randomized controlled trial. *Journal of the American Medical Association*, 277(1605-1611).
- Grottpeter, J. K., & Crick, N. R. (1996). Relational aggression, overt aggression, and friendship. *Child Development*, 67, 2328-2338.
- Hall, G.E. & Hord, S.M. (2001). *Implementing Change. Patterns, Principles, and Potholes*. Needham Heights, MA: Allyn and Bacon.
- Hall, J. A., Herzberger, S. D., & Skowronski, K. J. (1998). Outcome expectancies and outcome values as predictors of children's aggression. *Aggressive Behavior*, 24, 439-454.
- Hall, R. (2004). *Within-Subjects Design*. Retrieved 2/1/2004, from http://web.umn.edu/~psyworld/within_subjects.htm.
- Henggeler, S. W., Clingempeel, W. G., Brondino, M. J., & Pickrel, S. G. (2002). Four-year follow-up of multisystemic therapy with substance-abusing and substance-dependent juvenile offenders. *Journal of the American Academy of Child and Adolescent Psychiatry*, 41(7), 868-875.
- Henggeler, S. W., Rowland, M. D., Halliday-Boykins, C., Sheidow, A. J., Ward, D. M., Randall, J., et al. (2003). One-year follow-up of multisystemic therapy as an alternative to the hospitalization of youths in psychiatric crisis. *Journal of the American Academy of Child and Adolescent Psychiatry*, 42(5), 543-552.
- Henington, C., Hughes, J. N., Cavell, T. A., & Thompson, B. (1998). The role of relational aggression in identifying aggressive boys and girls. *Journal of School Psychology*, 36(4), 457-477.

- Huessman, L. R. (2002). *The Peer-Nomination Scale for Children*. Ann Arbor: Research Center for Group Dynamics.
- Huessman, L. R., & Guerra, N. G. (1997). Children's normative beliefs about aggression and aggressive behavior. *Journal of Personality and Social Psychology*, *72*(2), 408-419.
- Iverson, A.M. (2002). Best practices in problem-solving team structure and process. In A. Thomas and J. Grimes (Eds.), Best Practices in School Psychology, IV. Bethesda, MD: NASP Publications.
- Jarvinen-Keltikangas, L., & Pakaslahti, L. (1999). Developmental of social problem-solving strategies and changes in aggressive behavior: a 7-year follow-up from childhood to late adolescence. *Aggressive Behavior*, *25*, 269-279.
- Knoff, H.M. (2002). The Stop & Think Social Skills Program: Teaching Children Interpersonal, Problem-Solving and Conflict Resolution Skills. Retrieved 9/17/04, from <http://www.nasponline.org/publications/cq306stopthink.html>.
- Kratochwill, T.R., & Stoiber, K.C. (2002). Evidence-based interventions in school psychology: conceptual foundations of the procedural and coding manual of division 16 and the society for the study of school psychology task force. *School Psychology Quarterly*, *17*(4), 341-389.
- Lambert, S. (2001). *Test Review: The Aggression Questionnaire*. Retrieved 1/27/04, from <http://aac.ncat.edu/newsnotes/y01.sum.html>.
- Leff, S. S., Power, T. J., Manz, P. H., Costigan, T. E., & Nabors, L. A. (2001). School-based aggression prevention programs for young children: current status and implications for violence prevention. *School Psychology Review*, *30*(3), 344-353.

- Lipsey, M. W., & Wilson, D. B. (1998). Effective intervention for serious juvenile offenders: A synthesis of research. In R. Loeber & D. P. Farrington (Eds.), *Serious and violent juvenile offenders: Risk factors and successful interventions* (pp. 313-345). Thousand Oaks, CA: Sage Publications.
- Loudin, J. L., Loukas, A., & Robinson, S. (2003). Relational aggression in college students: examining the roles of social anxiety and empathy. *Aggressive Behavior*, 29, 430-439.
- MacDonald, C. D., D'Amico, L., & O'Laughlin, E. M. (2000). *Relational Aggression and Victimization in Middle-School Students*. Memphis: Conference on Human Development.
- Masten, A. S., Morison, P., & Pelligrini, D. S. (1985). A revised class play method of peer assessment. *Developmental Psychology*, 21(3), 523-533.
- Mattis, J.S. (2002). Grappling with culture, class, and context in cross-cultural research and intervention. *Prevention & Treatment*, 5(11),
- McEvoy, M. A., Estrem, T. L., Rodriguez, M. C., & Olson, M. L. (2003). Assessing relational and physical aggression among preschool children" intermethod agreement. *Topics in Early Childhood Special Education*, 23(2), 53-64.
- Nangle, D. W., Erdley, C., Carpenter, E. M., & Newman, J. E. (2002). Social skills training as a treatment for aggressive children and adolescents: a developmental-clinical integration. *Aggression and Violent Behavior*, 7(2), 169-199.
- Nelson, D. A., & Crick, N. R. (1999). Rose-Colored Glasses: Examining the social information-processing of prosocial young adolescents. *Journal of Early Adolescence*, 19(1), 17-38.

- Olweus, D. (1997). Bully/victim problems in school: facts and intervention. *European Journal of Psychology of Education, 12*(4), 495-510.
- Osterman, K., Bjorkqvist, K., Lagerspetz, K., M.J., Kaukiainen, A., Landau, S. F., Fraczek, A., et al. (1998). Cross-cultural evidence of female indirect aggression. *Aggressive Behavior, 24*, 1-8.
- Parker, J. G., & Asher, S. R. (1987). Peer relations and later personal adjustment: Are low-accepted children at risk? *Psychological Bulletin, 102*, 357-389.
- Phelps Roecker, C. E. (2001). Children's responses to overt and relational aggression. *Journal of Clinical Child Psychology, 30*(1), 240-252.
- Prinstein, M. J., Boergers, J., & Vernberg, E. M. (2001). Overt and relational aggression in adolescents: social-psychological adjustment of aggressors and victims. *Journal of Clinical Child Psychology 30, 4*, 479-491.
- Puig, M., Lambert, M., Rowan, G., Winfrey, T., Lyubansky, M., Hannah, S. & Hill, M. (2003). Behavioral and emotional problems among Jamaican and African American children, ages 6 to 11: teacher reports versus direct observations. *Journal of Emotional and Behavioral Disorders, 7*(4), 240-250.
- Rauste-von Wright, M. (1989). Physical and verbal aggression in peer groups among Finnish adolescent boys and girls. *International Journal of Behavioral Development, 12*(4), 473-484.
- Richardson, D. R., & Green, L. R. (1999). Social sanction and threat explanations on direct and indirect aggression. *Aggressive Behavior, 25*, 425-434.
- Rodkin, P.C., Farmer, T.W., Pearl, R., & Van Acker (2000). Heterogeneity of Popular

- Boys: Antisocial and Prosocial Configurations. *Developmental Psychology*, 36(1), 14-24.
- Rodkin, P. C., & Hodges, E. V. E. (2003). Bullies and victims in the peer ecology: four questions for psychologists and school professionals. *School Psychology Review*, 32(3), 384-400.
- Roosa, M.W., & Gonzales, N.A. (2000). Minority issues in prevention: Introduction to special issue. *American Journal of Community Psychology*, 28, 145-148.
- Rubin, K. H., Hastings, P., Chen, X., Stewart, S., & McNichol, K. (1998). Intrapersonal and maternal correlates of aggression, conflict, and externalizing problems in toddlers. *Child Development*, 69(6), 1614-1629.
- Salmivalli, C., Kaukiainen, A., & Lagerspetz, K., M.J. (2002). Aggression and sociometric status among peers: Do gender and type of aggression matter? *Scandinavian Journal of Psychology*, 41(1), 17-24.
- Schunk, D. H. (2004). *Learning Theories: An Educational Perspective* (Vol. 4). Upper Saddle River: Pearson Education, Inc.
- Second Step violence prevention curriculum: Steps for successful implementation.* (1998). from <http://www.cfchildren.org/comboimp.pdf>.
- Sheidow, A. J., & Woodford, M. S. (2003). Multisystemic Therapy: an empirically supported, home-based family therapy approach. *The Family Journal: Counseling and Therapy For Couples and Families*, 11(3), 257-263.
- Tapper, K., & Boulton, M. (2000). Social representations of physical, verbal, and indirect aggression in children: sex and age differences. *Aggressive Behavior*, 26, 442-454.

- Taub, J. (2002). Evaluation of the Second Step violence prevention program at a rural elementary school. *School Psychology Review*, 31(2), 186-201.
- Tilly, W.D. (2002). Best Practices in school psychology as a problem-solving enterprise. In A. Thomas and J. Grimes (Eds.), Best Practices in School Psychology, IV. Bethesda, MD: NASP Publications.
- Tomada, G., & Schneider, B. H. (1997). Relational aggression, gender, and peer acceptance: invariance across culture, stability over time and concordance among informants. *Developmental Psychology*, 33(4), 601-609.
- Toppo, G. (2003, October 21, 2003). School violence getting an early start this year. *The Salt Lake Tribune*.
- Turiel, E. (1983). *The development of social knowledge: morality and convention*. New York: Cambridge University Press.
- Van Acker, R., & Talbott, E. (1999). The school context and risk for aggression: implications for school-based prevention and intervention efforts. *Preventing School Failure*, 44(1), 12-28.
- Van Schoiack-Edstrom, L., Frey, K. S., & Beland, K. (2002). Changing adolescents' attitudes about relational and physical aggression: an early evaluation of a school-based intervention. *School Psychology Review*, 31(2), 201-216.
- Vernberg, E. M., & Gamm, B. K. (2003). Resistance to violence prevention interventions in schools: barriers and solutions. *Journal of Applied Psychoanalytic Studies*, 5(2), 125-138.
- Viadero, D. (2003). Tormentors. *Education Week*, 22(18), 24-27.
- Viadero, D. (2003, May 14). Two Studies Highlight Links Between Violence,

- Bullying by Students. *Education Week On the Web*. Retrieved April 14, 2004, from <http://www.edweek.org/ew/ewstory.cfm?slug=36bully.h22>.
- Wampold, B.E. (2002). An examination of the bases of evidence-based interventions. *School Psychology Quarterly*, *17*(2), 500-507.
- Watson, M. (2002). *Longitudinal study of the antecedents of aggression*. Retrieved April 30, 2002, from http://www.gse.harvard.edu/~hci/hci/programs/cav/cav_m_4_30_02.html.
- Werner, N. E., & Silbereisen, R. K. (2003). Family relationships quality and contact with deviant peers as predictors of adolescent problem behaviors: the moderating role of gender. *Journal of Adolescent Research*, *18*(5), 454-480.
- Xie, H., Cairns, R. B., & Cairns, B. D. (2002). The development of social aggression and physical aggression: a narrative analysis of interpersonal conflict. *Aggressive Behavior*, *28*, 341-355.
- Yoon, J. S., Hughes, J. N., Cavell, T. A., & Thompson, B. (2000). Social cognitive differences between aggressive-rejected and aggressive-nonrejected children. *Journal of School Psychology*, *38*(6), 551-570.
- Young, E. L., & Raffaele Mendez, L. M. (2003). The mental health professional's role in understanding, preventing, and responding to student sexual harassment. *Journal of Applied School Psychology*, *19*(2), 7-23.
- Yu, D.L., & Seligman, M.E.P. (2002). Preventing depressive symptoms in Chinese children. *Prevention & Treatment*, *5*, Article 9. Available on the World Wide Web: <http://www.journals.apa.org/prevention/volume5/pre0050009a.html>.

APPENDICES

Appendix A: Passive Parent Consent Form

Parent Passive Consent Form University of South Florida Information for People Who Take Part in Research Studies

The following information is being presented to you to help you decide whether or not you want to allow your child to participate in a minimal risk research study. Anything you do not understand, please ask the principal investigator.

Title of Study: Helping Break the Cycle of Youth Violence and Aggression: A Program Evaluation of the *Owning Up* Curricula

Principal Investigator: Stephanie T. Mihalas

Study Location: _____

Dear Parents:

We are writing to request permission for your son/daughter to participate in a research study. The purpose of this study is to evaluate the effectiveness of a program, *Owning Up*, which is designed to promote social skills and empower students to effectively deal with violence and aggression. Your son/daughter has been asked to participate because she/he has been identified as a participant in the “PET” program and this pool of children will be receiving the intervention as part of their curriculum during the school day.

This study is being conducted for Ms. Mihalas’s Ed.S. thesis requirement through the University of South Florida. The *Owning Up* curriculum is designed in a manner that engages your child in a series of activities including discussions, role-play, and journal reflections. No particular discomforts, inconveniences, or risks are expected as a result of participating in this study; however, if your son/daughter becomes uncomfortable at any time during the research study, he/she may discontinue participation. An alternative of the study would be to have your child not participate in this study.

If you agree to allow your child to participate in the research study, he/she will be administered 3 self-report measures related to attitudes and beliefs about overt and covert aggression. Additionally your child will also be administered a peer-nomination report as part of the assessment process. This will occur once before your child has participated in the *Owning Up* program and once after the implementation of the program. Again, there are no expected risks associated with this process.

If you agree to allow your child to participate in the research study, there are several anticipated benefits. These benefits include increasing your child’s awareness of gender-based violence issues, providing your child with skills and tactics to deal with perpetrators, empowering your child to cope with aggression at school and in your community.

If you decide to allow your child to participate in the research study, s/he will receive a lesson each week for 50 minutes over an 8-week period. Your child will receive the intervention at the same time s/he usually receives specializing programming during the “PET” period. Therefore, your child will not be missing any academic or social time during the school day. Your child will not be paid for participation in this study.

Any information gathered from the study and can be identified with either you or your child will remain confidential except under specific circumstances (e.g., reports of harm to self or others), which are required by law. In addition, authorized research investigators, the Department of Health and Human Services and the USF Institutional Review Board, and its staff, and any other individuals acting on behalf of USF may inspect your child’s records from this research project. The information will not be shared with anyone in your child’s school. The information gained from this study will be of use in determining if the *Owning Up* curricula is effective in changing violent and aggressive

Appendix A (continued)

attitudes and behaviors. The study is not concerned with individual responses, but rather the collection of data from the entire group. The results of this study may be published, but they will in no way identify your child.

You are free to decide to discontinue your child’s participation in the study at any time without negative consequences. Participation is strictly voluntary. If you have any questions about this research study contact Stephanie Mihalas at (813) 971-1838 or Dr. Linda Raffaele Mendez at (813) 974-1255.

If you decide to allow your son/daughter to participate in this study, you are agreeing to the following statements:

By signing this consent form I agree that:

- I have fully read or have had read and explained to me in my native language this informed consent describing this research project.
- I have had the opportunity to question one of the persons in charge of this research and have received a satisfactory answer.
- I understand that my child is being asked to participate in research. I understand the risks and benefits, and I freely give my consent for my child to participate in this study as outlined in this form, under the conditions indicated in it.
- I have been given a signed copy of this informed consent form, which is mine to keep.

If you agree to allow your child to participate in this study please sign below and return the form as soon as possible.

I agree to have my child participate in the study as described in this document.

Signature of Parent or Legal Guardian

Date

Print your name here

Print your child’s name here

I certify that participants have been provided with an informed consent form that has been approved by the University of South Florida’s Institutional Review Board that contains the nature, demands, risks, and benefits involved in participating in this study. I further certify that a phone number has been provided in the event of additional questions.

Stephanie T. Mihalas
Research Investigator

Printed name

Date

Linda Raffaele Mendez, Ph.D.
Faculty Sponsor

Printed name

Date

Appendix B: The Aggression Questionnaire

Publisher does not permit re-print of this measure. If you have any questions, please contact WPS.

Appendix C: The Normative Beliefs about Aggression Scale

Research Number: _____

The following questions ask you about whether you think certain behaviors are **WRONG** or are **OK**. Circle the answer that best describes what you think. Circle **ONE** and only one answer.

Suppose a guy says something bad to another guy, John.

1. Do you think it's OK for John to scream at him?

It's perfectly OK Sort of OK It's sort of wrong It's really wrong

2. Do you think it's OK for John to hit him?

It's perfectly OK Sort of OK It's sort of wrong It's really wrong

Suppose a guy says something bad to a girl.

3. Do you think it's wrong for the girl to scream at him?

It's perfectly OK Sort of OK It's sort of wrong It's really wrong

4. Do you think it's wrong for the girl to hit him?

It's perfectly OK Sort of OK It's sort of wrong It's really wrong

Suppose a girl says something bad to another girl, Mary.

5. Do you think it's OK for Mary to scream at her?

It's perfectly OK Sort of OK It's sort of wrong It's really wrong

6. Do you think it's OK for Mary to hit her?

It's perfectly OK Sort of OK It's sort of wrong It's really wrong

Suppose a girl says something bad to a guy.

7. Do you think it's wrong for the guy to scream at her?

It's perfectly OK Sort of OK It's sort of wrong It's really wrong

Appendix C (continued)

8. Do you think it's wrong for the guy to hit her?

It's perfectly OK Sort of OK It's sort of wrong It's really wrong

Suppose a guy hits another guy, John.

9. Do you think it's wrong for John to hit him back?

It's perfectly OK Sort of OK It's sort of wrong It's really wrong

Suppose a guy hits another girl.

10. Do you think it's OK for the girl to hit him back?

It's perfectly OK Sort of OK It's sort of wrong It's really wrong

Suppose a girl hits another girl, Mary.

11. Do you think it's wrong for Mary to hit her back?

It's perfectly OK Sort of OK It's sort of wrong It's really wrong

Suppose a girl hits a guy.

12. Do you think it's OK for the guy to hit her back?

It's perfectly OK Sort of OK It's sort of wrong It's really wrong

13. In general, it is wrong to hit other people.

It's really wrong It's sort of wrong It's sort of OK It's perfectly OK

14. If you're angry, it is OK to say mean things to other people.

It's perfectly OK It's sort of OK It's sort of wrong It's really wrong

15. In general, it is OK to yell at others and say bad things.

It's perfectly OK It's sort of OK It's sort of wrong It's really wrong

16. It is usually OK to push or shove other people around if you're mad.

It's perfectly OK It's sort of OK It's sort of wrong It's really wrong

Appendix C (continued)

17. It is wrong to insult other people.

It's really wrong It's sort of wrong It's sort of OK It's perfectly OK

18. It is wrong to take it out on others by sharing mean things when mad.

It's really wrong It's sort of wrong It's sort of OK It's perfectly OK

19. It is generally wrong to get into physical fights with others.

It's really wrong It's sort of wrong It's sort of OK It's perfectly OK

20. In general, it is OK to take your anger out on others by using physical force.

It's perfectly OK It's sort of OK It's sort of wrong It's really wrong

21. It is wrong to leave out someone from an activity, just to make them feel bad.

It's really wrong It's sort of wrong It's sort of OK It's perfectly OK

22. In general, it is OK to spread rumors about someone to make other kids not like them.

It's perfectly OK It's sort of OK It's sort of wrong It's really wrong

Appendix D: Behavior Scale-Relational aggression

Research number: _____

For the following questions, please circle the answer that is most appropriate for you.

Think about the past two weeks. How often have you:

1. Spread rumors that weren't true:

Never Once Twice 3 or 4 times 5 or more times

2. Tried to exclude someone from group activities:

Never Once Twice 3 or 4 times 5 or more times

3. Tried to get other people to stop playing with or stop liking somebody you were mad at:

Never Once Twice 3 or 4 times 5 or more times

4. Threatened to stop being somebody's friend in order to hurt them or to get what you wanted from them:

Never Once Twice 3 or 4 times 5 or more times

5. Had someone spread untrue rumors about you:

Never Once Twice 3 or 4 times 5 or more times

6. Had someone intentionally ignore or stop talking to you because they were mad at you:

Never Once Twice 3 or 4 times 5 or more times

7. Been excluded from group activities by someone who was mad at you:

Never Once Twice 3 or 4 times 5 or more times

8. Had somebody try to get others to stop playing with or stop liking you:

Never Once Twice 3 or 4 times 5 or more times

9. Had someone threaten to stop being your friend in order to hurt you or to get what they wanted from you:

Never Once Twice 3 or 4 times 5 or more times

10. Started a fist fight or shoving match:

Never Once Twice 3 or 4 times 5 or more times

11. Hit someone because you did not like what they said or did:

Never Once Twice 3 or 4 times 5 or more times

Appendix D (continued)

- | | | | | | |
|---|-------|------|-------|--------------|-----------------|
| 12. Been in a fight with members of a gang: | Never | Once | Twice | 3 or 4 times | 5 or more times |
| 13. Carried a weapon, like a knife or gun: | Never | Once | Twice | 3 or 4 times | 5 or more times |
| 14. Hit someone because you did not like what they said or did: | Never | Once | Twice | 3 or 4 times | 5 or more times |
| 15. Had a serious fight at school: | Never | Once | Twice | 3 or 4 times | 5 or more times |
| 16. Gotten in an argument with schoolmates: | Never | Once | Twice | 3 or 4 times | 5 or more times |
| 17. Threatened to hit or beat up others: | Never | Once | Twice | 3 or 4 times | 5 or more times |

Appendix E: The Peer Nomination Scales for Children

Instructions (to be read by proctor): “We are going to ask you to do a number of things for us today. These things are not tests. We just want to know how you feel about some things. All of the answers you give us today will be confidential. You will know your answers and we will know your answers, but no one in the class will know, not even your teachers. Now there are some rules that we all must follow. The first rule is: don’t talk to other students while you are doing these things. Second rule: don’t shout your answers; we don’t want you to tell anyone your answer. Third rule: keep your eyes on your own papers. Fourth rule: don’t turn any pages until we tell you to. Once again the rules are (repeat them).

We will pass out the booklets now. Please put a check mark next to the name of the child that you feel best answers the questions I am about to ask. If you think that the question fits more than one student, you can put check marks to as many student names as you would like. Also, if you feel like no one is the answer to one of these questions, please put a check mark next to “NO GIRL” or “NO BOY.” If you have any questions, please ask them now.

Questions (to be asked by proctor):

1. Please write your research number on the first page. Please remember that I would like you to answer all of these questions thinking about what has happened **in the past 2 weeks.**
2. Who has shared what they have with others?
3. Who has not follow directions?
4. Who has engaged in fights with authority figures like teachers and the principal?
5. Who has been getting along well with others?
6. Who has usually been chosen last to join in group activities?
7. Who has been picked on by other kids?
8. Who has given dirty looks or gestures to other kids?
9. Who has made up stories and lies to get other kids in trouble?
10. Who has done things to bother other kids?
11. Who has helped other kids?
12. Who are the kids in your class that you really don’t like?
13. Who has been hit and pushed around by other kids?
14. Who has started a fight over nothing?
15. Who has pushed or shoved other children?
16. Who has been getting into trouble more than a few times?
17. Who has said mean things?
18. Who has stolen?
19. Who has done nice things for helping other people?
20. Who has not retaliated back when picked on to fight?
21. Who has not been invited to parties or out-of-school activities?
22. Who has lost friends because other kids are talking behind their back?
23. Who has spread rumors to ruin someone’s reputation?

Appendix F: Consumer Satisfaction Survey-Student



So you have finally completed the *Owning Up* curriculum! We want to understand your experience with the program, so please provide us with some honest feedback so we can make it even better in the future. Please note that all your answers will be held in confidentiality and used for research purposes only. However, please provide your own personal identification number here _____ for tracking purposes.

1. Overall, how would you rate your interest in the program (1=not interested and 5=very interested)?

1 2 3 4 5

Please tell us briefly why you scored it the way you did:

2. Do you feel that you had a chance to learn new things about violence and aggression (1=I learned nothing and 5=I learned A LOT)?

1 2 3 4 5

Please tell us briefly why you scored it the way you did:

3. Do you feel that you would recommend other students your age to participate in this program (1= I would not recommend and 5= I would highly recommend)?

1 2 3 4 5

Please tell us briefly why you scored it the way you did:

4. If there was one thing that you would definitely keep the same about the program what would it be? Please explain your answer.

5. What about one thing that you definitely would want to change about the program? Please explain your answer.

Appendix F (continued)

6. What lesson in *Owning Up* did you find most valuable to you? What about least valuable?

Appendix G: Owning Up Implementation Checklist

Name: _____ Date: _____
 School: _____ Lesson # and name: _____
 Number of students in attendance: _____

Directions: At the end of the lesson, please place an “X” by items to indicate events that occurred during the lesson. Thank you for your participation!

| Check “X” here | Statement of completion | Comments |
|----------------------|--|----------|
| | I asked students to generate or evaluate solutions to a social problem | |
| | I encouraged students to write in their journals during class | |
| | I took the time in the beginning of class to talk about last week’s lesson | |
| | At the end of this lesson, I asked students how they felt about the lesson overall | |
| | I encouraged students to think about questions and issues that were not resolved in class and to write them down for next week | |
| | I maintained a positive and constructive learning environment | |
| | I explained the objectives of the lesson in the beginning of class | |
| | I conducted at least one role-play or group activity | |
| | I was able to manage behavior problems effectively | |
| | I passed out any handouts that were associated with the lesson | |
| | I encouraged whole group participation | |
| | I used materials that were designated in the handbook for the lesson | |

Please use the back of the pages if you need more room.

How did you feel about your instruction today?

Did any major issues arise either with your students or with the curriculum?

Appendix G (continued)

Is there anything else you would like to share?

On a scale from 0-5, with 0 being completely unengaged to 5 very engaged, how would you rank your students today? _____. Please provide a short rationale for your rank choice.

Appendix H: Exit Interview

Owning Up Evaluation Exit Interview

Name of facilitator:

Date:

Time:

Questions to be posed by the researcher in an informal face-to-face interview

1. Overall how successful do you think the program was in altering student's beliefs and behaviors related to aggression and violence?
 - a. What could have made the program content or process more successful?
2. What were some key inhibiting factors that may have limited your effectiveness in teaching the different lessons?
 - a. How might these be resolved in future implementation efforts?
 - b. On the flip side, what factors helped to propel the curriculum forward?
Query for parent support, student interest, administration support
3. In terms of the program content, what were the primary areas of strengths? What about weaknesses?
 - a. If you could add or delete anything to the curricula, what would it be?
Why?
 - b. Do you think this curriculum was appropriate for the demographics of the population you were working with? If not, who would it be more suitable for?
4. Did you experience any issues with fitting the lesson content into 8 sessions? Was it feasible? Do you think that minimizing the sessions into 8 may have decreased the benefit for the students?
5. How do you personally feel about your role as the facilitator?
 - a. Is there anything you personally would have done differently?
 - b. What are your own personal strengths and weaknesses that may have effected your instruction?
6. Any other comments, questions, concerns?

Appendix I: Observer Review

Name of facilitator:

Date:

School:

Lesson # and name:

Name of observer:

Directions: During your observation, please place a check next to what you observe the facilitator doing. Additionally, please rate the quality by which the facilitator is meeting the designated criteria on a scale of (1-5); 1, suggested that the quality is extremely low and 5, suggesting that the quality is superior. Please use the back of this sheet if you need more room for your answers.

| Check "X" here; note quality (1-5) | Statement of completion | Comments |
|------------------------------------|--|----------|
| | Facilitator asks students to generate or evaluate solutions to a social problem | |
| | Facilitator encourages students to write in their journals during class | |
| | Facilitator took the time in the beginning of class to talk about last week's lesson | |
| | At the end of this lesson, the facilitator asked students how they felt about the lesson overall | |
| | Facilitator maintained a positive and constructive learning environment | |
| | Facilitator managed behavior problems effectively | |
| | Facilitator encouraged whole group participation | |
| | Facilitator acknowledged and respected multicultural issues | |
| | Facilitator handled sensitive topics appropriately | |

How did you feel about the overall quality of the facilitation?

Did any major issues arise either with how the students were taught in the class or how the facilitator handled situations?

Appendix I (continued)

Is there anything else you would like to share about your observation?

On a scale from 0-5, with 0 being completely unengaged to 5 very engaged, how would you rank the students today? _____. Please provide a short rationale for your rank choice.

Appendix J: Project Timeline

Month of February

Week of February 16th

- Meet with implementation team to discuss process and procedures (date TBD)
- February 20: Proposal Meeting

Week of February 23rd

- Attempt to expedite IRB approval
- Pilot measures on volunteers (date TBD)
- February 26th-28th: Facilitators are trained in D.C.

Month of March

Week of March 1st

- Packets should be copied and distributed to proctors by end of week
- Meeting with parents (date TBD)

Week of March 8th

- March 8th: Parent permission forms distributed to students
- March 12th: Touch base with SIS to determine rate of return of forms
- Pilot measures on 5 volunteers

Week of March 15th:

- Researcher and facilitator meet with participants (date TBD)
- Administration of pre-test (date TBD)

Week of March 22nd

- Participants spring break

Week of March 29th

- Intervention begins at all schools

Month of April

Week of April 19th

- Deadline for first outside observation to take place
- Facilitators to send checklists to researcher

Month of May

Week of May 10th

- Researcher to set up exit interviews with facilitators
- Last week for second observation to occur

Week of May 17th

- Last day of intervention

Appendix J (continued)

- Facilitators to send checklists to researcher
- Exit interviews take place (date TBD)

Week of May 24th

- Administration of post-test measures

Appendix K: Child Assent Form

I want to be a part of the study that is taking place at my school during the time that I participate in my regular “PET” class. I understand that I will be participating in lessons that discuss topics like violence, aggression, and harassment. I was told that I will participate in different kinds of surveys twice during the time that I will be participating in the “*Owning Up*” lessons. I understand that I can change my mind at any time and remove myself from participating in the program or the study and I will not get into trouble.

Your Name

Date

Witness (Print Name)

Witness (Sign Name)

Appendix L: Control School Lesson Plan Information

Name:

School:

Date:

Number of students in attendance:

Please provide a brief explanation of your lesson plan today.

Did you have any discussions related to gender-based violence or aggression?

_____ (Yes/No). If yes, please describe the discussion in detail.

Appendix M: Correlations of Measures

Figure 2A. Correlation between AQ-Total and Nbttotal for pre-test.

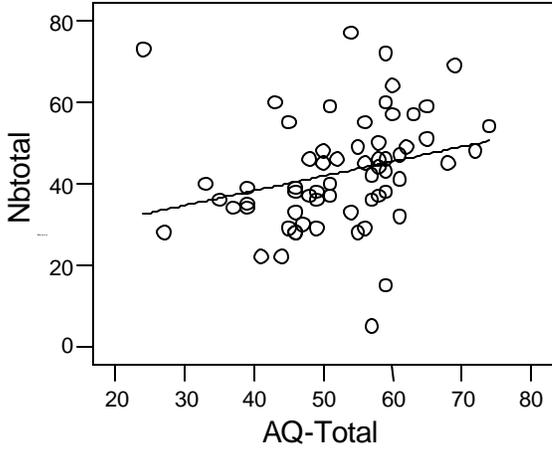


Figure 2B. Correlation between AQ-Total and Nbttotal for posttest.

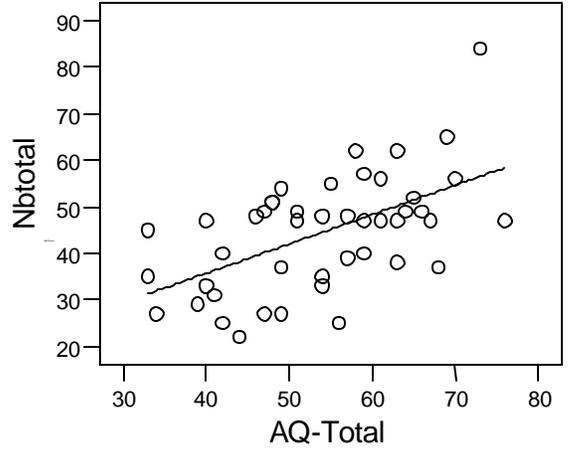


Figure 2C. Correlation between AQ-Total and RbTotal for pre-test.

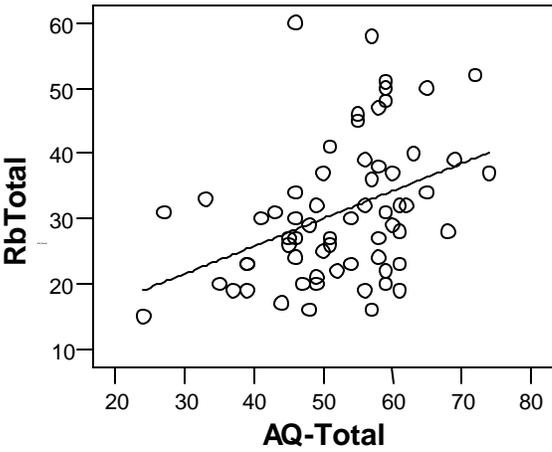


Figure 2D. Correlation between AQ-Total and RbTotal for posttest.

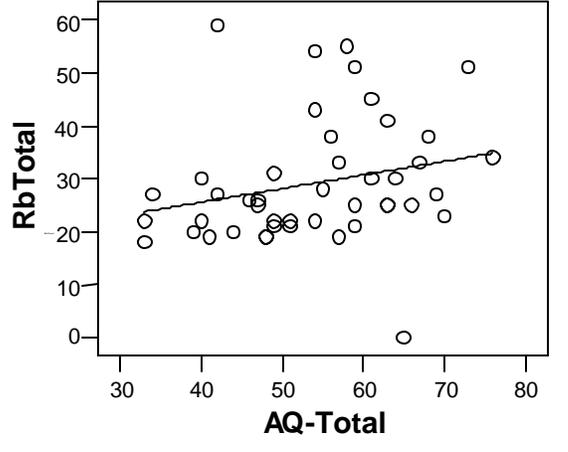


Figure 2E. Correlation between AQ-Total and PNPR+ for pre-test.

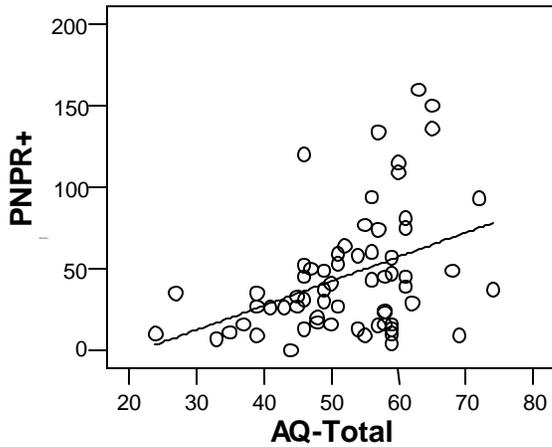


Figure 2F. Correlation between AQ-Total and PNPR+ for posttest.

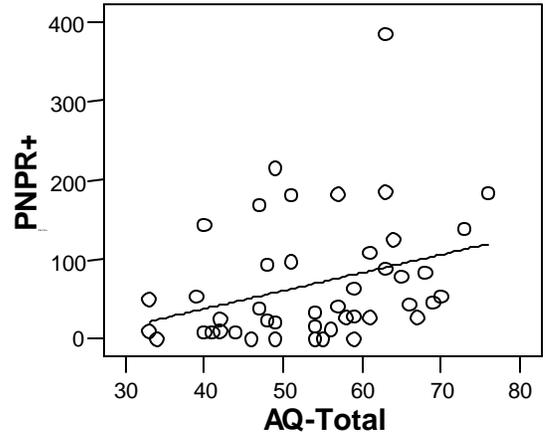


Figure 2G. Correlation between AQ-Total and PNPR- for pre-test.

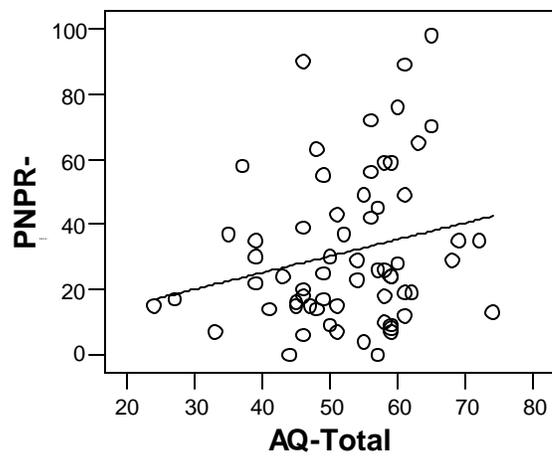


Figure 2H. Correlation between AQ-Total and PNPR- for posttest.

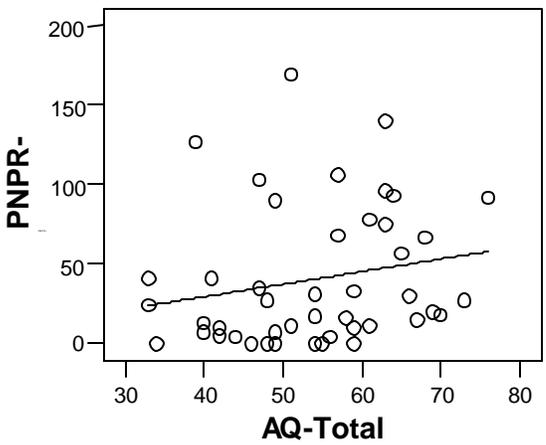


Figure 2I. Correlation between AQ-Total and PNVIC for pre-test.

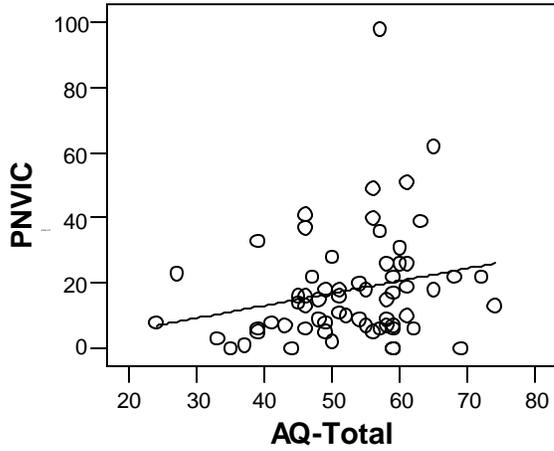


Figure 2J. Correlation between AQ-Total and PNVIC for posttest.

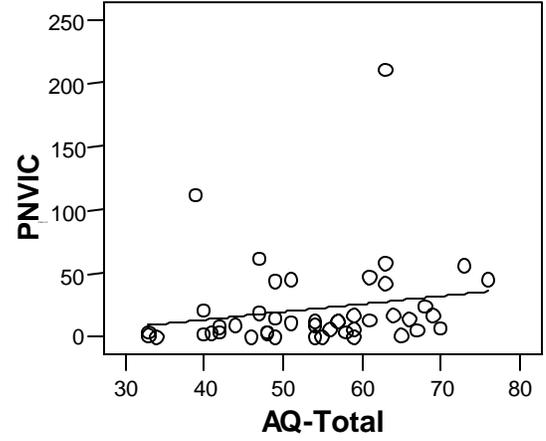


Figure 2K. Correlation between Nbttotal and RbTotal for pre-test.

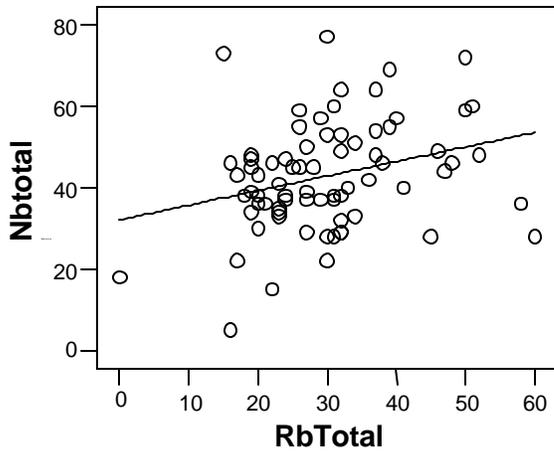


Figure 2L. Correlation between RbTotal and Nbttotal for posttest.

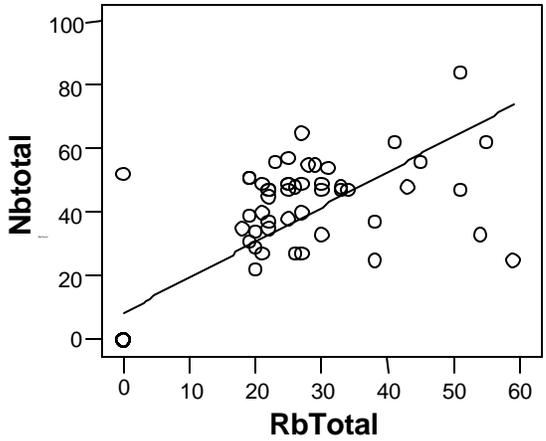


Figure 2M. Correlation between Nbtotal and PNPR+ for pre-test.

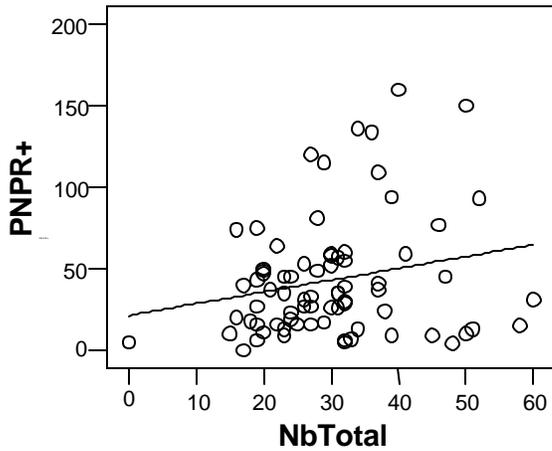


Figure 2N. Correlation between Nbtotal and PNPR+ for posttest.

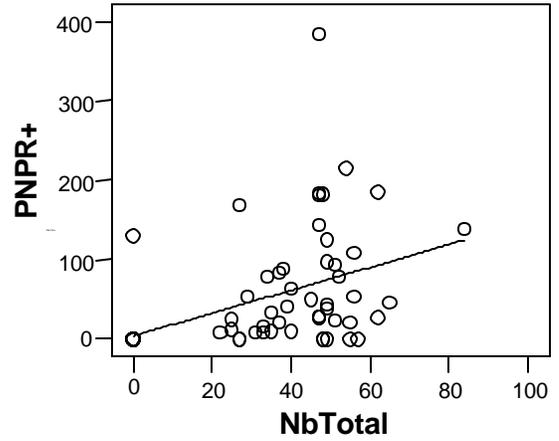


Figure 2O. Correlation between Nbtotal and PNPR- for pre-test.

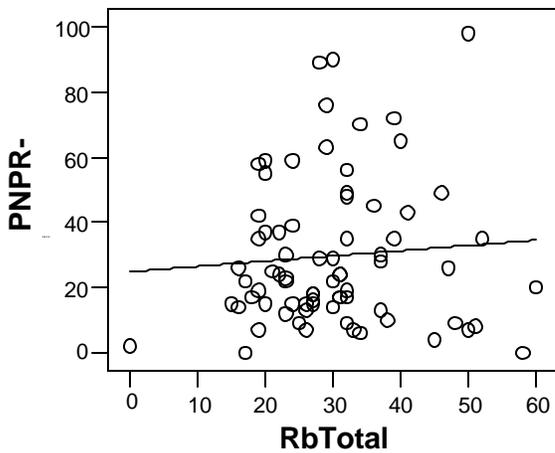


Figure 2P. Correlation between RbTotal and PNPR-.

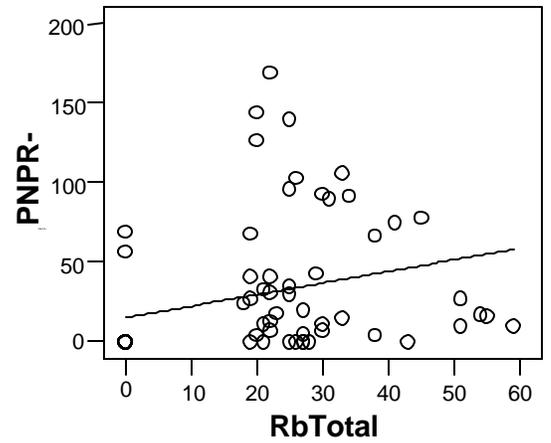


Figure 2Q Correlation between Nbtotal and PNVIC for pre-test.

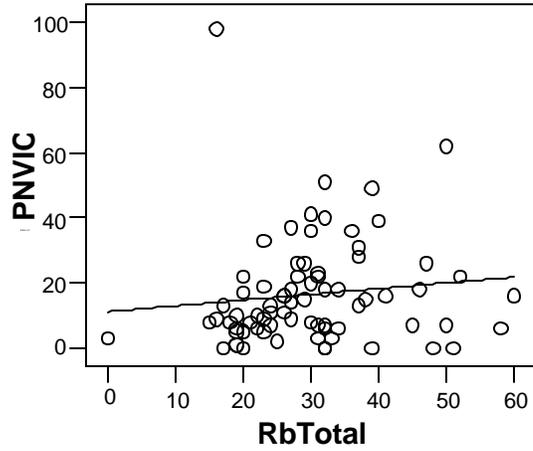


Figure 2R. Correlation between Rbtotal and PNVIC for posttest.

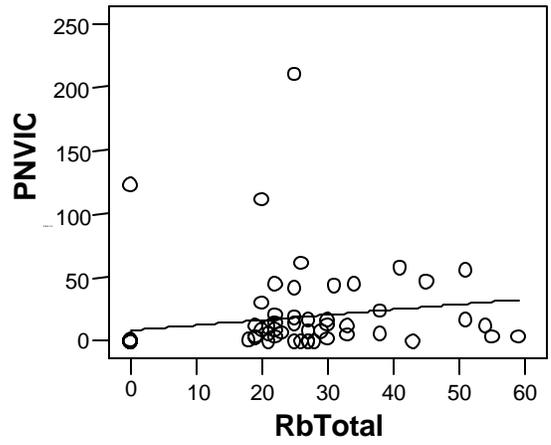


Figure 2S. Correlation between Nbtotal and PNPR+ for pre-test.

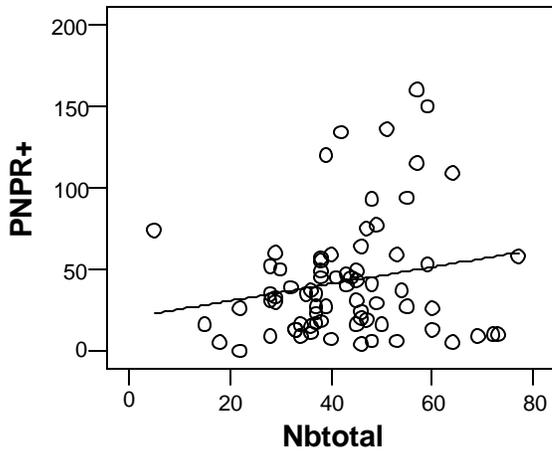


Figure 2T. Correlation between Nbtotal and PNPR+ for posttest.

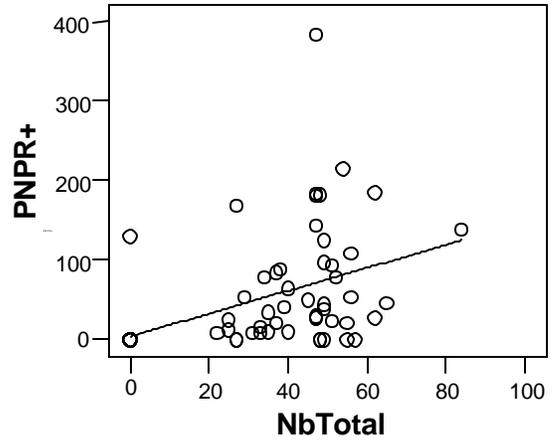


Figure 2U. Correlation between PNPR+ and PNPR- for pretest.

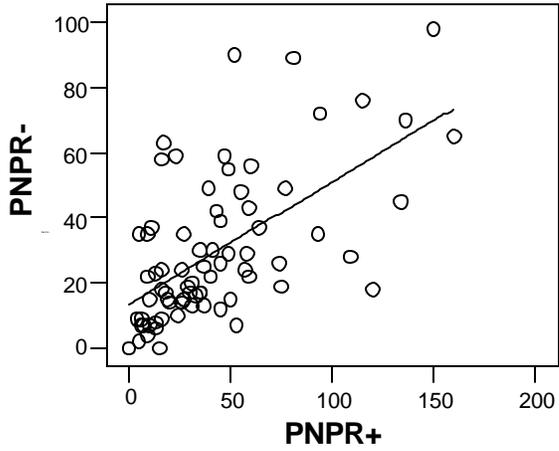


Figure 2V. Correlation between PNPR+ and PNPR- for posttest.

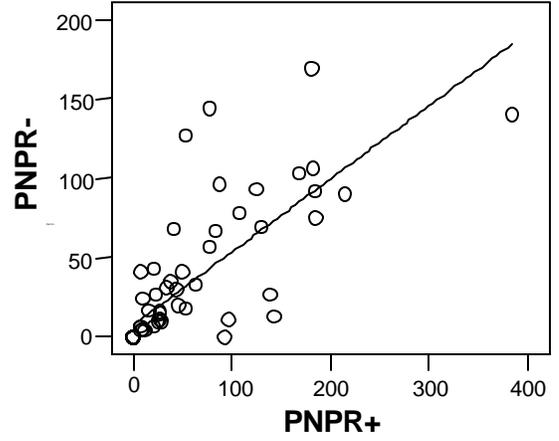


Figure 2W. Correlation between PNPR+ and PNVIC for pretest.

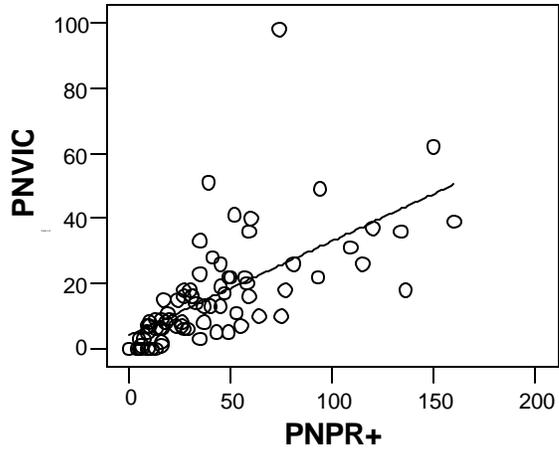


Figure 2X. Correlation between PNPR+ and PNVIC for posttest.

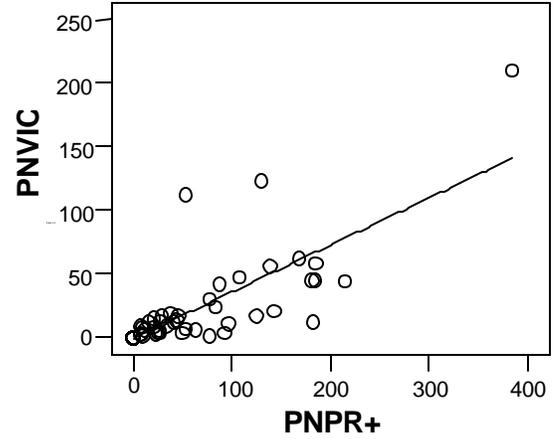


Figure 2Y. Correlation between PNPR- and PNVIC for pretest.

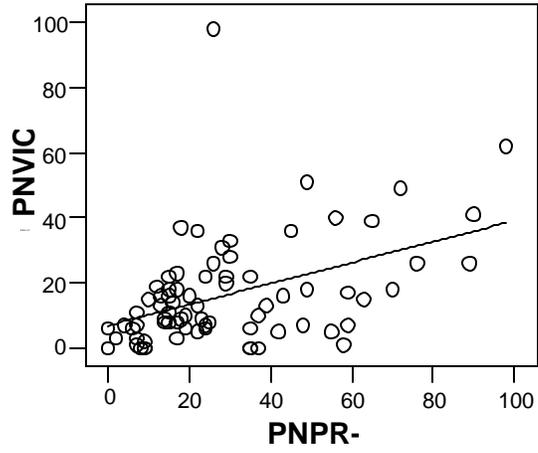


Figure 2Z. Correlation between PNPR- and PNVIC for posttest.

