What is Bullying? An Empirical Investigation into the Construct Validity of Bullying as Measured and Defined.

Margaret Jane McGugan

Thesis submitted to the

Faculty of Graduate and Postdoctoral Studies

in partial fulfilment of the requirements for the degree of

Doctor of Philosophy degree in

Clinical Psychology

School of Psychology

Faculty of Social Sciences

University of Ottawa

© Margaret McGugan, Ottawa, Canada, 2015

Table of Contents

List of Tables

List of Figures

Abstract

Bullying is a subset of aggressive behaviour that is characterised by three features: intention to cause harm, repetitiveness, and an imbalance of power between the perpetrator of the aggression and his or her victim (Olweus, 2010). This definition is widely accepted and widely used in academic research on bullying (Cornell & Bandyopadhyay, 2010), but lacks empirical evidence of construct validity (Finkelhor, Turner, & Hamby, 2012). This dissertation explores the construct validity of bullying in two ways. The first study examines the validity of the definition of bullying in terms of its ability to predict student functioning and identify a distinct group. The second study examines the validity of a widely used bullying measurement strategy in terms of its ability to identify aggression that meets the definition of bullying. Marginal evidence for the construct validity of bullying as currently measured and defined was found. While the characteristics of bullying do predict several measures of functioning above and beyond the presence of generic aggression or victimization, the differences predicted appear to be differences of magnitude only. This indicates that bullying may be related to more severe, but not unique, outcomes. In addition, the individual characteristics of bullying contained within the overall definition (repetition, intention to harm, power imbalance) are shown to be highly related and better thought of indicators that bullying has occurred rather than additive constructs that define a certain subclass of aggressive behaviour. Finally, regarding measurement, the results of this study show that peer-report responses using definitional measurement strategies are not strongly related to the definitional criteria of bullying. Implications for the field of bullying research are discussed. Theoretically informed empirical work on clarifying the bullying construct is identified as a research priority.

Acknowledgements

I would like to thank my friends and family for their support during my doctoral studies. Thank you especially to Laura Hanson, who understood, and John Piccoli, who is my everything.

This research was funded by a CIHR Fredrick Banting and Charles Best Canada Graduate Scholarship Doctoral Award.

What is Bullying? An Empirical Investigation into the Construct Validity of Bullying as Measured and Defined

Bullying is a subset of aggressive behaviour that is characterised by three features: intention to cause harm, repetitiveness, and an imbalance of power between the perpetrator of the aggression and his or her victim (Olweus, 2010). This definition, which has been widely accepted and continues to be widely used in academic research on bullying (Cornell & Bandyopadhyay, 2010; Greif & Furlong, 2006; Jimerson, Swearer, & Espelage, 2010a), was created by Dr. Dan Olweus in the early 1980s in order to facilitate the first large-scale and well-known academic investigations into bullying. Dr. Olweus also created the first version of the Olweus Bully/Victim Questionnaire (BVQ), a widely used self-report method of measuring bullying among school-aged children and adolescents at that time (Cornell & Bandyopadhyay, 2010; Olweus, 2010). This effectively marked the beginning of the academic study of bullying as we know it today.

Over the last three decades the phenomenon of school bullying has been increasingly recognized as an important social and clinical problem and has been the subject of extensive research (Espelage & Swearer, 2003; Salmivalli, 2010; Schäfer, Werner, & Crick, 2002). Approximately 800 peer-reviewed articles on the topic have been published between the late 1970s and 2010 (Cook, Williams, Guerra, Kim, & Sadek, 2010), and a review of the research database PsycINFO confirms that many more are published every month. A considerable body of knowledge has accumulated on the topic. Prevalence studies have documented the occurrence of bullying in 40 countries, in both sexes, and across developmental stages (Craig et al., 2009; Jimerson et al., 2010a; Salmivalli, Peets, & Hodges, 2011). Descriptive studies have shown that bullying can take multiple forms and have identified characteristics that put individuals at risk for bullying behaviour and/or victimization (e.g. Cook, Williams, Guerra,

Kim, et al., 2010; Stassen Berger, 2007). Outcomes studies have shown that involvement in bullying is associated with negative outcomes for students who bully, students who are victimized by bullying, and the classroom at large (e.g. Rigby, 2003; Swearer, Espelage, Vaillancourt, & Hymel, 2010). Headway has even been made on the difficult topic of how best to intervene and prevent bullying problems (Jimerson et al., 2010a), although results in this area have been called "disappointing" (Stassen Berger, 2007).

The increased recognition of bullying as an important issue in the academic literature has been accompanied by increasing public interest and regulatory actions on the topic. Bullying is the subject of regular discussion for popular and news media in North America (e.g. a collection of videos and articles on bullying collected by the television talk show "Ellen" "#UnitedAgainstBullying," 2014) and increased public awareness has led to the introduction of policy and programs related to the topic in many educational institutions and legal jurisdictions (Brank, Hoetger, & Hazen, 2012; Cascardi, Brown, Iannarone, & Cardona, 2014). Specific bullying incidents have even been investigated as human rights violations by human rights tribunals in Canada (CBC News, 2014). Occurrences such as these indicate that bullying has become a major social issue. Academic research on bullying directs cultural discussion, intervention efforts, and legal and other regulatory actions around the issue. As such, it is especially important that research on bullying is empirically based and as scientifically sound as possible.

Statement of the Problem

Despite the considerable interest and body of knowledge that has been accumulated about bullying, the research literature still has some significant areas of weakness. Recently, a number of authors have expressed concern about conceptual issues related to the meaning and utility of the bullying construct (Finkelhor et al., 2012; Hanish et al., 2013; Hellström,

Beckman, & Hagquist, 2013; Rigby, 2012). Questions have been raised about the validity of the current definition of bullying and the way that bullying is typically measured. These concerns can be summarised as concerns about the construct validity of bullying as currently defined and understood. Construct validity is an important theoretical consideration that has two aspects: whether the construct itself is valid and whether a measure being used accurately assesses the construct (Goodwin, 2003). Two areas of concern related to construct validity are highlighted here.

First, the current definition of bullying lacks empirical support. Though the current definition is widely accepted and continues to be widely used (Cornell & Bandyopadhyay, 2010), there is little empirical evidence showing that bullying as defined in this way is distinct from other forms of aggression in terms of causes, outcomes, or optimal intervention methods (Finkelhor et al., 2012; Hanish et al., 2013; Hellström et al., 2013; Rigby, 2012). This is because direct comparisons between bullying activity and other forms of aggressive behaviour are rare. There has also been little research on the definitional criteria themselves, resulting in a lack of empirical evidence that the three characteristics of aggression contained within the definition of bullying would be individually necessary or collectively sufficient for identifying a unique form of aggression known as bullying (Finkelhor et al., 2012). This lack of empirical evidence erodes confidence that the definition accurately defines a valid construct.

Second, there is a concern that methods for measuring bullying may not accurately discriminate the construct from other forms of aggression. A revised version of the Olweus Bullying Questionnaire is the most commonly used assessment tool of bullying involvement (Cornell & Bandyopadhyay, 2010), despite a lack of evidence that it actually assesses the subset of aggressive behaviour that is specified in the definition of bullying (Greif &

Furlong, 2006). Bullying is also commonly assessed using methods that make no attempt to assess definitional criteria and differentiate bullying from other forms of aggression (Furlong, Sharkey, Felix, Tanigawa, & Green, 2010). Problems such as these erode confidence that current measurement strategies accurately capture bullying as defined.

One might expect that by now, after 30 years of research, the validity of both Olweus' definition of bullying and his widely used measuring tool would be well established, but perhaps it is not surprising that empirical work on bullying has, to this point, focused more on practical rather than theoretical concerns. The urgency of identifying and aiding young people who are at risk for harm because of their involvement in peer aggression may outweigh what initially might be seen as purely academic concerns. Intervention efforts, however, rely on the research literature, and clarity of constructs is an important element of ensuring the best research possible. As several prominent researchers on bullying stated, "advances in measurement and conceptualization [of bullying] simultaneously advance researchers' ability to study the factors that increase or decrease the likelihood of bullying, responding to bullying, and the potential for harm from bullying" (Hanish et al., 2013, p. 290). The lack of empirical support and precision surrounding the concept of bullying has the potential to undermine efforts to understand and intervene in the bullying behaviour. Unique correlates and consequences of bullying may be missed if bullying is not defined correctly or if it is not differentiated from other form of aggression when measured. This is especially important moving forward, as our understanding of the possible functions and consequences of aggression within the peer groups of young people become more sophisticated, and we begin to investigate differences between bullying and other types of aggression in a social contexts (e.g. Hawley, Little, & Rodkin, 2007; Salmivalli, 2010). Ensuring that the construct of bullying has construct validity will also help ensure that results across research studies are

comparable and that they aid policy and lawmakers in consistently applying information from the research in real-life settings.

Purpose of the Study

The fundamental purpose of the current research is to empirically explore the construct validity of bullying. Current definitions of bullying distinguish bullying from other forms of aggression on the basis of three characteristics of the aggressive behaviour: that it is repeated, that it is intentionally harmful, and that it occurs in the context of a power imbalance. Despite the widespread adoption of this definition, it is not well validated. This paper will examine the validity of this definition in two ways:

1. By examining the validity of the definition in terms of its ability to predict student functioning and identify a distinct group,
2. By examining the validity of a widely used measurement strategy in terms of its ability to identify aggression that meets the definition of bullying.

Construct validity cannot be established or denied in a single study (Goodwin, 2003). This study, however, should contribute empirical data that, in combination with similar studies, will either increase confidence in the construct validity of bullying or perhaps spur efforts to refine and clarify the meaning and measurement of the construct.

Terminology

In the interest of clarity, I will endeavour to be consistent in my use of terms throughout this research project. Two distinctions will be very important: 1) The distinction between aggression and bullying, and 2) the distinction between victimization and being victimized by bullying. These terms are often used interchangeably in the literature on bullying, contributing to the lack of precision in the bullying concept. The aggressive behaviour of many young people who perpetrate aggression does not meet the criteria for

bullying when strict definitional criteria are applied. Similarly, many young people are victimized by aggressive behaviour but are not victimized repeatedly and intentionally harmed within the context of a power imbalance, as they would have to be to meet the criteria for having been victimized by bullying. I will use the term "bully," "bullying" or "bullying behaviour" to refer to individuals or actions that meet the full criteria for bullying. I will use "aggressor," "aggression" or "aggressive behaviour" to refer to aggressive individuals and actions that do not necessarily meet the criteria for bullying. The phrase "victimized by bullying" will be used to indicate that an individual has been the victim of bullying, whereas "victimized" means that they have been the victims of some type of aggression (which may or may not meet the criteria for bullying). Throughout the project, I will strive to use person-centred language where possible (i.e., person who bullies instead of bully) but at times, terms such as bully, aggressor, victim, victim of bullying, or bully/victim will be used when referring to researcher-created groups defined by specific survey responses. No offence or indication that these are primary characteristics of individual participants is intended.

Chapter 2. Literature Review

Research on Aggression

Aggression, an aspect of human behaviour that is as old as humanity itself (Vaughn & Santos, 2007), has been an area of interest in psychology for well over a century (Dodge, Coie, & Lynam, 2006). Many consistently replicated findings about aggression are "as close as we are likely to come to psychological canons" (Parker, 2007, p. x). Within the last decade, however, the generally accepted view of aggression has begun to shift (see Hawley, Little, & Rodkin, 2007). Where aggressive behaviour was once viewed as universally maladaptive, there is now a growing recognition that some types of aggression may be

adaptive for some individuals in some situations (Vaughn & Santos, 2007). Presented here is a brief review of research on aggression, including research of individuals who may use aggression to achieve higher social status.

What is aggression?

Aggression is defined as any behaviour directed towards another individual that is carried out with the immediate intent of causing harm (Anderson & Bushman, 2002). Others would add to this definition the idea that the victim of the aggression must be motivated to avoid the aggression (Baron & Richardson, 2004). There are many types of aggressive behaviour that have been of interest to researchers. Bullying, as mentioned, is a subtype of aggressive behaviour. So is violence, which is aggression that has extreme harm (e.g., death) as its goal (Anderson & Bushman, 2002). Researchers have also been interested in the effects of aggression on those who are victimized. Peer victimization, "a form of peer abuse in which a child is frequently the target of peer aggression" (Kochenderfer & Ladd, 1996, p. 1305) is a commonly studied concept. Peer victimization is commonly confused with victimization by bullying, despite the fact that in peer victimization, while the individual is repeatedly aggressed against, it may not always be by the same person or within the context of a power imbalance.

Physical damage to the victim is not essential for behaviour to be considered aggressive. In fact, aggression is known to have many forms (Bushman & Huesmann, 2010; Coyne, Nelson, & Underwood, 2011). Physically aggressive acts, such as hitting, biting, and destroying someone's property, are often the most easily observable form of aggression. Two other forms of aggressive behaviour are also commonly recognized: verbal aggression (i.e., hurting people with words) and a more manipulative and subtle form of aggression known alternatively as indirect aggression, relational aggression, or social aggression. This final

form of aggression, which will be called "relational aggression" throughout this paper, operates by harming the victim's social standing or relationships. To call relational aggression indirect aggression (i.e., aggression that does not occur in the presence of the victim) is not necessarily wrong, as it highlights the often covert nature of this type of aggression; however, physical and verbal aggression can also be indirect (e.g., vandalism and written threats), and relational aggression can be, and often is, direct (i.e., aggression that occurs in the presence of the victim), especially among younger children (e.g., telling someone you will not be friends with them unless they invite you to their birthday party).

Just as aggression can have different forms, aggressive behaviour can also have different functions (i.e., purposes for inflicting harm). One prominent distinction is between reactive and proactive aggression (Anderson & Bushman, 2002; Card & Little, 2007; Dodge, 1991). Reactive aggression is impulsive and angry behaviour that is typically a direct response to a perceived threat. Its function is thought to be defensive or retaliative. Proactive aggression is premeditated, deliberate, and calculated. It is thought to function as a method for meeting goals. It is sometimes difficult to distinguish between reactive and proactive aggression because the associated behaviours are similar and the two forms of aggression are often highly correlated within an individual, perhaps because motives for aggression are often mixed (Bushman & Anderson, 2001). Card and Little (2007), however, argue that this high correlation is an artifact of measurement and that different forms of aggression are related to different outcomes and may arise from different underlying processes.

Who is affected by aggression and what are its consequences?

Different forms of aggression are more common among individuals at particular developmental stages and of particular gender. In general, males display more physical aggression than females, while evidence for any gender discrepancy in levels of relational

aggression is mixed (Archer, 2004; Card, Stucky, Sawalani, & Little, 2008). For both sexes, physical aggression peaks around two years of age (Côté, Vaillancourt, Barker, Nagin, & Tremblay, 2007), while the prevalence of verbal and relational aggression increases along with children's verbal and social skills (Coyne et al., 2011; Vitaro, Brendgen, & Barker, 2006). Overall prevalence levels of aggressive therefore depend on contextual factors. One study showed that 46% of sixth graders reported personally being the victim of peer harassment (defined as one student being picked on, made fun of, insulted, bullied, threatened, shoved, or hit by another student) within a four-day span (Nishina & Juvonen, 2005). In the same study, 42% of students reported witnessing peer harassment within the same time span. As such, it appears that exposure to aggression is a part of the school experience for many youth.

At the same time as broad developmental trends in the level of different forms of aggression have been observed, there are also individual differences in levels of aggressive behaviour that remain consistent throughout development (Cook, Williams, Guerra, Kim, et al., 2010). Some individuals simply seem to be more aggressive than others (Côté et al., 2007). Factors that have been shown to contribute to these individual differences include biology, parenting and family characteristics, peer context, and exposure to violent media (Coyne et al., 2011). Aggression appears to be at least partly heritable (Moffitt, 2005), perhaps through a genetic influence on structural features of the brain (Raine, 2008). It is related to harsh parenting and potentially related difficulties with emotion regulation (Chang, Schwartz, Dodge, & McBride-Chang, 2003). Youth who are aggressive also seem to befriend other aggressive youth (Espelage, Holt, & Henkel, 2003) and think differently than other children and adolescents about social situations and the use of aggression (Toblin, Schwartz, Hopmeyer Gorman, & Abou-ezzeddine, 2005). Certain children and adolescents

are also more likely to be victims of aggression than the norm. These youth tend to be physically weaker, display socially unskilled behaviour and academic difficulties, suffer from internalizing problems, and be rejected by peers (Card & Hodges, 2008).

Initial compelling findings on the consequences of aggressive behaviour among children and adolescents showed that both victims of aggression and its perpetrators suffered negative consequences that required intervention (Card, Isaacs, & Hodges, 2009). Victimization often leads to lower self-esteem, increases in internalizing problems such as depression and anxiety, and social withdrawal (Egan & Perry, 1998; Hodges, Boivin, Vitaro, & Bukowski, 1999; Hodges & Perry, 1999; Kochenderfer & Ladd, 1996; Vernberg, 1990). Students who are victimized also tend to have lower grades and dislike school more than their non-victimized peers (Juvonen, Nishina, & Graham, 2000; Kochenderfer & Ladd, 1996). In terms of social outcomes, victimized youth tend to be disliked by most of their peers, have fewer friends than other children, and the friendships that they do have tend to be of poorer quality (Hodges et al., 1999; Hodges & Perry, 1999; Ladd & Troop-Gordon, 2003; Rigby, 2000; Vernberg, 1990). Socially, aggressive children also tend to be widely disliked and befriend mostly other aggressive youth (Cairns, Cairns, Neckerman, & Gest, 1988; Dishion, Andrews, & Crosby, 1995; Newcomb, Bukowski, & Pattee, 1993; Parker & Asher, 1987). Aggressive children are also often disengaged from school and have relatively poor literacy skills (Coie, Terry, Lenox, Lochman, & Hyman, 1995; Miles & Stipek, 2006; Moffitt, 1993), and some evidence suggests that they are also at risk for depressive symptoms (Morrow, Hubbard, McAuliffe, Rubin, & Dearing, 2006). The negative consequences of aggressive behaviour seem to be long term, with aggressive children being at high risk for further delinquency, substance use, school dropout, and criminal behaviour during the later teen years and early adulthood (Farrington, 1991; Nagin & Tremblay, 1999).

Clearly, involvement in aggression is associated with concerning outcomes; however, as our understanding of the consequences of victimization and aggressive behaviour are becoming more sophisticated, researchers are beginning understand that the consequences of aggression or victimization are not the same for all young people involved. Among youth who are aggressive, two subgroups with varying patterns of risk have been identified. It appears that some aggressive individuals may actually benefit from their aggressive behaviour. For some youth, aggressive behaviour can be a successful method of acquiring desired resources, changing other's behaviour, and maintaining group boundaries (Farmer, Xie, & Cairns, 2007; Hawley et al., 2007; Hawley, 1999). In line with this, a growing body of literature now reveals the presence of a small group of children who are both aggressive and popular with their peers (Bagwell, Coie, Terry, & Lochman, 2000; Estell, Cairns, Farmer, & Cairns, 2002; Estell, Farmer, & Cairns, 2007; Farmer, Estell, Bishop, O'Neal, & Cairns, 2003; LaFontana & Cillessen, 2002; Lease, Musgrove, & Axelrod, 2002; Rodkin, Farmer, Pearl, & Acker, 2006; Rodkin, Farmer, Pearl, & Van Acker, 2000). The fact that some aggressive young people are popular does not mean that their aggression isn't anti-social or harmful to their victims, but it does suggest that for some individuals, aggression is not universally maladaptive for the perpetrators in terms of achieving social power.

The existence of a group of popular aggressive youth was initially surprising, given the widely replicated finding that aggressive children are socially rejected (meaning that they are disliked by their peers) or controversial (meaning they are liked by some of their peers but disliked by others (e.g., Coie, Dodge, & Kupersmidt, 1990; Tomada & Schneider, 1997). In the last 10 to 15 years, however, we have come to understand that being liked is only one aspect of social acceptance. For a more complete understanding of a child or adolescent's position within their social context we must also consider popularity, a measure of social

centrality and influence over others which is only marginally related to being liked (Crick, Murray-Close, Marks, & Mohajeri-Nelson, 2009). Popularity can be seen as a measure of social power and dominance as it is associated with possessing resources or skills that others desire (Babad, 2001; Crick et al., 2009). Far from being socially rejected outsiders, youth who are aggressive and popular are leaders in their peer groups (Cillessen & Rose, 2005). In addition, peer-perceived popularity appears to buffer relationally aggressive students from unwanted outcomes like internalizing symptoms (Rose & Swenson, 2009). Given this, popular aggressive youth appear to be a distinct group among aggressive youth.

Now that it is clear that some aggressive youth achieve high-status positions within their social network, research has begun to focus on identifying factors that contribute to the variation seen in outcomes for aggressors. Walcott and colleagues (2008, p. 550) suggest "what separates low- and high-status students in not the presence of aggression per se, but how effectively their displays of aggression achieve their social goals." Several factors have been identified that seem to make the use of aggression more effective. As might be expected, since proactive aggression is goal directed, it appears that proactive aggression is related to less social and internalizing problems than reactive aggression (Card & Little, 2006, 2007; Morrow et al., 2006). Relational forms of aggression also seem to be less damaging to aggressors than other forms of aggression. In one study, overt aggression was associated with poorer adjustment and a socially unskilled behavioural profile; however, relational aggression did not show a similar negative relationship (Putallaz et al., 2007). Relational aggression is also the form of aggression most strongly related to popularity in adolescence (Crick et al., 2009). Use of relational aggression has even been shown to predict later popularity, suggesting that the use of relational aggression may be a strategy for achieving popularity in the future (Rose, Swenson, & Waller, 2004).

Individual characteristics of the aggressor also seem to play a part in whether use of aggression is an effective social strategy. First, it only seems to be effective for older children and adolescents, perhaps because young children have not yet developed sufficiently skilled aggressive behaviour to be effective at manipulating social networks (Crick et al., 2009). Second, it seems that youth who are admired by peers for other reasons have an easier time using aggression effectively, as peer-valued characteristics like attractiveness and athleticism moderate the relationship between the use of aggression and perceived popularity and power (Vaillancourt & Hymel, 2006). Finally, it appears that aggression is most effective for achieving social status when it is only part of an individual's behavioural repertoire (Bukowski & Abecassis, 2007). Hawley (2003) identified a group of youth she called Bistrategic Controllers, who use both pro-social (e.g, co-operative) and anti-social (e.g., aggressive) behavioural strategies in their social behaviour. This group appears to be highly socially skilled and to achieve the highest level of social dominance of all groups within a classroom.

To summarize, it appears that there are two distinct subgroups of aggressive youth. Some youth seem to be successful in using aggression to achieve a high social status. These adolescents likely possess a number of peer-valued characteristics and use proactive, relational aggression in the context of a larger repertoire of behavioural strategies that include pro-social strategies. This group appears to be quite different from aggressive youth characterized by emotion regulation difficulties, social-information processing deficits, and socially unskilled behaviour, in terms of both the context of aggressive behaviour and social status.

How has aggression been understood?

There are four major theoretical perspectives that have informed research on aggression. The earliest was frustration-aggression theory (Dollard, Doob, Miller, Mower, & Sears, 1989), which proposes that aggression is always preceded by frustration, an unpleasant emotion that arises when a person is blocked from achieving a goal. While this theory seemed to explain a large number of everyday occurrences of aggression, it has become apparent that the theory is too broad, as not every episode of frustration leads to aggression, and not every occurrence of aggression is preceded by frustration (Vaughn & Santos, 2007). While frustration can be closely linked to reactive aggression, it is not as clearly linked to more strategic, proactive aggression (Card & Little, 2007). Frustration-aggression theory does allow, however, that aggression has a functional value, even as it labels aggression in general as maladaptive (Rodkin & Wilson, 2007).

Social learning theories of aggressions arose out of the frustration-aggression theory when this theory was revised to suggest that aggression was one of a number of ways of managing frustration that people learned through experience (Bushman & Huesmann, 2010). Research shows that children can be taught to behave aggressively through reinforcement, and that they can also learn to discriminate between situations in which aggression is likely to be a successful means of goal attainment and when it is not (Bushman & Huesmann, 2010). What's more, children can learn these associations through observation and imitation (social learning; Bandura, 1977). According to social-cognitive learning theory, three cognitive components are related to aggressive behaviour: self-efficacy for aggression (confidence in one's ability to enact aggression), outcome expectations for aggression (beliefs that aggression will lead to positive outcomes), and outcome values for aggression (valuing the outcomes that could be obtained through the use of aggression; Card et al.,

2009). In other words, children and adolescents may choose to behave aggressively because they believe it to be an effective method that they are capable of achieving for obtaining desired ends. They believe this because they have learned it by observing others and possibly also through direct reinforcement. Unlike frustration aggression theory, social-cognitive learning theory has been more closely related to instrumental aggression than reactive aggression (Card & Little, 2007) and is congruent with the empirical results showing that some perpetrators of aggression are popular with their peers.

The social information-processing model of aggression (Crick & Dodge, 1994) also focuses on the cognitions of perpetrators of aggression. The model delineates six steps of information processing that can account for aggressive behaviour: encoding of social cues, interpretation of these cues, selecting goals for social encounters, generating or accessing response choices in social situations, choice of response, and behavioural enactment of the response (Card et al., 2009). Often, this model has been interpreted as implying that perpetrators of aggression have a deficit in social information-processing skills (e.g. Sutton, Smith, & Swettenham, 1999), and indeed, it appears that perpetrators of aggression, as a group, do have certain deficits. For example, children who display aggressive behaviour generate fewer response options than their peers and may interpret ambiguous behaviour as aggressive (Crick & Dodge, 1994). Early on, however, Crick and Dodge (1996) differentiated between the social information-processing patterns of perpetrators of reactive aggression and perpetrators of proactive aggression. They found that perpetrators of reactive aggression demonstrated hostile biases in their attributions of peers' intentions in provocation situations (which is known to lead to reactive anger) and that perpetrators of proactive aggression evaluate aggression and its consequences in relatively positive ways. This makes sense because proactive aggression is goal directed and would only be the selected response

choice if it were expected to lead to positive outcomes. For perpetrators of proactive aggression, the characteristic patterns in social information processing are best thought of as different from other children, rather than inherently deficient (Salmivalli et al., 2011). Popular children and adolescents who perpetrate aggression may have positive expectations about the use of aggression as a behavioural strategy because of social learning or direct experience of achieving social power through aggressive means.

A remaining question is why aggression should ever be an effective strategy for achieving social status. This is best understood through social dominance theory. Human beings, like all animals, exist in a world where valued resources are limited. According to social dominance theory, access to resources such as material goods, influence over others, and positive attention from peers, is determined by social status or rank. Social dominance hierarchies, which organize individuals in terms of their social status, emerge in early childhood (Hawley, 1999). Given its importance for survival and reproduction, achieving high social status is a valued goal for many individuals (Rowell, 1974). There are two main strategies for achieving high social status. Pro-social strategies such as co-operation and persuasion allow individuals to access resources through the co-operation of others. Coercive strategies such as aggression allow individuals to access resources through force, without regard for others (Hawley, 1999). As has already been stated, individuals who are able to use both strategies in combination, bi-strategic controllers, appear to be the most successful in achieving high social status (Hawley, 2003). Individuals who only use coercive strategies, or use coercive strategies in an unskilled manner, may find that their behaviour is not well tolerated by the peer group, leading to peer rejection and low status (Vaughn & Santos, 2007). Thus, social dominance theory is able to explain both high- and low-status perpetrators of aggression.

Research on Bullying

What is bullying?

Bullying is defined as a subset of aggressive behaviour that is intentionally harmful, repeated, and characterised by an imbalance of power (Olweus, 2010). As such, all bullying is aggressive, but not all aggression meets the criteria for bullying. In particular, the frequently studied construct peer victimization, defined as the repeated victimisation of one child by peers, would not necessarily meet the criteria for bullying. The characteristics of a power imbalance and the requirement that the repetition of aggressive acts occur between the same aggressor/victim dyad distinguish these concepts. Bullying is also distinct but overlapping with harassment (Cascardi et al., 2014). Harassment occurs when others perceive actions as creating a hostile or threatening atmosphere, regardless of the harasser's intent. Harassment laws protect classes of individuals that have historically been stigmatized and disempowered, meaning that there is implicitly a context of power imbalance in the case of harassment, but, unlike with bullying, specific types of power imbalance have been identified (e.g., race, sex, sexual orientation, disability) and codified into harassment laws in most countries (IBID). Harassment does not need to be repeated. Thus, harassment is distinguished from bullying in terms of the requirement of intent and repetition, and the breadth of the types of power imbalance that may exist.

Dr. Olweus' definition of bullying as aggression that is repeated, intentionally harmful, and occurring in the context of a power imbalance has been widely used and accepted in the field of bullying research since 1983 when he became one of the first researchers to turn academic attention to the subject (Cornell & Bandyopadhyay, 2010; Smith & Brain, 2000). He created this defintion based on his observations of a type of aggression that occurred among school children, in order to differentate this type of

aggression from other types, particularly group aggression against an individual (Smith & Monks, 2008). This form of aggression was singled out because it was thought to characterise a type of aggression which was particualry difficult to escape and was thus especially harmful (Finkelhor et al., 2012). Olweus' defintion was appropriate for the level of knowledge about bullying available at the time, allowed for collection of prevalence data, and essentially began the academic study of the subject. Even Dr. Olweus himself acknowledges, though, that his defintion was created on the basis not of overwhleming evidence for construct validity, but out of necessity caused by the urgency of beginning research on the construct (Olweus, 2013). Nevertheless, the concept of bullying that was identified seems to have a universal resonance, not just within one culture, but internationally (Jimerson et al., 2010a; Smith & Monks, 2008).

The degree to which Olweus' defintion has been accepted and remained relatively stable within the body of academic research on bullying can be seen in the results of a recent Centre for Disease Control and Pervention (CDC) and the United States Department of Education effort to establish a uniform defintion of bullying for use in public health research (Gladden, Vivolo-Kantor, Hamburder, & Lumpkin, 2014). This defintion, arrived at by 12 expert panel members with feedback from external researchers and practitioners, is as follows (bold font theirs):

> Bullying is any **unwanted aggressive behaviour(s)** by another youth or group of youths who are not siblings or current dating partners that involves an **observed or perceived power imbalance and is repeated multiple times or is higly likely to be repeated**. Bullying may inflict harm or distress on the targeted youth including physical, psychological, social, or educational **harm**. (p. 7)

The biggest change between this definition and Olweus' defintion is that intention to harm on the behalf of the perpetrator is not singled out as a specific criterion. The

explanation given for this is that aggressive behaviour is intentionally harmful, and that instead of attempting to assess intention to harm, "intentionality can be captured by assessing the perpetrator's intent to use harmful behaviours against the targeted youth" (Gladden et al., 2014, p. 8). This would seem to imply that the authors think that intention is still important, but that it is simply already captured through the measurement of the intention to perpetrate aggressive behaviour (although it is not clear whether the intention to behave aggressively or the presence of aggressive behaviour is actually to be assessed). Other changes appear to be refinements: the aggressive behaviour has to be unwanted, certain types of relationships are excluded, the power imbalance can be observed by outsiders or perceived by the individuals involved, and repetition can be actual or threatened. On the whole, however, core ideas of Olweus' original defintion are largely retained. This includes the idea that the aggression must be perpetrated repeatedly (or that repetition appears likely) by a single youth or group of youths and that power imbalance is an important contextual factor.

The fact that government organizations like the CDC and the United States Department of Education are interested in establishing a uniform defintion of bullying is not surprising. Public agencies around the world have begun to establish formal policy about how bullying schould be recognised and punished. For example, all Canadian provinces and territories and 49 US states have already enacted anti-bullying legislation or policy (Brank et al., 2012; "Bullying Policy & Legislation," 2014). In this context, problems with the way that bullying is defined and measured have become salient as they result in immediate consequences. For example, the defintion of bullying enacted in policy differs across Canadian provices ("Bullying Policy & Legislation," 2014), and recent comprehensive reports on the contents of American state anti-bullying statues indicate that these statues tend to be formulated around an expanded definition of bullying that includes all forms of peer

aggression and harassment (Sacco, Silbaugh, Corredor, Casey, & Doherty, 2012; Stuart-Cassel, Bell, & Springer, 2011). This is problematic because a non-specific definition of bullying may lead to redundencies where laws already exist regarding specific forms of aggression (e.g., harassment), put undue stress on public resources to investigate all episodes of aggression, discourage reporting or lead to all aggressive incidents being treated the same way by schools, and/or lead to unpredictable consequences for behaviour (e.g. two students may experience consequences ranging from criminal charges to repremands from a teacher for the same behaviour; Cascardi et al., 2014).

Problems with the way that bullying has been defined become salient when the definition is applied in real-world context, but there have been growing concerns about the definition from within the research literature as well. Much of this criticism revolves around a lack of empirical foundation. Finkelhor, Turner and Hamby (2012) point out that, though the definition has been useful for prompting research and has been widely adopted, "the bullying concept is in reality currently only *a hypothesis*. It is a hypothesis that peer episodes characterized by repetition and power imbalance have a special seriousness and commonality that deem them worthy of special attention" (Finkelhor et al., 2012, p. 217). Similarly, Bauman, Underwood and Card (2013) criticise the use of the traditional bullying definition because "there is a danger that definitional features are misperceived as empirical findings" when there has in fact been little effort to evaluate the definition empirically. There has been a lack of empirical evidence particularly in comparing bullying to other forms of aggression in order to support theories about a unique etiology, function, impact, or intervention strategy for bullying behaviour. This, and associated concerns about clarity of the underlying construct, how the construct is operationalized and measured, and whether the definition unjustifiably excludes some serious forms of peer aggression, has caused some researchers to

question the validity and utility of separating out bullying as a form of aggression for separate study (Bauman et al., 2013; Finkelhor et al., 2012; Hellström et al., 2013; Rigby, 2012). This is a rather surprising and serious question for a research area with a thirty-year history and robust interest. However, the seriousness of the question underscores the seriousness of the issue. Without conceptual and definitional clarity on the nature of bullying, all research on the construct of bullying is compromised. Given this, gathering evidence for the validity of the current definition of bullying has been identified as a research priority (Hanish et al., 2013).

Empirical evidence related to the current definition of bullying as a whole.

Research on bullying has rarely explicitly examined the differences between bullying and other forms of aggressive behaviour (Hellström et al., 2013). There are, however, a few exceptions to this rule. Four studies have compared the outcomes associated with bullying and peer victimization. These studies are important because, to the extent that bullying is related to different outcomes than peer victimization, we can be assured that it is a distinct experience.

Hunter, Boyle, and Warden completed the most well known study comparing bullying and peer victimization in 2007. For this study, peer victimization is defined as "a form of peer abuse in which the child is frequently the target of peer aggression (p. 798)." Note that in this case, the repetition of aggression is not necessarily by the same individual. Interestingly, the way Hunter and colleagues measure peer victimization through a list of behaviours is the same as behavioural approaches to the measurement of bullying used in some bullying research. This underscores the difficulties in distinguishing the outcomes associated with peer victimization and bullying – while in this study a clear distinction is made, typically one is not, meaning that many studies presented as being related to bullying

may actually be measuring peer victimization (see section on the measurement of bullying for more details). Hunter and colleagues identify bullying involvement through explicit measurement of all three aspects of aggressive behaviour that define bullying (repetition, intention to harm, and power imbalance, operationalized as the aggressor being "physically stronger, in bigger groups and/or more popular" p. 801). The results of the study support the discriminative validity and conceptual distinctiveness of bullying in a number of ways. First, they show that victimization by peers, even repeated victimization, does not always constitute bullying. Victims of bullying made up only 38.1% of the peer-victimized group. The results also suggest that those who are victims of bullying perceive the victimization differently than those who are victims of peer victimization. Victimization by bullying was associated with feeling more threatened and less in control of the aggressive episode. Victimization by bullying was also associated with different coping methods than peer victimization. While the use of active coping strategies was similar between groups, people who were victimized by bullying used more coping strategies focused on regulating emotional reactions. Finally, victimization by bullying was associated with more concurrent depressive symptoms than peer victimization, suggesting either that the two types of aggressive involvement are associated with different participant outcomes or that existing differences in depressive symptoms predict different types of victimization by classmates. Taken together, these results led the authors to conclude that peer-victimization and bullying are qualitatively different experiences. The results also suggest that bullying is perceived as more difficult to defend against, supporting Dr. Olweus' theory.

The Hunter, Boyle, and Warden (2007) study takes an important step in validating the current definition of bullying. However, sometimes this study is cited as the only piece of empirical evidence in papers arguing that the definition is valid (Smith, del Barrio, &

Tokunaga, 2013) or that the effects of bullying are more detrimental than peer victimization (e.g. Cascardi et al., 2014). No one study alone should ever be taken to have conclusively proven such an important point. The following studies appear to be less well known, but also compare outcomes associated with bullying and peer victimization.

Identifying itself as the first empirical investigation of its kind, an earlier study by Schäfer, Werner and Crick (2002) also examines differences between the concepts of peer victimization and bullying, this time in a German sample. This study measured peer victimization in terms of the frequency that participants were the target of peers' relational and physical aggression and measured bullying using Olweus' BVQ (Olweus, 1996). Like Hunter and colleagues, Schäfer found that that victimization by bullying and peer victimization were related but not identical categories. Their results with respect to the two approaches' ability to provide unique information about participants' functioning, however, are more equivocal. Schäfer and colleagues examined the ability of information about victimization to predict peer acceptance and rejection. They found that bullying was uniquely related to peer rejection only, and only for female participants, suggesting that peer victimization and bullying were not as distinct as expected. Further, the authors point out that for female participants, being bullied was associated with being victimized through relational aggression; thus, the interaction may in part reflect this association. Further research to distinguish the unique effects of peer victimization and bullying is recommended.

Two studies published after Hunter, Boyle and Warden in 2007 (Felix, Sharkey, & Green, 2011; You et al., 2008) have investigated the unique effects of bullying in the context of validating a new measure of bullying called the California Bully Victimization Scale (CBVS). The CBVS is a behavioural measure that asks respondents to report about their experiences of eight forms of victimization, and then to indicate the frequency, intended

harm, and existence of a power imbalance in those experiences. The studies have different foci, but appear to be based on the same data and report similar findings in our area of interest. One important difference is that in the You et al study (2008) respondents are classified as victims of bullying if they report experiencing at least one form of victimization, at least two or three times in the past month, and with at least one form of power imbalance in relation to the main person victimizing them. This operationalization includes consideration of frequency of victimization and power imbalance only. Thus in actuality, the full definitional criteria of bullying are not assessed in this study (as intention to harm and repetition of aggression by the same individual are not required). In the Felix et al. (2011) study, intention to harm is also included in the operationalization of bullying. The reason for this discrepancy between studies is not explained. Both studies report results indicating that, as in the preceding research, it was possible to create separate groups of victims of peer aggression and victims of bullying. The numbers presented in the You et al. study (2008) indicate that 40.8% of their sample reported some form of victimization, and slightly more than half of these participants, 21.1% of the total sample, met their criteria for victimization by bullying. Both studies also report differences in outcomes between victim groups. Victims of bullying had lower scores on self-reported school connectedness, hope, and life satisfaction than the peer victim group. The You et al. (2008) study further reports that despite this difference, both victim groups were also similar in that they scored lower on the measured outcomes than the non-victim group, and that school connectedness was not found to be a significant mediator of the other two outcomes measured, unlike for the non-victim group. One interpretation of these findings is that all people who are victimized by peer aggression share some negative outcomes in these areas, but people who are victimized by bullying are more negatively affected in these areas than other types of people who are

victimized. Felix and colleagues (2012) see these findings as indicating that victims of bullying represent a distinct group who are more at risk for poor psychosocial development at school.

A final study that examines how the definitional characteristics of bullying are related to participant outcomes was conducted by Ybarra, Espelage and Mitchell (2014). In this study, bullying was assessed through an online survey asking how often participants had been bullied or harassed in the past year. The definitional characteristics of said bullying and harassment were then assessed through follow-up questions about repetition and differential power (intention to harm was not assessed). This introduces some ambiguity, as bullying and harassment are conflated, and technically, an experience would not meet the criteria for bullying unless repetition and power imbalance exist. However, the results of the study do indicate that, as with other studies, an identifiable subset of victimization experiences appear to meet the criteria of bullying. Sixty percent of participants that reported being victimized (18% of the total sample) were victimized by bullying. In addition, as with other studies, experiences that met the criteria for bullying appeared to be especially harmful. Repetition and power imbalance were shown to be independently associated with interference in daily functioning, but those who meet both criteria for bullying experienced the highest level of interference.

Overall, the empirical evidence related to whether bullying is associated with different outcomes than other forms of aggression is limited, but it does suggest that outcomes associated with victimization by bullying are more severe than outcomes associated with different forms of aggression. Results are strongest when the behavioural components of the definition of bullying are assessed explicitly. No research to date examines whether there are also different outcomes for people who perpetrate bullying.

Empirical evidence and theoretical considerations related to the individual behavioural characteristics contained within the definition of bullying.

Repetition.

Besides the literature cited above validating the bullying concept as a whole, most of the empirical evidence which suggests that repetition could be an important characteristic of bullying comes from outside of the bullying literature. Peer victimization research has shown that a trajectory of chronic victimisation is the most strongly associated with maladjustment and affective distress, suggesting an additive effect (Biggs et al., 2010; Kochenderfer- Ladd & Wardrop, 2001). Research on depression has also shown that multiple negative life events can have an aggregate effect on depression symptoms (Kessler, 1997). There is also evidence from studies within the bullying literature, from research focused on validating bullying assessment tools, that more negative outcomes are seen in students who have been victimized two or three times in the past month than in those who were victimized less often (Solberg & Olweus, 2003). Although some would argue that evidence for a dosage-response relationship between bullying and adverse outcomes is still lacking (Hanish et al., 2013), it seems probable that repeated victimization by aggression is in general more harmful than a single incident. More empirical evidence is needed.

Two criticisms of including repetition as a criterion for bullying have been made. One is that repetition is really just a proxy for the severity of aggression or victimization and is an indicator of only a quantitative, not qualitative difference (Hanish et al., 2013; Ringrose & Renold, 2010). This would be supported by the fact that in the above-cited research, chronic victimization or repeated negative events are associated with only quantitatively worse, but not qualitatively different, outcomes. As such, some researchers suggest that repetition may be useful in standardized, self-report measures of bullying to avoid over-

estimating the bullying problem by excluding incidents below a certain threshold, but may be less useful practically in addressing specific incidents in schools, where clear definitional boundaries may be less important than the assessment of harm to the victim (Carrera, DePalma, & Lameiras, 2011). In this thought, the first criticism overlaps with the second, which is that in some cases, single incidents of victimization may be severe enough and cause enough harm to the victim that we would not wish to exclude them from the category of bullying. Even Olweus, despite defending the inclusion of repetition in the definition of aggression, has recognized that one single incident of "serious" abuse can occasionally be legitimately defined as bullying (Olweus, 2010).

There are, however, arguments to be made on theoretical grounds that repetition is included in the definition of bullying based not only on practical considerations, but also because it is key to the bullying construct. The first argument would be that repetition, or the threat of repetition, fundamentally changes the experience of victimization by aggressive behaviour. It has been argued that repetition, or the possibility of it, may lead to anticipatory fear, which may in turn increase the impact of the aggression itself (Hanish et al., 2013) in much the same way that the impairment in Panic Disorder (American Psychiatric Association, 2013) is associated both with the experience of panic attacks and the anticipatory fear of experiencing panic attacks. Repetition may also be associated with Olweus' conceptual cornerstone of bullying, which is that it is a form of aggression that is difficult to escape (2013). Repetition of aggression by a single perpetrator could be associated with feelings of helplessness and loss of agency or control in the person victimized (Sercombe & Donnelly, 2013). Both of these feelings have been associated with negative outcomes and could represent the pathway through which bullying becomes particularly harmful. There are also those who argue that bullying is fundamentally a

relationship problem, unique because of the repetition of one individual asserting their power over another through aggression and because of the lessons this teaches children involved about relationships going forward (PREVNet, 2014). While bullying can be shown as part of a developmental trajectory of violence in relationships (Pepler, Craig, Connolly, Yuile, & Mcmaster, 2005), neither of these theories regarding the importance of aggression have to this point been conclusively demonstrated empirically. As such, the underlying construct that is being measured through the repetition characteristic (severity of involvement, anticipatory fear, loss of agency, learning about relationships) is still an open question.

Intention to harm

Intention to harm has the least research support of the three aspects of the bullying definition in terms of its importance in predicting outcomes. This is probably because intentions are rarely measured explicitly (Bukowski & Abecassis, 2007). We may still expect that intention to harm is important, however, since it is the very factor that defines aggression (Anderson & Bushman, 2002). Perceived intentions form the basis for moral judgements about an action (Rule, Nesdale, & McAra, 1974) and events seem to be more distressing when an aggressive intention is perceived (Crick, 1995). People who are victimized by bullying see the intention of the aggressor to harm them as a determining factor in their perception that bullying has occurred (Cuadrado-Gordillo, 2012; Olweus, 2013). As such, there is reason to believe that behaviours intended to inflict harm may be especially impactful on individuals who are victimized.

A counter argument to the inclusion of intention to harm in the definition of bullying would be that some behaviour could inflict harm even when not intended to do so (i.e., jokes, failure to invite someone to something; Carrera et al., 2011; Sercombe and Donnelly, 2013). Therefore, the inclusion of the intention to harm criteria may exclude some people

experiencing distress from study. Another counter argument is that since intentionality is a shared characteristic of bullying and aggression, it is not useful for distinguishing between these constructs and is redundant (Cascardi et al., 2014; Hanish et al., 2013; Hellström et al., 2013). Probably the biggest criticism of the intention to harm criterion is that intent is difficult to determine in someone else and as such, the construct is subjective and difficult to measure (Bauman, 2013; Carrera, DePalma, & Lameiras, 2011; Hellström et al., 2013). In Hunter, Boyle, and Warden's (2007) study, participants were asked directly about the intent of the person who aggressed against them through the question "Do you think the kid(s) [who aggressed against you] were trying to upset you?" Forty-eight percent of participants indicated that they did not know. This would seem to confirm the difficulty of judging intent in others. Bauman, Underwood, and Card (2013) argue that intention to harm is essentially infeasible to measure. They cite the suggestion, made by Smith, del Barrio, and Takuaga, (2013, p. 30) that intention to harm can be inferred if "a reasonable person would judge that the action could be foreseen as likely to cause harm to the intended recipient," and suggest that this would place an unreasonable burden on researchers since they have difficulty imagining who this objective third party that could report on bullying incidence could be. The reasonable person principal has been applied in many legal contexts however (Smith et al., 2013), and one could perhaps accept that peer reporters would qualify as reasonable persons. This would not, however, eliminate subjectivity from the judgement.

A review of the recent literature reveals a trend of intention to harm being de-emphasized as an important characteristic of bullying (e.g. Cascardi et al., 2014; Green, Felix, Sharkey, Furlong, & Kras, 2013; Juvonen & Graham, 2014; e.g. Ybarra et al., 2014; You et al., 2008), just as it was de-emphasised in the recent CDC definition. The reasons for this general trend have not been explicitly stated. It could be because of the above-described

criticisms and difficulties implementing the construct. It could also be because intention to harm has not been the focus of much theoretical discussion about the unique functions and implications of bullying. Either way, without empirical evidence of its importance, intention to harm seems to be at risk for becoming forgotten as a characteristic defining bullying.

Power Imbalance

Power imbalance between the aggressor and the victim has been called the hallmark feature of bullying (Hanish et al., 2013; Ybarra et al., 2014). It also has the most research support of any of the three defining characteristics. The results of studies discussed above comparing outcomes associated with bullying and peer victimization support it as an important feature. In addition, there have been two studies documenting that a power imbalance does seem to exist between bullies and their victims (Rodkin & Berger, 2008; Veenstra et al., 2007). How power difference is operationalized, however, does seem to matter. In a paper presentation, Card and Hodges (2005) presented data indicating that the larger the difference in physical strength between a person who bullies and his or her victim, the more likely the victim was to feel less socially competent over time (in Salmivalli & Peets, 2009). In contrast, Chan (2009), who operationalized power imbalance as age difference, argued that the power imbalance may not be an observed characteristic of bullying on the basis of finding that most bullying was carried out by same-aged peers. This operationalization of power imbalance seems overly simplified.

The multifaceted nature of power imbalance can make this construct difficult to measure. It has been suggested, and would likely be generally accepted, that power can be manifested based on age, physical strength, social status or popularity, socio-economic status, being a member of a dominant group, intelligence, attractiveness, athletic ability, leadership skills, verbal fluency, self-confidence, having friends, or social support (Chan,

2009; Green et al., 2013; PREVNet, 2014; Smith et al., 2013). Which variables are relevant is likely to vary by the context of specific episodes of aggression. This makes power imbalance difficult to capture using standardized questions (Cascardi et al., 2014). To date, power imbalance has often been operationalized in terms of size, strength, and popularity (Finkelhor et al., 2012). Given the above list of potential sources of power, this seems inadequate; however, trying to explicitly capture all potential sources of power imbalance becomes impractical. Even if measurement of all potential sources of power advantages were possible, there would still be the methodological challenges of determining what to do when different indicators of power favour different members of the interaction or of deciding how to determine which situational factors had the most influence in given situations. It is also unclear whether all types of power imbalance would lead to qualitatively similar experiences for victims and aggressors. In Hunter, Boyle and Warden's (2007) study, there were gender differences in the types of power imbalances that were perceived as threatening. Both genders perceived more physically imposing aggressors as particularly threatening, but males also perceived aggressors in a group as more threatening, whereas females perceived more popular aggressors as more threatening. All of this illustrates that the concept of a power imbalance is complex, and it is unclear whether our current ways of measuring and discussing power imbalance in the literature capture all of the subtleties.

Another methodological issue with the power imbalance characteristic is that once an aggressive relationship has begun, it is difficult to know whether the power imbalance predates the start of the aggressive relationship. "Once someone has aggressed against you or beaten you in a fight, you may be legitimately intimidated by them, but that may not mean that a power imbalance existed initially" (Finkelhor et al., 2012, p. 272). Bullying is typically assessed after the fact, so the relative standing of participants in bullying incidents after the

incident may be more salient than their standing prior to the incident. As such, measures of bullying may lead to false positives, with incidents in which there is a pre-existing power imbalance, and incidents that result in a power imbalance being captured.

Some, however, may argue that capturing aggression with both pre-existing and resulting power imbalances is not problematic, as they consider pre-existing power imbalance as a criterion of bullying to be too restrictive. Schafer Werner and Crick (2002) argue that friendships are typically horizontal relationships, but that victimization within friendships has been shown to be harmful. As such, requiring a pre-existing power imbalance may result in excluding from the category of victims of bullying individuals who experience significant negative outcomes.

Having a restrictive definition of bullying that excludes some people who are harmed by aggressive involvement is not necessarily negative, however, as long as the reasons for including some incidents and excluding others are clear. The question is whether the reasons that the pre-existing power imbalance is important have been adequately articulated and empirically supported to justify excluding some people experiencing harm from the bullying category.

Several theoretical arguments have been made for the importance of including power imbalance criterion in the definition of bullying. Olweus (2010) argues that the power imbalance aspect of the bullying definition is important because otherwise the “winners” of aggressive relationships will also be identified as victims when their opponents retaliate against them, introducing unhelpful variability into the victim category. However, given the existence of a large group of youth who both bully others and are victims of bullying themselves (and thus the “losers” in some aggressive episodes), and the fact that most studies do not capture who is bullying whom, we cannot say for sure whether the power imbalance

criterion is an effective manner of capturing this distinction between the "winners" and "losers" of aggressive episodes.

Arguments in support of the inclusion of the power imbalance criterion can also be made on theoretical grounds, in that the experience of being systematically the subject of aggression by a more powerful other is a qualitatively different experience than being aggressed against by a friend, which could be painful for different reasons. Olweus originally separated out bullying as a type of aggression that is difficult to escape. He apparently sees this as so central to the meaning of the power imbalance that, in the definition of bullying which is used with the Olweus BVQ measures of bullying involvement, Olweus has translated the power imbalance concept into the phrase "it is difficult for the student being bullied to defend himself or herself" (Olweus, 1996). More powerful aggressors may be socially central and able to enlist the support of others in their aggression (Cillessen & Rose, 2005), and/or physically stronger than their victim (Smith & Brain, 2000), both which may reduce a victimized person's chances of defending themselves. As such, victimization by a powerful aggressor may lead to feelings of entrapment, helplessness, and hopelessness, which have been associated with the development of depressive symptoms (Gilbert, 1992).

Some see a total loss of agency in the victim of bullying as the key feature for defining bullying behaviour (Sercombe & Donnelly, 2013). Certainly, it is the idea that victims of bullying are unable to defend themselves that is cited as the aspect of bullying that creates a moral imperative for research and intervention. Smith, del Barrio, and Takuaga (2013) argue that it makes bullying a human rights issue. Therefore, the generally accepted hypothesis would be that a pre-existing power imbalance is important for this reason.

There is, however, a lack of empirical evidence documenting that difficulty defending oneself or loss of agency are the mechanisms that makes bullying relationships particularly

distinct or harmful. For this reason, the conflation of power imbalance and difficulty defending oneself have been criticised on the grounds of introducing "squishiness" (Finkelhor et al., 2012, p. 272) into the definition. There are other possible reasons that aggression by a more powerful other could be particularly harmful (e.g., more powerful aggressors able to inflict more harm and/or are more likely to be the "winner" of aggressive interactions). There are also multiple reasons why a victim of aggression may have difficulty defending oneself (e.g., surprise, context; Bauman et al., 2013; Finkelhor et al., 2012). As such, using the concept of power imbalance to measure difficulty defending oneself does seem to introduce unnecessary confusion. Clarification of what is intended to be measured and research on the mechanisms by which bullying causes increased harm would allow for a clearer distinction.

How has bullying been measured?

Measurement has been called the Achilles heel of bullying research (Cornell, Sheras, & Cole, 2006). The way that bullying has been measured has been criticised on multiple grounds. Two major problems with measurement of bullying will be reviewed here. First, there is poor correspondence between bullying measures completed by different reporters. Second, there is little evidence that bullying measures actually capture the construct as defined. Some measurement approaches appear to ignore behavioural criteria and the distinction between bullying and aggression altogether. These problems in the measurement of bullying are highly troubling, as they call the construct validity of bullying into question on the grounds that the underlying construct is not consistently measured. When measurement of a construct is inconsistent, it is hard to compare results across studies, as there is little confidence that the same construct has been assessed in each study even if the term bullying is used (Hanish et al., 2013). As a result, findings from different studies risk

becoming a collection of interesting individual results instead of forming a coherent body of accumulating knowledge (Bauman et al., 2013). The lack of consistency in assessing bullying behaviour has been identified as a primary concern in bullying research (Swearer, Siebecker, Johnsen-Frerichs, & Wang, 2010).

Measuring bullying and the choice of reporter.

Researchers have several choices about whom to ask to report on bullying activity. The two most commonly chosen strategies are self- and peer-report (Cornell & Bandyopadhyay, 2010). In self-report methods participants are asked to report directly on their own experiences of perpetrating and being victimized by bullying. In peer-report, a nominating group (usually a classroom) is asked to name peers who are perpetrators and/or victims of bullying. Researchers may also use adult reporters such as teachers or parents to identify youth who are involved in bullying, but unfortunately these types of reporters may miss many aggressive incidents, especially in older children and adolescents, as they are often subtle or occur in settings with minimal adult supervision (e.g., school playgrounds and bathrooms; Stassen Berger, 2007). Observational techniques such as recording videotape of school playgrounds (Pepler, Craig, & Roberts, 1998) have also been used to assess bullying, but are very labor intensive to carry out.

If there is a distinction to be made between how bullying and other forms of aggression are measured, it would be that bullying has typically been measured through self-reports, while aggression and general victimization have commonly been measured through peer- or multiple-reporters (Branson & Cornell, 2009; Schäfer et al., 2002). Peer-report methodologies are quite common in the United States for collecting information about peer relationships, however, (Rubin, Bukowski, & Laursen, 2009) and are increasingly seen in the bullying research, especially as researchers focus on investigating the potential social causes

and consequences of bullying behaviour (e.g. Reijntjes, Vermande, Olthof, et al., 2013; Rodkin & Berger, 2008). There are arguments to be made in support of both peer- and self-report approaches in terms of their appropriateness for measuring bullying behaviour.

Supporters of using self-report methodology, including, most notably, Dr. Olweus, who has written on the topic several times (Olweus, 2010, 2013; Solberg & Olweus, 2003), argue that the subjective experience of being victimized is of primary importance. "Considering the usually painful and somewhat subjective nature of being bullied, it is natural to maintain that the students themselves, rather than their peers, are likely to be the best informants on such experiences (Olweus, 2010, p. 19)". This argument, however, appears somewhat contradictory with the idea, also supported by Olweus, that bullying can be defined in objective terms (i.e., by the presence of repeated, intentionally harmful aggression in the context of a power imbalance, Bovaird, 2010).

Olweus criticises peer-report methodologies on several grounds (2010, 2013). He argues that bullying may be too subtle or secretive to be observed by peers. While it is true that the individual being victimized is necessarily always present when victimization occurs and peers are not (Scholte, Burk, & Overbeek, 2013), it is likely that peers are aware of the majority of bullying behaviour. Motivation for bullying is related to social positioning (Houghton, Nathan, & Taylor, 2012; Thornberg, 2013), which suggests that bullies would like it to be observed, and it has been documented that peers are present for 85-88% of all bullying episodes (Atlas & Pepler, 1998; Hawkins, Pepler, & Craig, 2001).

Olweus also criticises peer-report methods for limiting the possible pool of people who may be identified as bullies or victims of bullying to one's own grade or classroom, and at times to same-sex participants, when bullying may occur outside of those contexts. Some limitation to the pool of people who may be nominated does indeed seem necessary for

practical reasons and because the peers do not necessarily observe all of the contexts of their classmates' lives. The degree to which this affects the validity of identification of bullying involvement has not been documented.

Finally, Olweus also criticises peer-report measures because of their reliance on extreme group methodology (2013). Often in peer-report research, an individual is identified as a bully or victim of bullying if they receive a number of nominations for such a role that is above a certain cut point (Cornell & Bandyopadhyay, 2010). Olweus (2013) argues that frequency of nominations may not directly translate to a suitable level of seriousness or frequency of problem behaviours. In addition, these cut-off points are often determined statistically (e.g., one standard deviation above the mean for number of nominations) and standardized within classroom. Olweus argues that, in practice, this means that the cut-off for bullying involvement varies by classroom, resulting in different numbers of nominations being required for different participants to be considered bullies or victims of bullies and meaningful differences between classrooms in total levels of bullying being obscured. Other approaches to quantifying peer nominations that do not have these limitations are available, however, such as a two-nomination cut-off (Rodkin & Berger, 2008).

Supporters of peer-report methodology argue that peer reports are more reliable, less subjective, and less prone to errors than self reports of bullying because they are based on multiple reporters (Bouman et al., 2012; Ladd & Kochenderfer- Ladd, 2002). Self reports of victimization could be affected by careless responding, biased relational schema, perceptions and social cognitions in the reporter, distortions caused by individual influences like depression symptoms, or efforts to minimize experiences in order to protect one's self-esteem (Bouman et al., 2012; Cornell & Bandyopadhyay, 2010; Crick & Dodge, 1994; Graham & Juvonen, 1998; Ladd & Kochenderfer- Ladd, 2002; Rosen, Milich, & Harris,

2007; Scholte et al., 2013). In one study, Cross & Newman-Gonchar (2004) showed that the proportion of students who report having been bullied was reduced from 45.7% to 25% when surveys were screened for inconsistent responses and suspect surveys were removed. Self-report methods have also been criticised for not documenting evidence of reliability and validity (Cornell & Bandyopadhyay, 2010; Felix et al., 2011). For example, self reports of aggressive behaviour (but not victimization) have been shown to lack criterion-related validity, whereas peer reports were significantly correlated to observations of aggressive behaviour (Henry, 2006).

With potential sources of error attached to all methods of assessing bullying (Stassen Berger, 2007), researchers continue to disagree about which reporter is the best source of information. Some recommend greater use of peer-report methods (Cornell & Bandyopadhyay, 2010) while others recommend self-report methods (H. J. Thomas, Connor, & Scott, 2014). No consensus gold standard choice between peer- and self-report methods for measuring bullying involvement has yet been identified.

One reason why there has been so much discussion about which method of assessing bullying is preferable, peer- or self-report, is likely that the correlation between these two methods has been documented to be quite low. While some would argue that the seriousness of this problem has been exaggerated (Olweus, 2013), correlations between self- and peer-reported bullying statuses are only in the range of .14 to .42 (Cornell & Bandyopadhyay, 2010). Such low convergence indicates that there are significant discrepancies in perceptions between individuals and their peers about their behaviour. The two methods are not identifying the same people are as being involved in bullying. This lack of convergent validity is troubling.

One relatively positive way of interpreting this lack of correspondence is to consider the idea that self and peer reports are complementary methods that assess different constructs – self-report captures students' self-perceptions of bullying involvement while peer reports captures students' social reputations for bullying involvement (Juvonen & Graham, 2001). The idea that both peer- and self-report methods capture a meaningful aspect of victimization is supported by the fact that, despite their low correspondence, both are associated with negative outcomes. In one study, both types of measure made independent, statistically significant contributions to the prediction of depression, perception of school staff responsiveness to bullying, and grades. Moreover, the correlations between these outcomes and the two measurement methods did not differ (Branson & Cornell, 2009). The idea that peer and self reports capture different, complementary aspects of the bullying experience is supported by the now well-replicated finding that social outcomes are more related to peer reports and internalizing outcomes to self reports (Bouman et al., 2012; Scholte et al., 2013). As a result, multi-informant research strategies are now recommended, with peer- and self-report measures to be included in the same studies (IBID).

In practice, however, multi-informant assessment of bullying is relatively rare. It is also somewhat unclear how it would be implemented in practice. How would disagreement between informants be managed? It does seem that participants identified through both methods have the worst outcomes (Scholte et al., 2013), but does that mean that participants identified by only one method of assessing bullying should not be considered to be involved? In reality, it seems that the choice of reporter is typically guided by the aim and purpose of the research study (Bouman et al., 2012; Felix et al., 2011; Greif & Furlong, 2006). This may be appropriate for capturing the aspect of bullying of most interest in a given study, but it means that most studies are assessing only one component of the bullying experience (self-

perception or social reputation). It also means that comparing results across studies still often means comparing findings from only moderately related groups.

Measurement strategies used to assess bullying.

Despite the lack of consensus about the best informant for measuring bullying involvement, the most widely used measure of bullying involvement continues to be the self-report Olweus Bully/Victim Questionnaire (BVQ; Cornell & Bandyopadhyay, 2010; Olweus, 1996). In this measure, a definition of bullying is presented to participants, who are then asked to anonymously report on whether they have bullied others or been bullied by others "in the past couple of months." Typically, if students report that they have bullied or been victimized two or three times in that time frame, they are identified as bullies, victims of bullying or bully/victims (i.e., students who both victimize others and are victimized themselves; Solberg & Olweus, 2003).

The BVQ is the prototypical example of what has been called the definitional method of measuring bullying involvement. Other examples of this method include the Multidimensional Peer-Victimization scale, (Mynard & Jospeh, 2000) the Bully Survey (Swearer, Turner, Givens, & Pollack, 2008) and the Who Bullies Whom Measure (Rodkin & Berger, 2008) used in this study. In the definitional method, students are presented with a detailed description of what is meant by the term "bullying" and then asked to indicate whether they or peers have been bullied or bully others (Green et al., 2013). The inclusion of a definition of bullying is considered important as it is meant to distinguish acts of bullying from other forms of peer victimization by creating a shared meaning of the term "bullying" between the researcher and the participants (Furlong et al., 2010; Green et al., 2013). The definitional method has been the dominant method of assessing bullying for two decades (Furlong et al., 2010).

A second method of assessing bullying, called the behaviour-based or behavioural method, emerged in the 1990s (Furlong et al., 2010). In this strategy, lists of specific aggressive behaviours are presented to participants who are asked to specify whether they have committed them or experienced them as a victim. Typically, the behaviours listed cover multiple forms of aggression. The behavioural strategy for assessing bullying has also become very popular. Hamburger, Basile, and Vivolo (2011) reviewed 33 existing self-report measures of bullying and peer aggression for the CDC. Only four of these explicitly define bullying (Cascardi et al., 2014).

One reason that the behavioural method of assessing bullying has risen to prominence is that it avoids the stigma or bias associated with the term "bullying" (Furlong et al., 2010). A criticism of definitional methods is that use of the term "bullying" may unintentionally lead to under-reporting due to socially desirable responding and the emotion-provoking nature of labeling oneself as a bully or victim of bullying (Espelage & Swearer, 2003; Furlong et al., 2010). Indeed, prevalence rates of bullying do seem to be lower when measures include the word "bully" or "bullying" (Cook, Williams, Guerra, & Kim, 2010; Kert, Codding, Tryon, & Shiyko, 2010). However, it could be argued that there is another reasonable explanation for why including the word "bullying" may lead to lower prevalence rates. Definitional methods of bullying at least attempt to distinguish it from other forms of aggression. Behaviour-based measures generally measure only the presence of an aggressive act, and perhaps repetition, and neglect other aspects of bullying such as intention to harm and a power differential (Cascardi et al., 2014; Furlong et al., 2010). Lower prevalence rates found with definitional measures of bullying may reflect that the term "bullying" represents to respondents a narrower class of behaviour than that accounted for by behavioural methods.

As such, the adoption of behavioural methods of measuring bullying involvement may have worsened an existing issue in the measurement of bullying – namely, that it is unclear the extent to which measurement of bullying actually captures bullying as defined, and not merely peer aggression (Hellström et al., 2013). Recently, there have been several studies that have attempted to address this problem by using a behavioural approach and asking explicit follow-up questions about definitional criteria. This strategy is not commonplace, but can be seen in the CBVS (Felix et al., 2011) and other work (discussed earlier) that attempt to differentiate bullying from aggression or peer victimization by investigating whether definitional criteria predict participant outcomes (Hunter et al., 2007; Ybarra et al., 2014). I will refer to this measurement strategy as the explicit behavioural strategy, because repetition of aggression, power imbalance, and sometimes intention to harm are explicitly assessed by asking about them directly. There is reason to believe that this measurement strategy improves on existing methods. The word "bullying" and any associated stigma are avoided, and the definitional criteria of bullying are operationalized in a more direct way than either existing measurement strategy. It is from studies using an explicit behavioural measurement strategy that we get the best evidence for the uniqueness of the bullying construct (Felix et al., 2011; Hunter et al., 2007; Ybarra et al., 2014; You et al., 2008). The one study that compared peer aggression to bullying measured through the definitional strategy (specifically the BVQ; Schäfer et al., 2002) found more ambiguous results about whether bullying and peer victimization are associated with different outcomes. Thus, the explicit behavioural strategy is a promising extension of existing measurement approaches.

Assessment of the definitional characteristics of bullying.

Both behavioural and definitional methods of assessing bullying have been criticized on the grounds that there is little evidence that they actually capture the construct as defined (Bauman et al., 2013; Cascardi et al., 2014; Furlong et al., 2010). Traditional behavioural methods of assessing bullying are indistinguishable from methods used for assessing aggression and do not attempt to capture definitional characteristics. When compared with definitional methods of bullying, self-reported rates of victimization by bullying are significantly higher using this method (Kert et al., 2010). Some may argue that this is because of stigma associated with the word "bullying," but an alternative compatible explanation is that bullying involvement is over reported using this method, which really just assesses aggression. Supporting this is the finding from Hunter, Boyle and Warden (2007) that only 31.7% of people who were identified as peer victims using a list of behaviours met criteria for bullying when criteria were assessed. Hunter, Boyle and Warden are unusual in that they distinguish carefully between the peer victimization and bullying constructs. Functionally, the list of aggressive behaviours used to assess peer victimization in their study is identical to behavioural methods of measuring bullying.

Even when bullying is defined, however, it is unclear whether participants take the definitional criteria into account. In one study, only 10 of 19 students who self reported that they were victims of bullying actually meet definitional criteria when assessed through a follow-up interview, and this was after an intervention program in which the definition of bullying had been explicitly taught to participants (Cornell, McDade, & Biasiolli, 2007 as cited in Cornell & Bandyopadhyay, 2010). This suggests that the definitional methods may also result in over-inclusive reporting of bullying.

When we examine correspondence between definitional and explicit behavioural methods of measuring bullying, there is further evidence that definitional methods are not strongly related to the definitional criteria for bullying. Two studies have examined the correspondence between self-reported bullying involvement measured using definitional methods and the CBVS (representing explicit definitional methods). Felix and colleagues (2011) found that when compared to the Swearer Bullying Survey, the CBVS identified roughly the same number of victims of bullying overall, but there was only a 24-29% correspondence (depending on grade) about which specific individuals experienced bullying (κ = .34-.49). Green and colleagues (2013) examined concordance between the CBVS and the BVQ (Olweus, 1996) and again found that correspondence was only "fair" (AUC, a measure of correspondence = 71.3%). This indicates again that somewhat different groups of respondents were identified as victims of bullying using the different measurement methods. In this study, the CBVS identified more participants as victims of bullying than the definitional measure, so it is not clear whether the definitional method overestimates bullying involvement in this instance. It is likely, however, that the CBVS classifications are more strongly related to definitional characteristics than the definitional classifications, given that the definitional criteria form the basis for decision making in explicit behavioural methods. This suggests that definitional bullying identifications are strongly influenced by other factors.

Green and colleagues (2013) examined which components of the aggressive behaviour (as measured by the CBVS) most strongly predicted BVQ responses. Repeated victimization was definitively the most important predictor. The authors speculate that this is because repeatedly being victimized is a main contributor to developing an identity as a victim. The other significant predictors were all different forms of aggression. If

victimization included teasing, ignoring or hitting, participants were more likely to self identify as victims of bullying. Therefore, form of aggression seems to be an important influence over whether victims of aggression self identify as a victim of bullying. No measures of power imbalance emerged as strong predictors of BVQ victim-of-bullying status. This indicates that the BVQ may not be sensitive to capturing the presence of power imbalance in peer victimization, a critical concern given the centrality of power imbalance to the bullying construct.

Researchers have identified two reasons that definitional strategies for measuring bullying may not accurately capture bullying as defined. One is a question of whether child and adolescent reporters can comprehend the complex concept of bullying and apply it appropriately. Bieber (2013) showed that responses to bullying questions are influenced by developmental level and cognitive abilities. Older children (i.e., grades six, seven, and eight) were in general more accurate at differentiating bullying from aggression than younger students (i.e., grades three four, and five), and increased accuracy was related to lower prevalence rates. There is some evidence that a written definition of bullying may be insufficient to be retained in working memory and applied accurately even by adolescent reporters of bullying involvement. Additional elaboration of the definition, as in Baly and Cornell's (2011) study in which participants were provided with an educational video outlining the differences between bullying and other forms of peer victimization as well as a written definition, also resulted in lower prevalence rates.

A second potential reason is that children and adolescents define bullying differently than researchers (Frisén, Holmqvist, & Oscarsson, 2008; Monks & Smith, 2006; Smith et al., 2002; Vaillancourt et al., 2008). Most studies show that adolescents do not simultaneously consider the three defining characteristics of bullying in their own understanding of the term.

Some studies have also shown that the definitional criteria that do seem important vary by role in the bullying interaction (Cuadrado-Gordillo, 2012). Given that the technical definition of bullying seems to be at odds with the usage of the term among reporters (Finkelhor et al., 2012), it seems likely that a reporter's natural understanding of the term, rather than, or in addition to the presented definition, influences responding.

Whatever the reason, what empirical evidence there is suggests that identification of bullying involvement through definitional methods is not strongly related to meeting the behavioural criteria for involvement. The correspondence with explicitly measured definitional criteria is only fair and it is not necessarily the definitional characteristics that are the strongest predictors of definitional identification for bullying involvement. Behavioural methods are also suspected to be flawed in terms of differentiating bullying from other forms of aggression. At the moment, explicit behavioural strategies seem to be the most promising method for assessing bullying that meets the definitional criteria, but these methods are new, relatively unknown, and need more research (Cascardi et al., 2014). Overall, the validity of measurement of bullying continues to be an area of significant concern and no gold standard method for measuring bullying as defined has been established.

Who is affected by bullying and what are its consequences?

What can be said about who is affected by bullying and the consequences of bullying involvement is necessarily affected by the previously presented concerns about the construct validity of bullying. Discriminate validity rests on the ability of a concept to measure or predict something distinct and unique from other concepts, and there are few studies that examine this directly. Bullying and other forms of aggression are often treated as interchangeable in outcomes studies (Bauman et al., 2013; Cascardi et al., 2014). For

example, in Hawker and Boulton's (2000) meta-analysis on the relationship between peer victimization and psychosocial adjustment, studies that measure victimization by bullying are combined with studies on general victimization, and differences that are related to the form of victimization measured are not explored in published paper. In another example, a study linking childhood bullying to adult outcomes (Copeland, Wolke, Angold, & Costello, 2013) defines bullying in a way that would not be distinct from general aggression and measures bullying only in terms of aggressive behaviours. Therefore, it is often difficult to state definitively that research purporting to document outcomes associated with bullying has in fact measured outcomes that are related to bullying specifically. In addition, poor concordance between measures of bullying that use different reporters or measurement strategies suggests that who is actually identified as being involved in bullying varies depending on how bullying is measured. This may lead to significant differences in the population being studied across samples, which could certainly introduce variability into results. Despite these weaknesses, a significant amount of research on the outcomes of bullying involvement has been completed. With an understanding of the research context behind them, these findings are reviewed below.

It appears that bullying is a truly international phenomenon. It has been documented in at least 40 countries (Jimerson, Swearer, & Espelage, 2010b), although the exact meaning of the bullying construct varies slightly by country (Smith & Monks, 2008). Measured prevalence rates of bullying vary widely, depending on the country in which the study is completed, the developmental stage of participants, and the measurement strategy employed, with between 9% and 72% of young people shown to be involved in bullying in some way (IBID; J. C. Cole, Cornell, & Sheras, 2006; Ybarra et al., 2014). Recent Canadian statistics show that 23.3% of boys and 17.0% of girls aged 11, 13 and 15 are directly involved in

bullying in some way (i.e., as a perpetrator/bully, a victim, or both), (Craig et al., 2009), and a nationally representative survey of American students in Grades six through ten showed that 17% of students had been bullied several times in the past semester and that 19% of students had bullied others (Nansel et al., 2001). Clearly, a significant minority of North American youths are directly involved in bullying. Youths who are not directly involved in bullying are still likely exposed to bullying in the role of witness. Studies that used naturalistic methods to observe bullying in the schoolyard found that peers are witnesses in 85-88% of all bullying episodes (Atlas & Pepler, 1998; Hawkins et al., 2001). As such, bullying is understood to be a widespread and frequently observed form of aggression among school-aged youth.

As with other forms of aggression, bullying behaviour can take multiple forms: physical, verbal, relational, and even cyberbullying (i.e., bullying via electronic communication tools; Salmivalli et al., 2011). Gender differences in forms of bullying seem to follow the same pattern as aggressive behaviour in general, with boys using more physical aggression in their bullying and negligible gender differences in the use of relational aggression in bullying (Card et al., 2008). Most research finds that boys bully more than girls, but that both genders are more likely to bully those of the same gender than the opposite gender (Ladd, 2005). Developmental trends are also expected to follow similar patterns to other forms of aggressive behaviour, with physical bullying declining with age but other forms of bullying increasing into adolescence (Stassen Berger, 2007).

As a subtype of aggression, bullying should be related to similar risk factors and outcomes as aggression, and the research clearly shows that it is. Personal and interpersonal characteristics that place children at risk for involvement in bullying have been identified and are similar to risk factors for aggression in general. Students who are submissive, disliked by

classmates, physically weak, and have an anxious/preoccupied attachment style are more likely to be victims of bullying than their peers (Hodges & Perry, 1999; Perry, Hodges, & Egan, 2001; Salmivalli & Peets, 2009; Schwartz et al., 1998). Familial risk factors, such as high levels of parental conflict and an authoritarian parenting style, as well as a tendency to view aggression as an acceptable and effective method of problem solving, differentiate youth who bully from those who do not (Salmivalli & Peets, 2009; Toblin et al., 2005). Research also shows that young people who are victimized by bullying show poor social and academic functioning, psychological distress, and physical ill-health (Glew, Fan, Katon, Rivara, & Kernic, 2005; Hawker & Boulton, 2000; Rigby, 2003; Salmivalli, 2005; Swearer, Espelage, et al., 2010).

Similarly, research has shown that the perpetration of bullying is linked to negative outcomes such as later anti-social behaviour (Rigby, 2003) and depression (Salmon, James, & Smith, 1998). Many early assumptions about the type of outcomes experienced by perpetrators of bullying, however, were based on research about aggression in general. As explained by Salmivalli and Peets (2009, p. 329):

> For a long time, researchers and practitioners regarded bullies as individuals who lack social skills and have low self-esteem, deficiencies in social information processing, low social standing in the peer group, and other adjustment problems. Although many of these factors seem to be associated with aggression in general, or with reactive aggression, there is little empirical support for their being related to bullying specifically.

In fact, contrary to stereotypes, there appears to be more evidence that at least some perpetrators of bullying experience largely positive social outcomes. Children and adolescents who bully may be disliked by classmates, but their behaviour is associated with popularity, dominance, and socially skilled behaviour (de Bruyn, Cillessen, & Wissink, 2009; Peeters, Cillessen, & Scholte, 2009; Pellegrini & Long, 2002; Rodkin & Berger, 2008;

Salmivalli, 2010; Vaillancourt, Hymel, & McDougall, 2003). This would suggest that children who bully belong to the popular aggressor subtype of aggressive children. Volk, Camilleri, Dane and Marini (2012) suggest that these types of bullies, whose aggressive behaviour is goal driven and sensitive to contextual factors, may experience individual benefits associated with their behaviour and suffer few costs for their aggressive behaviour. In support of this argument, Reijntjes, Vermande, Olthof and colleagues (2013) found that most bullies do not demonstrate high levels of internalizing symptoms or suffer problems in the social domain.

There is some evidence, however, of variability in the social status of the bullying group. This evidence comes from the same studies that link bullying with high-status in the first place. Vaillancourt, Hymel and McDougal (2003) divided children who bully into three groups based on the amount of power that they held in their classroom. The largest group was the highly powerful group with 51 out of 102 participants from the bullying group. There were, however, also 10 students who bullied others in the low-power group, and the remaining 41 were averagely popular. As a group, Rodkin and Berger (2008) found that bullies were near average in popularity, with a mean *Z*-score of .07 on perceived popularity. This was even after separating students who bully from students who both bully and are victimized by bullying, who scored much lower on perceived popularity with a mean *Z*-score of -0.71. This amount of variability is noteworthy considering that when young people who bully are discussed, they are often treated as a uniform group. Indeed, some expected similarity between children and adolescents who bully others is the reason we measure the concept in the first place.

There are at least two possible explanations for the variability that has been found in the social status of young people who bully. It may be that there are two distinct groups of

children who bully, some high status and some low status, just as we see in other types of aggression. In this case, we could say that some individuals are more successful at using bullying behaviour to meet status goals than others. Another possibility is that this variability is a result of measurement error. It could be that traditional measurement strategies used to identify bullies actually identify both high-status true bullies and other types of aggressors, who may be low status, as part of the bullying group. This second possibility would be more consistent with social dominance theories of bullying.

How has bullying been theoretically understood?

Unfortunately, despite decades of research, a theoretical understanding of bullying is still underdeveloped (Salmivalli & Peets, 2009). Social information processing and social-cognitive learning theories are helpful in understanding different thought processes which may underlie bullying behaviour, but only to the extent that the theories help to understand the processes behind all aggressive behaviour. Olweus' conceptualization of bullying as a type of aggression that is difficult to escape (Olweus, 2010) remains largely unexplored. The most promising area of theoretical development in bullying research is in the attempts that have been made to conceptualize bullying from a social dominance perspective as a goal-directed form of aggression used to achieve and maintain social status (e.g. Pellegrini & Long, 2002; Salmivalli & Peets, 2009; e.g. Salmivalli, 2010; Veenstra et al., 2007). These attempts, like other areas of bullying research are, at this point, largely driven by expert opinion; however, empirical evidence supporting such a position is beginning to accumulate.

Early support for a social dominance theory of bullying comes from the finding that bullying increases during the transition between primary and secondary schools, a time when social dominance hierarchies are in flux. Moreover, bullying mediates the relationship between dominance in primary school and dominance in secondary school, suggesting that

bullying is one strategy of achieving or maintaining dominance in a new school environment (Pellegrini & Long, 2002). Bullying has also been associated with having strong social status goals, valuing dominance and the use of proactive aggression (Sijtsema, Veenstra, Lindenberg, & Salmivalli, 2009), which supports the idea that perpetrators of bullying could be using aggression to meet their status goals. Evidence is accumulating that bullying can be associated with power, popularity and dominance (de Bruyn et al., 2009; Peeters et al., 2009; Pellegrini & Long, 2002; Rodkin & Berger, 2008; Salmivalli, 2010; Vaillancourt et al., 2003), although variations in the social success of bullies still needs to be accounted for. Variability in the social status of bullies undermines the idea that people who bully are using aggression strategically to achieve social dominance, however, it is possible that measurement and definitional issues in bullying research account for the variability that is seen. More research will be required before the relationship between bullying and social dominance can be fully understood.

Chapter 3. Overview of the Current Study

As stated in the Introduction, the purpose of this study is to empirically explore the construct validity of bullying. Construct validity includes consideration of content, criterion and construct validities and is "an overall evaluative judgement of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and action of the basis of test scores or other modes of assessment (Messick, 1995, p. 741)." In this study, two aspects of construct validity will be examined. For ease of understanding, analyses and discussion of these two aspects of construct validity have been split into two studies. The first study examines the validity of the definition of bullying in terms of evidence that the definition defines a meaningful subset of aggressive behaviours.

The second study examines the validity of common measurement approaches to bullying in terms of their ability to capture bullying as defined.

Study Rationale and hypotheses

Study One

The goal of the first study is to examine the construct validity of the current definition of bullying in terms of whether the construct as defined identifies a meaningful subset of aggressive behaviour. Bullying as defined would be considered to have construct validity if it identified a distinct group and was meaningfully related to functioning. It is also important to determine whether each of the components included in the definition is necessary, or whether some may be redundant or unhelpful. Existing research, while scarce, suggests that being victimized by bullying is associated with different outcomes than victimization by other forms of peer aggression. No research has yet demonstrated this difference associated with the perpetration of bullying compared to the perpetration of other forms of peer aggression. There is also very little research on the importance of each of the three behavioural characteristics (i.e., repetition, intention to harm, and power imbalance) contained within the definition or their combined effect. Though each of the three characteristics in the definition may be important for theoretical reasons, each has also been criticised on theoretical and/or methodological grounds.

This study will determine whether the definitional characteristics of bullying identify patterns of functioning distinct from patterns associated with aggressive behaviour or victimization in general. It will also examine whether each of the definitional characteristics is individually necessary for identifying a distinct group. These goals are embodied in the following four questions that will be addressed in this study.

1. Do the definitional characteristics of bullying predict functioning above and beyond the presence of aggressive behaviour or victimization by peers?
2. Is meeting full definitional criteria for bullying meaningful in that the presence of all three definitional criteria predicts functioning above and beyond the presence of the individual critieria?
3. Are each of the three definitional characteristics of bullying independently significant for predicting participant functioning?
4. What type of functioning is associated with the definitional criteria of bullying?

I will address these questions through a series of sequential regression analyses. My hypotheses are as follows:

1a. Information about the characteristics of the aggression or victimization will predict functioning above and beyond the presence of aggressive behaviour or victimization by peers.

1b. The presence of all three definitional behavioural characteristics of bullying will predict functioning above and beyond the presence of aggression or victimization that has some or none of these characteristics.

1c. Each of the three definitional characteristics of bullying will independently predict participant functioning.

1d. Victimization by bullying will be associated with worse social and individual outcomes than victimization that does not meet the criteria for bullying. The relationship between perpetration of bullying and social and individual outcomes is less clear.

Hypotheses 1a-1c are based on the premise that the definition of bullying has construct validity and that each aspect of the definition (i.e., all three behavioural characteristics included in the definition plus the requirement that all three characteristics occur together) is important. This is a reasonable position to take, despite a surprising lack of specific research on this topic, because the construct has shown predictive ability and is able to be measured in many cultures (Craig et al., 2009; Rigby, 2003; Swearer, Espelage, et al., 2010). Further, the assumption that bullying as defined is a meaningful construct underlies most bullying research. Criticisms of the definition of bullying are more recent and, though important, also lack empirical backing.

Hypothesis 1d is clear with regards to the predicted functioning associated with victimization by bullying. Bullying is conceptualized as an especially harmful form of aggression and has been associated with poor social and academic functioning, psychological distress, and physical ill-health (Glew et al., 2005; Hawker & Boulton, 2000; Rigby, 2003; Salmivalli, 2005; Swearer, Espelage, et al., 2010). Though direct comparisons between the types of functioning associated with peer victimization and victimization by bullying are rare, evidence from these studies also largely supports the idea that victimization by bullying is more harmful (Felix et al., 2011; Hunter et al., 2007; Ybarra et al., 2014; You et al., 2008). With regards to bullying, however, initial assumptions that perpetrators of bullying also suffer adjustment difficulties are being brought into question by findings that show that bullying can be associated with popularity and social dominance (Salmivalli and Peets, 2009). Since no study has yet directly compared the types of functioning associated with perpetrating aggression and bullying, no specific hypothesis with regards to the functioning of perpetrators of bullying will be made. This part of the analysis is more exploratory.

Study Two

As with Study One, the purpose of Study Two is to examine the construct validity of the current definition of bullying. This time, the aspect of construct validity of interest is our ability to measure the construct as defined. Research has shown that there is considerable variation in who is identified as involved in bullying, depending on the reporter and measurement strategy used. This study will examine the relationship between the widely used definitional method of measuring bullying involvement and the behavioural characteristics of bullying as measured through the explicit behavioural method. In the definitional method, participants are presented with a definition of bullying and then asked to identify their classmates who perpetrate and are victimized by bullying. In the explicit behavioural method, participants assess the presence of aggression, and then the definitional criteria are explicitly assessed as participants are asked whether this aggression was repeated, intentionally harmful, and/or occurred in the context of a power imbalance. Though the explicit behavioural method of measuring bullying is not a gold standard method for measuring bullying, in this study the explicit behavioural method will be assumed to assess the behavioural characteristics of bullying more accurately than the definitional method. The definitional method for measuring bullying will be tested against the explicit behavioural method to determine to what extent the definitional method also reflects the definitional characteristics of bullying without assessing them separately.

The assumption that the explicit behavioural methods better capture the definitional characteristics of bullying, though acknowledged as an assumption, is justified in that bullying is defined in behavioural terms and the explicit behavioural method assesses these behaviours explicitly, whereas the definitional method relies on participants' understanding and application of these criteria to a general question. It is also from studies using an explicit

behavioural measurement strategy that we get the best evidence for the uniqueness of the bullying construct, (as discussed in the literature review; (Felix et al., 2011; Hunter et al., 2007; Ybarra et al., 2014; You et al., 2008), suggesting that explicit behavioural methods capture a specific subset of aggressive behaviour. On the other hand, while there is relatively little empirical research on this topic, existing evidence suggests definitional measures may be sensitive to only one definitional criteria of bullying (repetition) while being insensitive to power imbalance and intention to harm (Green et al., 2013).

Three sets of analyses will examine different aspects of the relationship between definitional and explicit behavioural methods of assessing bullying. First, the agreement between explicit behavioural and definitional methods of assessing bullying when they are used categorically will be examined. Next, the extent to which behaviourally measured characteristics of aggression predict definitional bullying status will be tested. Finally, participants who are definitionally identified as being involved in bullying but do not (according to the explicit behavioural method) meet full criteria for bullying involvement will be compared with those who do meet full behavioural criteria to see if the difference between measurement techniques seems to be meaningful in terms of the types of outcomes associated with identification. This study adds to the currently relatively scant existing literature by examining these questions for the first time that this author is aware of in the people who bully as well as those who are victimized by bullying, by examining these questions using a peer report strategy, and by examining how groups identified through different methods differ in terms of social and individual functioning.

Agreement between definitional and explicit behavioural methods of measuring bullying involvement.

This set of analyses will concern the amount of agreement observed between two methods of measuring bullying involvement: the definitional method and the explicit behavioural method. Disagreement between the two methods would indicate that they are not capturing the same construct. My hypothesis regarding this analysis (Hypothesis 2a) is that the explicit behavioural and definitional methods of assessing bullying will be significantly related, but the level of concordance between the two will suggest meaningful differences in what is being captured by each measurement strategy. Based on past research, it is unclear whether the definitional method will either over-estimate bullying involvement as measured through the explicit behavioural method (as in Green et al., 2013) or whether the two methods will simply identify different individuals as involved in bullying (as in Felix et al., 2011).

The explicit behavioural and definitional methods of assessing bullying are expected to be related because the definitional and explicit behavioural methods of assessing bullying used here theoretically assess exactly the same construct using the same sample and similar peer-report methodology. The assumption that the definitional method assesses the behavioural markers that are explicitly assessed with the explicit behavioural method underlies the widespread adoption of the definitional method in the field. This hypothesis is therefore consistent with the practice of bullying research. However, previous research studies that investigated this assumption found concordance rates between explicit behavioural and definitional methods to be between poor and fair (Felix et al., 2011; Green et al., 2013). Both of these studies used self-report strategies and the same explicit behavioural measure for comparison. It is unclear whether concordance rates found in this study (which

uses peer-report and different definitional and explicit behavioural measures) will be in the same range; therefore, no specific hypothesis is made about the overall level of concordance. It does appear likely, however, that there will be some level of disagreement between methods. Therefore, it is predicted that a significant minority of participants will be classified differently across methods.

Predicting definitional bullying status with behaviourally measured definitional characteristics.

This set of analyses will examine whether the observed definitional characteristics of bullying (measured through the explicit behavioural strategy) predict whether an individual will be identified as involved in bullying through the definitional method. The definitional characteristics would be expected to predict definitional bullying status if definitional status reflects the presence of these characteristics. The analyses will seek to answer, in sequence, three questions. These questions are similar to the questions in study one of this paper, but seek to understand whether the definitional characteristics predict definitional bullying status instead of participant functioning. The questions are as follows:

1. Do the definitional characteristics of bullying predict definitional bully or victim of bullying status above and beyond the presence of aggressive behaviour or victimization by peers?
2. Is meeting full definitional criteria for bullying meaningful in that the presence of all three definitional criteria predicts definitional bullying status above and beyond the presence of the individual criteria?
3. Are each of the three definitional characteristics of bullying independently significant for predicting definitional bullying status?

My hypotheses are as follows:

2b. The three behavioural characteristics of aggression or victimization that define bullying, as a group, will predict definitional bullying status above and beyond aggressive involvement or peer victimization.

2c. Each behavioural characteristic of aggression or victimization that defines bullying will uniquely predict definitional bullying status.

2d. The presence of all three behavioural characteristics of aggression or victimization that define bullying will also be uniquely related to definitional bullying status.

2e. Some variance in definitional status will not be accounted for by the definitional characteristics of bullying for both perpetrators and victims of bullying.

Bullying is understood to be a unique form of aggressive behaviour. What makes it unique is operationalized in the definition of bullying. The premise of the definitional method of bullying is that including a definition of bullying in the questions that assess it ensures that this response reflects bullying as defined. The wide adoption of this type of measurement strategy suggests that many researchers in the field share the assumption that this measurement method does indeed capture bullying as defined. Only one study that is known to this author examines this assumption. It suggests that, of the three definitional criteria for bullying, only repetition is predictive of definitional bullying status (Green et al., 2013). Given the limited nature of this empirical evidence and the widespread acceptance of definitional methods of bullying, the first three hypotheses reflect the assumption that definitional criteria will predict definitional bullying status.

Despite this, it is likely that there is some variance in definitional bullying status not accounted for by the definitional characteristics of bullying. We know that adolescents define

bullying behaviour differently than researchers do and their own conceptions of bullying may influence their responses in addition to the definition given (Frisén, et al, 2008; Monks & Smith, 2006; Smith et al, 2002; Vaillancourt et al., 2008). Questions have also been raised about the ability of child and adolescent reporters to comprehend and apply the bullying definition appropriately (Beiber, 2013). Part of the purpose of this research is to attempt to measure how much of the variance in definitional bullying status is accounted for by such unknown factors.

Comparing indicators of social and individual functioning of participants identified as involved in bullying through the definitional and explicit methods.

This set of analyses will examine whether those who meet the explicit behavioural criteria for bullying involvement differ from those who are identified as being involved in bullying using the definitional method but do not meet the behavioural criteria. Differences in terms of social and individual functioning will be examined. This will contribute to our understanding of whether the current definition of bullying is meaningful, since, to the extent that those who meet full criteria differ from those who do not, we can be confident that meeting full criteria for bullying is an important distinguisher.

This set of analysis will consist of a series of MANOVA analyses comparing participants who meet the behavioural criteria for bullying or victimization by bullying (as measured through the explicit behavioural method) against participants who were identified as perpetrating or being victimized by bullying through the definitional method, but do not meet the behavioural criteria. These groups will be compared on a number of measures of social and individual functioning. My hypotheses are as follows:

2f. Participants who meet the behavioural criteria for victimization by bullying will have worse outcomes than participants who are identified as victims of bullying but do not meet the behavioural criteria.

2g. There will be a difference in level of functioning between participants who meet the behavioural criteria for perpetrating bullying and those students who are identified as perpetrating bullying but do not meet behavioural criteria.

My hypotheses are based on the premise that bullying as defined is a valid construct. If so, people who meet the definitional criteria should therefore be a distinct group. In this study, the explicit behavioural method is assumed to be the best indicator of whether definitional criteria are met. Therefore, those who are identified as being involved in bullying by definitional methods but do not meet definitional criteria as captured through the explicit behavioural method are believed to be incorrectly identified as involved in bullying. As such, they should be distinct in terms of functioning from individuals who do meet the full criteria for bullying involvement. For victims of bullying, current thinking is clear that victimization by bullying seems to be especially harmful, so participants who meet the full criteria for victimization by bullying should have worse indicators of functioning than other participants. For perpetrators of bullying, the difference is less clear and thus my hypothesis is more exploratory. It may be that the behavioural features that define bullying are associated with the skilled use of aggression to gain social status. If so, we would expect perpetrators of bullying to have more positive outcomes (at least social outcomes) than other perpetrators of aggressive behaviour. However, it could also be that the behavioural features that define bullying indicate a type of aggression that has particularly negative outcomes for the perpetrator. Our analyses should provide some information on this issue.

Assessment of Bullying through Peer Report in this Study

While peer reports have several acknowledged weaknesses (see discussion in literature review), a peer-report methodology was chosen for this study for several reasons. First, this study is interested in examining the construct validity of the current definition of bullying. Bullying has been defined in objective terms – certain behavioural characteristics are either present or they are not. If aggression occurs repeatedly between two individuals and the aggression is intentionally harmful and occurs in the context of a power imbalance between the perpetrator and the victim, then bullying has occurred. Since this definition is objective, there was a desire to also measure the bullying and aggressive behaviour in as objective a manner as possible. Self report is considered to capture a more subjective experience of bullying involvement than peer report (Olweus, 2013). While some may consider this an advantage, it is not seen that way for this study.

Second, a peer-report methodology was chosen for this study because peer reports of bullying capture individual social reputations for bullying involvement and are more strongly related to social outcomes than self-report measures (Bauman et al., 2013; Juvonen & Graham, 2001; Scholte et al., 2013). The social function of bullying is increasingly seen to be an important part of what makes bullying unique (Salmivalli, 2010) and this study examines the relationship between bullying involvement and several social outcomes. Several non-social outcomes are also included in this study, but it is the social outcomes that are the most interesting for theoretical reasons and that are most unique to this study compared to previous similar work (Felix et al., 2011; Green et al., 2013; Hunter et al., 2007; Schäfer et al., 2002; Ybarra et al., 2014; You et al., 2008).

Third, the use of peer-reports in this study is unique for studies on the construct validity of the definition of bullying (Felix et al., 2011; Green et al., 2013; Hunter et al.,

2007; Schäfer et al., 2002; Ybarra et al., 2014; You et al., 2008). This uniqueness will hopefully allow this study to contribute new insight to the research literature by comparing the construct validity of bullying when measured using peer report to the construct validity of bullying measured using self report.

Fourth, we felt that including both peer and self reports of bullying was not practical for this study. While multi-informant assessment of bullying may capture more aspects of the bullying experience (Cornell & Bandyopadhyay, 2010), it was unclear how multiple reports would be handled. Already this study compares two measurement strategies for assessing bullying and two methods for operationalizing the peer nominations for those methods (see "Methods of operationalizing peer reports" below). Adding another assessment method for comparison was beyond the scope of this project. As such, one assessment method was chosen, and peer report was felt to be the better choice for this project for reasons outlined above.

Chapter 4. Method

Participants

Data for this study were collected from grade six, seven, and eight students across 29 classrooms in 14 schools in an Ontario, Canada school board. The schools were located in both rural areas and small cities.

In total, data were collected about 535 participants. Peer and teacher report procedures meant that data was collected about all students in participating classrooms with permission to participate, even if they were absent on the day of testing. Of the 535 participants, 488 were present on the day of testing and thus were also able to provide self-report data. The percentage of these students who were male (49%) and female (51%) was close to even.

Data was collected in seventh grade classrooms at participating schools. As a result, most of the student participants were in the seventh grade (n = 352). Because of split classes, however, there were also a number of students in the sixth (n = 16) and eighth (n = 120) grades who participated. Grade seven students were chosen as the targets of this research because the relationship between popularity and the use of relational aggression appears to become stronger as youth enter adolescence, and achieving dominance becomes a more important social goal (Rose et al., 2004). The average age of participants was 12.46 years (*SD* = 0.65).

The most commonly reported ethnicity of participants was white, at 84.8% of the sample. The next-largest groups were self-identified native Canadian/First Nations Canadian at 5.8% of the sample, and Other at 4.0%. The remaining 5.4% of the sample were made up of people who identified as, in descending order of commonality, Filipino, Metis, Black, Latin American, South Asian, Chinese, Inuit, and Southeast Asian. This mostly Caucasian sample appeared to be typical of the area in which data collection was completed and is roughly in line with Canadian national data from the 2006 census (Statistics Canada, 2006).

Students and teachers had to be able to read and understand English to participate as the measures used in this study are written in English.

Procedure

Data collection took place in the participants' school classrooms. Ethical approval for this study was granted by the University of Ottawa Office of Research Ethics and Integrity (see Appendix A for a copy of the Ethics Approval Notice). All ethical guidelines were followed in obtaining student and parent permission to participate in the study. Students were compensated for returning study permission forms with a $5 gift certificate to a local store (e.g., Tim Horton's, Subway, movie rental). Data collection

proceeded only in classrooms where at least 60% of students were willing and able to participate. This was to ensure the integrity of the peer report by ensuring a sufficient pool of nominators. Of the classrooms approached to participate in the study, 29 of the 34 completed the data collection.

During data collection, the researcher guided all participants as a class through the peer nomination section of the survey using a standardized script (see Appendix B). Participants were then asked to answer questions about themselves on their own. During the time that students were completing questionnaires classroom teachers were also asked to complete questionnaires about their students. Once the 40 minutes allotted for data collection was over, all surveys were be collected. Five minutes was spent debriefing with participants.

Measures

Bullying involvement

This study measured bullying involvement using two different measurement paradigms. With the definitional method, participants were given a definition of bullying and then asked explicitly who in their class bullies others and who is bullied by others. With the explicit behavioural method, participants were asked about their peers' aggressive behaviour and then asked follow-up questions to determine whether the aggression meets the definitional criteria for bullying. The names of these methods are somewhat misleading, as definitional criteria is assessed directly only in the explicit behavioural method, but were kept so as to map as directly as possible the two major methods of assessing bullying that are seen in the literature. The definitional method gets its name from the fact that participants are presented with the definition of bullying before responding.

The definitional method of identifying bullying involvement.

Definitional bullying involvement was measured through modified versions of Rodkin and Berger's (2008) *Who Bullies Whom?* measure. In the original measure, participants are asked to identify up to three bullies from their class, and also the people that each one of these students is most likely to bully. This measure was chosen for this study because it allows paired nominations (i.e., it allows us to pair students who bully with the students that they victimize). Though it is widely recognised that bullying typically occurs within specific dyads in a classroom (Coie et al, 1999), specific bullying relationships are rarely identified in bullying research. Paired nominations allowed us a fuller understanding of bullying by allowing us to investigate outcomes affected by specific characteristics of bullying relationships.

Using their paired nomination methodology, Rodkin and Berger (2008) were able to obtain prevalence rates for bullying that are within the range expected based on studies using other methodologies, showing that fourth and fifth grade children were capable of naming bullies and victims conjointly. Moreover, bully and victim categories showed strong stability from fall to spring of the school year, and behavioural characteristics of bullies, victims, and bully/victims were typical of those obtained using traditional bullying measures (IBID).

In the original *Who Bullies Whom* measure, participants were not presented with a definition of bullying before answering the question. This is what was modified for the current study. Participants in the current study were given a definition of bullying before being asked to identify bullies. We chose to modify this measure to produce a peer-report questionnaire about bullying that used a definitional method and resulted in paired nomination. No existing measure that met these criteria was found. Most peer report

measures of bullying use a behavioural method and the paired nominations procedure is also quite rare.

The definition presented to teachers and student participants prior to the *Who Bullies Whom* measure is as follows:

> A student is being bullied, or picked on, when another student, or group of students, says nasty or unpleasant things to him or her. It is also bullying when a student is hit, kicked, threatened, locked inside a room, sent nasty notes, when people won't talk to them, and things like that. These things may happen a lot and it is difficult for the student to defend himself or herself. It is also bullying when a student is teased a lot in a nasty way. It is not bullying when two students of about the same strength have the odd (rare) argument.

This definition is based on Olweus (1996) and has been slightly modified to reflect Canadian dialect, as seen in Vaillancourt and colleagues (2008). It is similar to the definition used in the Olweus Bully/Victim Questionnaire, the most widely used instrument of bullying (Furlong et al., 2010). More recently modified versions of this definition exist that more explicitly address the intention aspect of the definition (Olweus, 2010). To capture this aspect of bullying in our definition, we added the text, "Bullying is meant to hurt people physically, emotionally, or through damaging their reputation."

The definition of bullying was presented to participants in written and verbal form. After reading the complete definition, the investigator emphasized the three definitional aspects of bullying further by verbally instructing participants to remember that bullying is meant to hurt people, it may happen a lot, and it is hard to escape because the person doing the bullying is usually more powerful. In this way, the three definitional characteristics of bullying were further emphasized.

The question used to identify bullies on the *Who Bullies Whom* measure is, "Are there some people in your class who bully other people? Who are they, and who do they bully?" This question was printed across the top of the page with two columns below. Each column

contained a list of participating students in the class with a checkbox next to their name. One column was labelled “Who is a bully?” and the other column was labelled “Whom do they bully.” There was also a checkbox at the bottom of the page for indicating that no one on the list was a bully. If participants felt that someone on the list was a bully, they were asked to check that person’s name in the first column and the names of the students that that person bullied in the second column. Only one bully could be nominated per page. Nine pages with the same question were provided to approximate unlimited nominations as closely as possible without the practical difficulty of providing unlimited answer sheets. Since not every student in the classroom (or school) participated in the study, participants were told that if they felt that someone on the list bullied someone that was not on the list, or that someone was bullied by someone not on the list, they could check a name in one column and leave the other one blank, allowing us to assess bullying involvement in a way that was not limited by the sample.

Most participants appeared to be able to use this paired nomination measure correctly to identify bully/victim pairs. Ninety-two percent of the sample (444 out of 484 participants) completed the measure correctly, meaning that they stated either that no bullying occurred in the classroom, that they identified one or more bully/victim pairs, or that they identified bullies or victims without being able to identify the other member of the pair (presumably because that person was not on the list of names they were given). When mistakes were made, it was usually that multiple bullies and victims were identified on one page, making the intentions of the reporter unclear.

Two sets of scores were calculated for the definitional method of identifying bullying involvement: continuous and categorical scores. Continuous scores equalled the number of nominations that participants received for bullying behaviour and victimization by bullying

on the modified *Who Bullies Whom* measure. Bullying scores ranged between 0 and 15, with an average score of 1.23 nominations (*SD* = 2.04). Victimization by bullying scores ranged between 0 and 17 with an average score of 1.82 nominations (*SD* = 2.24). For categorical scores, participants were divided into four definitional bullying status groups: the Bully group, made up of those who were nominated by peers as perpetrating bullying; the Victim group, made up of those who were nominated by peers as being victimized by bullying; the Bully/Victim group, made up of those who were nominated by peers as perpetrating bulling and being victimized by bullying; and the Uninvolved group, made up of those who were not nominated for either bullying behaviour or victimization by bullying. A two-nomination cut-off was used to determine group membership (i.e., if a person was nominated by (at least) two peers to be a victim of bullying, they were placed in the victim of bullying group). See Appendix C for a discussion on the choice of cut-off used. Using the definitional method, 77 (14%) participants were identified as part of the Bully group, 162 (30%) participants were identified as part of the Victim of Bullying group, 66 participants (12%) were identified as part of the Bully/Victim group, and 230 participants (43%) were identified as part of the Uninvolved group.

The Explicit Behavioural Method of identifying bullying involvement.

Our behaviourally-based measures of bullying involvement is a newly developed extension of Rodkin and Berger's (2008) paired nomination procedure, which uses items from Bjorkqvist and Osterman's (1998) Peer Estimated Conflict Behaviour (PECOBE) Scale to measure aggressive behaviour and its characteristics. The PECOBE evaluates students' level of physical, verbal, and relational aggression using peer report. Their questions for identifying different forms of aggression were used, with language slightly modified to be clear to our intended audience. Response options were modified to allow for paired

nominations and, novel to this study, the determination of whether the aggression met the criteria for bullying. Behavioural methods of assessing bullying typically do not assess the definitional criteria for bullying directly (Furlong et al., 2010), but doing so was necessary in order to answer the research questions of the present study.

The questions presented to participants for the purpose of identifying physical, verbal, and relational aggression are as follows:

- "In your classroom, who hits, kicks, trips, shoves, or pushes others or takes others' things by force?"
- "In your classroom, who teases, insults or yells at other kids or calls people mean names?"
- "In your classroom, who gossips, says bad things about others behind their backs, or tries to get others to dislike a person (for example by trying not to let the person join their group)?"

As in the definitional bullying questions, participants have the option of not identifying anyone. If they choose to identify someone, they are to check the name of the perpetrator in the column that says "Who [does the actions specified in the question]?" Participants then check the names of the people who are victimized by this individual in the column labelled "To whom do they do these things?" As before, they are only able to identify one perpetrator per page and are provided with nine pages for nominating perpetrators for each type of aggression. The measure as described to this point allows us to identify participants involved in aggressive relationships.

The part of the measure that is novel to this study is as follows. To determine whether the aggression identified by the PECOBE questions meets the criteria for bullying, three more questions are asked for each perpetrator/victim pairing identified:

- "Did the person intend to hurt or embarrass the other person?"
- "Is the person who did these things more powerful than the person to whom they did them?"
- "Does this type of thing happen repeatedly?"

If the answer to all three of these questions is yes for a given perpetrator/victim pairing, then the observed behaviour meets criteria for bullying.

The measure of behaviourally-based bullying used in this study is novel and somewhat complex. However, there is reason to believe that participants were, in large part, able to use the measure successfully. Past research shows that children are able to successfully identify bully and victim pairs (Rodkin & Berger, 2008) and the PECOBE questions can be used to identify aggressive individuals (Butovskaya, Timentschik, & Burkova, 2007). Also, in asking participants to answer questions about the repetition, intention to harm, and power imbalance characteristics of aggressive behaviours, we are simply asking directly about each aspect of bullying behaviour that reporters are already asked to consider in definitional methods. If participants are expected to be cognitively able to retain and apply the definition of bullying when completing definitional bullying measures, then they should be able to consider each aspect of the definition individually.

It is possible, however, that the specific formatting of the questions as they appeared in this study made our measure difficult to complete. In order to reduce this potential, the explicit behavioural measure always followed the labeled bullying question in our data

collection – that way, participants were already familiar with the paired nomination aspect of the design and were only dealing with one novel aspect of the question format at a time. Response columns were clearly labeled with the relevant question. In addition, extensive verbal instructions on how to complete the measure were given and participants were shown a fictional example of a completed form. These measures seem to have been largely successful in helping to avoid errors, as 410 of 484 (85%) participants completed the measure correctly (i.e., pairs of students involved in bullying relationships could be identified and the definitional questions were answered in such a way that it was possible to tell which pair they referred to if more than one pair was identified on a page). Of the errors made in filling out our behaviourally-based bullying measure, 29 (39%) were errors with the paired nomination part of the procedure (usually identifying more than one bully per page), 21 (28%) were errors with the definitional criteria questions (usually answering them on the incorrect line, or only answering them once when multiple relationships were identified), and 24 (32%) were both types of errors. In total, only 45 out of 484 (9%) participants made errors with the definitional questions, which were the novel part of this measure.

Two sets of scores were calculated for the explicit behavioural method of identifying bullying involvement: continuous and categorical scores. Continuous scores were based on the number of nominations that they received for aggressive behaviour and victimization by aggression on the modified *PECOBE* measure, plus scores on the original characteristics of aggression question developed for this study. Participants were said to have one nomination as an explicit behavioural bully for each reporter that nominated them as displaying all three behavioural characteristics of aggression that define bullying. The same technique was used to calculate behavioural victim status. Explicit behavioural bully scores ranged between 0

and 8, with an average score of 0.66 nominations (*SD* = 1.29). Victimization by bullying scores ranged between 0 and 10 with an average score of 0.86 nominations (*SD* = 1.44).

For categorical scores, participants were divided into four definitional bullying status groups: the Bully group, the Victim group, the Bully/Victim group, and the Uninvolved group, made up of those who were not nominated for either bullying behaviour or victimization for bullying. A two-nomination cut-off was used to determine whether aggression or victimization was present (see Appendix C) and also whether the aggression or victimization met the behavioural criteria for bullying (see Appendix D – the two-nomination cut-offs referred to as the "Once Is Enough" method in this discussion). Using the explicit behavioural method, 68 (13%) participants were identified as part of the Bully group, 92 (17%) participants were identified as part of the Victim of Bullying group, 17 participants (3%) were identified as part of the Bully/Victim group, and 358 participants (67%) were identified as part of the Uninvolved group.

Aggression.

In addition to being used to measure bullying, the behavioural questions described above were also used to assess the presence of aggression of victimization by aggression. For these items only the aggression questions from Bjorkqvist & Osterman's (1998) PECOBE scale and the aggressor and victim identified in the paired response format were used. Continuous aggression scores ranged between 0 and 16, with an average score of 2.42 nominations (*SD* = 2.90). Victimization by aggression scores ranged between 0 and 17 with an average score of 3.63 nominations (*SD* = 2.83). Categorical aggression group scores were also calculated, using a two-nomination cut-off (see Appendix C). Using this method, 42 participants (8%) were identified as part of the Aggressor group, 187 participants (35%) were identified as part of the Victim group, 209 participants (39%) were identified as part of

the Aggressor/Victim group, and 97 participants (18%) were identified as part of the Uninvolved group.

Measures of participant functioning.

A goal of this study is to examine how characteristics of aggressive involvement and the way that bullying is measured are associated with a number of different indicators of functioning in participants. The types of functioning measured in this study fall into two broad categories.

Social functioning.

Data was collected on five outcomes related to social functioning (power, popularity, social comparison, submissive behaviour, and level of security and acceptance). These outcomes particularly focus on social functioning in terms of variables related to social rank or dominance. While acknowledging that there are other important aspects of social functioning, this study focuses on dominance-related variables in the interest of further exploring the link between bullying and social dominance.

Popularity.

In this study, information about popularity was collected through peer report. Participants were asked to identify students in their classroom that they perceive to be popular and unpopular. As is traditional (Cillessen & Rose, 2005), no definition of popularity was given. Participants answered according to their own intuitive understanding of the concept. Past research has shown that popularity measured in this way is associated with social centrality, the possession of peer-valued characteristics, and influence over peers (Adler & Adler, 1995; LaFontana & Cillessen, 2002; Vaillancourt et al., 2003). From a social dominance perspective, popularity is associated with social dominance, as peers are attracted to popular youth's resources (Crick et al., 2009). Popularity is often used as a proxy for

measuring social dominance in youth (e.g. Olthof, Goossens, Vermande, Aleva, & van der Meulen, 2011) and was used as such in this study.

The specific questionnaire item used to measure popularity reads, "In every school, some people are popular and other kids are not so popular. We want to know, who are the most and least popular people in your class?" Under this question two class lists appeared, one labelled "MOST POPULAR" and the other labelled "LEAST POPULAR." Students were asked, through both verbal and written instructions, to identify popular and unpopular students by checking the box next to the names of these students on the correct list. Each participant was allowed unlimited nominations. Similar procedures for assessing popularity are widely used and allow us to assess popularity as a group construct (i.e., reputation) rather than based on personal preference (LaFontana & Cillessen, 2002). Although the number of nominations has traditionally been limited in sociometric research (Newcomb & Bukowski, 1983), Terry (2000) showed that using unlimited nominations is associated with more desirable data characteristics such as a greater range of values obtained, less skewness and kurtosis in the distribution of nominations, and reduced measurement error. Thus, unlimited nominations were the chosen technique for this study.

There are several accepted methods of quantifying sociometric data (Cillessen, 2009). In this study the standardization method was used. Peer-perceived popularity was calculated as the standardized difference between the number of "most popular" votes, standardized within the classroom, and the number of "least popular" votes, standardized within the classroom. This method of quantifying popularity is the most common and easiest to use (IBID, 2009) and it best accounts for the needs of our specific data set. Unlike other methods, it allows for unlimited nominations and controls for nominating pools of varying size (e.g., different numbers of students participating in different classrooms) (IBID, 2009;

Veldman & Sheffield, 2010). Using this method, we calculated a popularity score for each participant along the *z*-distribution. Scores range from -2.4 to 2 with an average of 0 and a standard deviation of 1.

Power.

While popularity can be viewed as a form of social power, we also wished to assess power as a total concept, as it is crucial in the definition of bullying. No definition of power was given; rather, reporters were asked to make decisions based on their own ideas about what makes people powerful. Potential sources of power that have been discussed in terms of the source of power imbalances seen in bullying relationships include age, physical strength, social status or popularity, socio-economic status, being a member of a dominant group, intelligence, attractiveness, athletic ability, leadership skills, verbal fluency, self-confidence, having friends, or social support (Chan, 2009; Felix et al., 2011; Green et al., 2013; PREVNet, 2014; Smith et al., 2013). As such, power is seen as a superordinate category to popularity, with social status being one type of power that may be manifested.

The method for collecting power nomination was modelled on the same sociometric measurement paradigm used to collect popularity measures. It makes use of peer report. As seen in the study by Vaillancourt, Hymel and McDougall (2003), participants were asked, "Who seems to have a lot of power over others?" followed by a list of names where participants were to check the names of those who they felt had a lot of power. We also asked participants, "Who does not seem to have a lot of power over others?" so that a dimensional scale could be created in the same manner as the popularity scale. Peer-perceived power was calculated as the standardized difference between the number of "has a lot of power over others" votes, standardized within the classroom, and the number of "does not have a lot of power over other" votes, standardized within the classroom. Using this

method, we calculated a power score for each participant along the *z*-distribution. Scores ranged from -2.7 to 2.7 with an average of 0 and a standard deviation of 1. As expected, given that popularity is a form of power, power and popularity are highly correlated, $r = .86$, $p < .001$. Despite this high correlation, they are retained as distinct constructs in further analyses in this study because of their conceptual distinctness.

Self-perceived social rank.

Social rank is a measure of one's standing within a social dominance hierarchy. Those with higher social rank have more access to valued resources. Self-perceived social rank has two major dimensions: the cognitive dimension, characterised by feelings of superiority or inferiority (based on social comparisons) and the behavioural dimension, characterised by dominant or submissive behaviour (Gilbert & Allan, 1994). Feelings of inferiority (i.e., consistently negative social comparisons) and high levels of submissive behaviour are related to depressive symptoms and poorer quality peer relationships in adolescents (Puissant, Gauthier, & Van Oirbeek, 2011) and depression and social anxiety in adults (Gilbert, 2000). The cognitive and behavioural dimensions of self-perceived social rank were measured separately.

Social comparison - Participants' perception of how they compare to their fellow classmates was assessed using the Adolescent Social Comparison Scale-Revised (ASCS: Irons & Gilbert, 2005). This scale was adapted from the adult Social Comparison Scale (Allan & Gilbert, 1995) and is intended to measure the extent to which participants compare themselves favourably to their peers. The ASCS contains 10 items and answers are given on a 10-point scale. Participants are asked to compare themselves to their peers (specifically, in this study, their classmates) on bipolar scales that reflect particular areas of comparison. For example, participants are asked "Compared to your classmates, how *confident* do you feel?"

and they respond on a 10-point scale ranging from 1, "Less Confident" to 10 "More Confident." Scores on the ASCS range from 1 to 10, with higher total scores representing a more positive social comparison (i.e., participants feel more popular, attractive and accepted than their classmates). Higher social comparison scores can be seen as an indicator of higher self-perceived social rank and will be discussed as such. In our sample, the average score was 6.51 (*SD* = 1.48) and the scale showed good internal consistency (Chronbach's Alpha = .78). The distribution of scores was roughly normal.

Submissive behaviour - Participants' perception of their level of submissive behaviour with classmates was assessed using the Adolescent Submissive Behaviour Scale (ASCS: Irons & Gilbert, 2005). This scale was adapted from the adult Submissive Behaviour Scale (Allan & Gilbert, 1997). In this 12-item scale, participants rate how likely they would be to behave in a given manner. For example, one question reads "I do things because others are doing them, rather than because I want to." Answers are given on a five-point Likert Scale ranging from 1 "Never" to 5 "Always". Scores range from 1-5, with higher scores indicating more submissive behaviour. In our sample (*M* = 2.57, *SD* = .67), this scale showed good internal consistency (Chronbach's Alpha = .78). Scores were roughly normally distributed.

Sense of social security and acceptance.

Twelve items from the Secure, Non-striving subscale of the Striving to Avoid Inferiority Scale (SAIS; Gilbert et al., 2007) were used to assess how vulnerable individuals feel to loosing or lacking social power. Research has shown that feeling insecure in the social environment can increase vulnerability to depression, anxiety and stress, and focus individuals on a hierarchical view of their social environment (Gilbert, McEwan et al, 2009), so feeling socially secure should be a protective factor. Indeed, it has been shown that the

Secure, Non-striving Subscale is negatively related to measures of self-harm, stress, anxiety and depression (Gilbert et al, 2007; Gilbert et al 2009). Feeling socially secure and less pressured to compete for social status should also theoretically be related to fewer behavioural displays of dominance, so we will be interested to investigate its relationship with bullying behaviour.

The 12-item "Secure, Non-striving" subscale of the SAIS is answered on a five-point Likert scale. Participants are asked to read a number of statements (e.g., "People accept me whether I'm successful or not") and indicate how often they believe each statement to be true on a scale ranging from 1 = Never to 5 = Always. Calculated total scores range from 1-5, with higher scores indicating a stronger sense of acceptance and security in one's social position. The mean score in our sample was 3.77 (*SD* = .85). The scale showed good internal consistency (Chronbach's alpha = .92) and participants' scores were roughly normally distributed, although the range was somewhat restricted at the high end of the distribution.

Individual functioning.

Mood.

Participants' mood is assessed through the Positive and Negative Affect Schedule, Child's version (PANAS-C Laurent et al., 1999). This 27-item measure is an adapted version of the adult Positive and Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988). Both the PANAS and the PANAS-C are designed to measure two related but distinct aspects of emotional activation: positive and negative affect. Negative affect is a broad category of emotional distress that includes moods such as fear, sadness, anger and guilt. Positive affect reflects positive engagement with the environment and moods such as excitement, self-assurance and cheerfulness. Consistent with the tripartite model of anxiety and depression, high levels of negative affect are related to both types of internalizing

disorder symptoms in young people (Boulton, 1999; Lonigan, Phillips, & Hooe, 2003) but low levels of positive affect are only strongly related to depression symptoms. This makes positive affect useful for distinguishing between the two disorders. In adults, high levels of positive affect have been associated with many positive outcomes, including better social relationships and mental health (Lyubomirsky, King, & Diener, 2005).

There is some developing support for the idea that positive and negative affect can predict changes in anxiety and depression symptoms in youth. In one longitudinal study (Lonigan et al., 2003), measures of NA and PA predicted unique variances in anxiety and depression symptom measures. Therefore, in this study, positive and negative affect will be considered to reflect the participant's current emotional functioning, and also as indicators of possible future mental health functioning.

In the PANAS-C, participants are asked to report on how often different feeling words (e.g., sad, happy, calm) have applied to themselves over the past few weeks. Answers are given on a five-point Likert scale ranging from 1 (very slightly or not at all) to 5 (extremely). Previous studies have indicated that the positive and negative affect subscales generated from this survey have good internal consistency (Laurent et al., 1999), and that is consistent with what was found in this study. The 15-item negative affect scale (M = 1.66, SD = .68) had a Chronbach's alpha of .92, and the 12-item positive affect scale (M = 3.47, SD = .77) had a Chronbach's alpha of .88. Scores on the positive affect scale were roughly normally distributed (skewness = -0.55, kurtosis = -0.08), while scores on the negative affect scale were positively skewed and slightly restricted in range (skewness = 1.33, kurtosis = 1.16). Most participants did not express high levels of negative affect.

School functioning.

Participants' school functioning was assessed through assessing their psychological sense of school membership and teacher-reported level of conduct problems and academic performance. These variables have been associated with bullying and poor educational outcomes (Glew et al., 2005; Juvonen et al., 2000). School membership is the extent of personal belonging, respect, and support students feel in school. Levels of school membership may be used to identify students who are at risk for poor school adjustment and dropping out (Goodenow, 1993). School membership is also related to self-concept and ratings of school climate (Hagborg, 1994).

Psychological sense of school membership - Participants' sense of school membership was assessed through the 11-item Psychological Sense of School Membership Questionnaire – Brief (PSSM-B; Hagborg, 1998). On the PSSM-B, participants are asked to rate statements about the school (e.g., I feel a real part of my school) on a five-point scale indicating how true the statements are for themselves. This measure showed good internal consistency (Chronbach's alpha = .87) in this sample. The mean score was 3.71 with a standard deviation of .78. The distribution of scores was roughly normal with some restriction of range on the upper end.

Conduct problems - Participants' conduct problems at school were measured through teacher report using four items from the teacher-report version of the five-item conduct problems subscale of the Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997). The question on bullying was excluded so that the measure could be used to measure conduct problems independent of involvement in bullying. The SDQ is a widely used measure of psychopathology that can be used for clinical assessment or as a research tool (Goodman, 2001). The conduct problems scale measures externalizing behaviour such as lying, stealing,

disobeying adults and losing one's temper. Teachers were asked to indicate whether a statement such as "Often loses temper" was "Not True (1)," "Somewhat True (2)," or "Certainly True (3)" about a specific child. Higher scores reflect more externalizing behaviour. In our sample, the scale (M = 1.22, SD = .34) was internally consistent (Chronbach's alpha = .78) but the distribution of scores was positively skewed and highly peaked at the low end of the scale. Low levels of conduct problems were seen overall.

Academic performance - In a brief measure of academic performance, teachers were asked to evaluate students' academic performance on a seven-point scale ranging from "bottom of class (1)" to "average (4)" to "top of class (7)" as seen in Cole (1990). The distribution of scores centred on an average of 4.57 (SD = 1.66) and was roughly normal.

Beliefs about aggression.

Participants' views about the acceptability of aggression were assessed through the general beliefs subscale of the Normative Beliefs About Aggression Scale (NBAAS; Huesmann & Guerra, 1997). Beliefs about the acceptability of aggression influence the likelihood that a child will perform aggressive actions (Kikas, Peets, Tropp, & Hinn, 2009). Moreover, there appears to be a specific effect, with beliefs about physical aggression relating to levels of physical aggression, and beliefs about relational/indirect aggression relating to levels of relational/indirect aggression (Werner & Nixon, 2005). Social information processing theory suggests that beliefs about the acceptability of aggression influence behaviour because they are cognitive representations of past experiences that influence the processing of new stimuli and generation of responses in social problem solving situations (Huesmann, 1988). Those who have in the past experienced or observed the "successful" use of aggression in problem solving situations may be more likely see, think of, and use aggressive responses in conflictual interactions. It has been argued that

bullies learn to use aggression and power to control and distress others (Houghton et al., 2012). If that is the case, it is hypothesized that bullying would be associated with increased belief in the acceptability of aggression.

The original version of the NBAAS includes eight items that measure beliefs about physical aggression (e.g., In general, it is wrong to hit other people) and verbal aggression (e.g., It is wrong to insult other people). As has been seen in other studies (e.g. Kikas et al., 2009; Werner & Nixon, 2005), I created four additional items that measure beliefs about relational aggression (e.g., It is wrong to try to get other people not to hang out with someone you don't like). This resulted in a 12-item scale that could be further divided to measure beliefs about specific types of aggression. To avoid bias, half of the items are worded as in the examples above, while the other half are worded in the opposite direction, asking about how "OK" specific behaviours are (e.g., it is usually OK to push or shove other people around if you're mad). Participants responded to each belief item on a four-point scale by indicating whether the behavior was *really OK* (1), *sort of OK* (2), *a little OK* (3), or *not at all OK* (4) or *really wrong* (1), *sort of wrong* (2), *a little wrong* (3), or *not wrong at all* (depending on the working of the question). In creating the final scale score, some items were reverse scored so that higher scores indicate higher belief in the acceptability of aggression.

The resulting 12-item scale (M= 1.40, SD = .41) showed good internal consistency (Chronbach's Alpha = .83) in our sample. Reported belief in the acceptability of aggression was quite low, with most responses clustering around the bottom of the scale range. The distribution of scores is positively skewed.

Chapter 5. Results

Data Screening

With guidance from Tabachnick & Fidell (2007) and Norušis (2012), all data were inspected for data entry errors and missing data before further analysis. To correct for data entry errors, the range of data for each variable was examined and corrections were made until only data within the possible range for each variable was present. Means and standard deviations were checked for each variable and found to be plausible. Scale scores were then calculated from individual items as long as the participant completed 80% of the items making up the scale.

Next, the randomness of missing data was examined. First, students who were absent on the day of testing (and thus did not complete self-report measures) were compared to students present on the day of testing. They were compared on the teacher- and peer-report variables, which were collected for all participants. A one-way MANOVA test was completed to see whether students who were present differed from students who were absent in terms of peer-rated power and popularity, and teacher-rated conduct problems and academic performance. The test was not significant, Wilks' Lambda = .98, $F(4, 518) = 2.15$, $p = .07$. Two 2-way contingency tables analyses were then conducted to see whether the two groups differed on the categorical variables of labeled bullying status (bully, victim of bullying, bully/victim, uninvolved) and aggression status (aggressor, victim of aggression, aggressor/victim, uninvolved). Students who were absent did not differ from students who were present in terms of bullying status, Pearson's χ^2 (3, $N = 535$) = 0.45, $p = .93$, Cramér's $V = .038$; however, they did differ in terms of aggression status, Pearson's χ^2 (3, $N = 535$) = 7.815, $p = .05$, Cramér's $V = .121$. Students who were absent were slightly more likely to be peer nominated as uninvolved in aggression, Pearson's χ^2 (1, $N = 535$) = 6.05, $p = .23$,

Cramér's $V = .11$. One possible explanation for this is that the aggression of students who were not present did not come as easily to mind as other students were completing the survey. This effect, if it exists, is small and is not seen with bullying nominations. Overall, participants who were absent on the day of data collection were not found to differ significantly from those who were present.

A missing variable analysis was then run to see whether, among the students who were present in class on test day, any of the data missing was nonrandom. More than 5% of respondents had data missing for three scales, to the point that scale scores could not be calculated. These three scales were the Positive Affect Scale (7.9% missing), the Negative Affect Scale (9.6% missing), and the Secure Non-striving Subscale from the Striving to Avoid Inferiority Scale (5.5% missing). Recall that scale scores were not calculated if more than 20% of the items were not completed. Missing scale scores for all three variables were found to be related to teacher-rated classroom conduct problems (ps = .01, .002, and .02 respectively) and teacher-rated academic performance (ps = .006, .02, and .03 respectively). People with missing scale scores had higher conduct problem scores and lower academic performance scores. No significant relationship was found between missing data and variables of interest: power, popularity, or bullying status (labeled or otherwise). A possible explanation for missing scores of the Secure Non-striving Subscale is that this was the last scale on the questionnaire, so students who did not have time to finish all the questions often missed this test. Anecdotally, it is also of note that many participants (i.e., one or two in each classroom) asked about the meaning of words on the PA and NA scales. Perhaps students who had more difficulty (academic and behavioural) in school had more difficulty with the emotion vocabulary and thus skipped questions on these subtests. Importantly, missing data

was not found to be related to any aggression or popularity questions, so it is not believed to be meaningful in terms of the questions of interest in this study.

Summary and conclusions.

To ensure that data used for all future analyses was valid and error free, all data was inspected for data entry errors and missing data. Scale scores were calculated for each participant who had completed 80% of items making up the scale. Missing data was examined to determine whether it appeared random. Participants who were absent on the day of testing (and thus did not complete self-report measures) did not differ on peer-rated power and popularity or teacher-rated conduct problems and academic performance from those who were present on the day of testing. Three self-report scales (the Positive Affect Scale, the Negative Affect Scale, and the Secure Non-striving Subscale) could not be calculated for more than five percent of participants because of missing data. The suspected reasons for missing data on these scales include difficulty with vocabulary in the affect scales and the placement of the Secure Non-striving Subscale at the end of the project questionnaire. Missing data on these scales was related to higher teacher-rated conduct problem scores and lower teacher-rated academic performance scores. No significant relationship was found between missing data and variables of interest: power, popularity, or bullying status. As such, it is believed that missing data will not unduly affect the validity of results in further analyses.

Study 1 Results

The purpose of Study One is to examine the validity of the current definition of bullying in terms of its utility in identifying youth with unique outcomes. This will add to the empirical literature on whether bullying as currently defined identifies a distinct subtype of peer aggression and victimization. Analyses consist of two series of sequential regression

analyses, one examining whether characteristics of aggression predict participant functioning and the other examining whether characteristics of victimization predict participant functioning. Each series consists of 11 analyses, each with a different measure of functioning as the outcome.

Four questions of interest were addressed in these analyses:

1. Do the definitional characteristics of bullying predict functioning above and beyond the presence of aggressive behaviour or victimization by peers?
2. Is meeting full definitional criteria for bullying meaningful in that the presence of all three definitional criteria predicts functioning above and beyond the presence of the individual criteria?
3. Are each of the three definitional characteristics of bullying independently significant for predicting participant functioning?
4. What type of functioning is associated with the definitional criteria of bullying?

Characteristics related to the perpetration of aggression.

In order to determine whether behaviourally measured characteristics of aggression perpetration are related to social and individual functioning, I conducted a series of sequential regression analyses, one for each outcome measured. Eleven outcomes were measured for each participant: Popularity ($M = 0$, $SD = 1$), Power ($M = 0$, $SD = 1$), Social Comparison ($M = 6.51$, $SD = 1.48$), Submissive Behaviour ($M = 2.57$, $SD = 0.67$), Security and Sense of Acceptance ($M = 3.77$, $SD = 0.86$), Positive Affect ($M = 3.47$, $SD = 0.77$), Negative Affect ($M = 1.67$, $SD = 0.70$), Sense of School Membership ($M = 3.71$, $SD = 0.79$), Positive Beliefs about Aggression ($M = 1.41$, $SD = 0.44$), Conduct Problems ($M = 1.21$, $SD = 0.35$), and Academic Performance ($M = 4.57$, $SD = 1.66$).

Table 1

Model for Sequential Regression Examining the Relationship between the Characteristics of Aggressive Behaviour and Participant Outcomes

Step	Predictor	Description
1	Class size	Number of people in the participant's classroom who participated in the study (i.e., number of potential nominators)
2	Aggression	Number of nominations for aggressive behaviour
3	Intention	Number of nominations for aggressive behaviour being intentionally harmful
	Power Imbalance	Number of nominations for aggressive behaviour occurring in the context of a power imbalance
	Repetition	Number of nominations for aggressive behaviour being repeated
4	All Three Characteristics	Number of nominators that aggression and endorsed all three definitional characteristics measured

In each regression, predictors were entered in four steps, which are outlined in Table One. The first step contained a control variable: classroom size (M = 19.14, SD = 3.79). This is to control for differences in number of nominations for aggression caused by differences in the size of the nominating pool. The second step also contained one variable: the number of times an individual was nominated as being the perpetrator of aggression (M = 2.42, SD = 2.90). This step indicates to what extent aggression alone is related to student functioning. The third step of this model of predictors contained three variables, corresponding to the number of nominations for each of the three characteristics of aggressive behaviour that would indicate that aggression meets the criteria for bullying: aggression is repeated (M = 1.25, SD = 1.97), aggression is intentionally hurtful (M = 1.3, SD = 2.00), and aggression occurs in the context of a power imbalance (M = 1.25, SD = 1.86). This step indicates to what extent the specific characteristics of aggressive behaviour contained within the definition of bullying add to our ability to predict the outcomes of those participants who perpetrate aggression. The fourth and final set of predictors includes one variable: the

number of times that a participant's aggressive behaviour was nominated as having all three characteristics that define bullying (M = 0.86, SD = 1.44). According to the definition, all three characteristics must be present in order for the behaviour to be defined as bullying. Therefore, this step indicates to what extent, if any, the fact that an individual is aggressive in a way that meets the criteria for bullying adds to our ability to predict student outcomes.

An initial examination of the predictors to be included in this model indicated that the predictors related to the perpetration of aggression were highly related. See Table Two for correlations among predictors. These correlations are elevated partially because some participants (n= 159) were not nominated as displaying any aggressive behaviour, and these participants necessarily also received no nominations for other predictors which detail the characteristics of aggressive behaviour. When these participants were removed, however, correlations between the predictor variables were still quite elevated (e.g., the correlations between nominations for aggressive behaviour and repetition, intention to harm, and power imbalance were still significant at the .001 level and only slightly lower when participants with no nominations for aggression were excluded – .88, .85, and .81 respectively). All

Table 2

Correlation among Predictors in Regression Predicting Participant Functioning from Characteristics of Aggression

	Number of Students in Classroom	Aggression	Repeated	Intentionally Hurtful	Power Imbalance
Aggression	.11*	–	–	–	–
Repeated	.05	.89***	–	–	–
Intentionally Hurtful	.09*	.88***	.88***	–	–
Power Imbalance	.04	.84***	.83***	.85***	–
All Three Characteristics	.033	.76***	.81***	.84***	.91***

Note: * $p < .05$ *** $p < .001$

participants were included in the analysis despite the minimal increases in multicollinearity for two reasons. First, because including people who are not aggressive is necessary to determine whether the number of nominations for aggression alone is often enough to predict student outcomes in many cases, and second, because including people who are not aggressive is necessary to also determine whether any of the characteristics of aggression are differently related to participant outcomes other than generic aggression.

To further examine the extent to which our measures of the behavioural characteristics of aggression were related, we examined the endorsement of differential power, repetition, and frequency nomination combinations. Figure 1 illustrates the

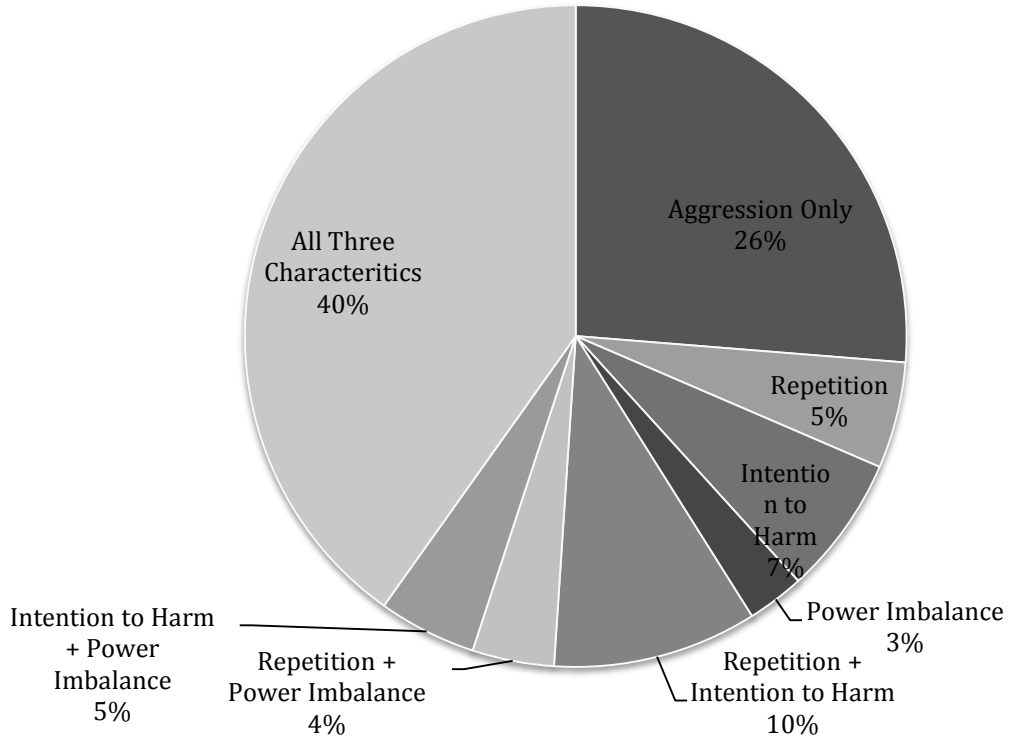

Figure 1. Percentage of participants with labelled combinations of aggression characteristics.

percentage of aggressive participants who met the two-nomination cut-off for each combination of variable characteristics. Roughly one quarter of participants displayed aggression only, one third displayed aggression with some characteristics that did not meet the definition of bullying, and two fifths met the full criteria for bullying. Thus, despite the high correlation, this was considered to be enough differentiation to consider the characteristics variables not completely identical.

The analysis was conducted as planned, despite the high level of correlation among predictors. It was important to examine the importance of all variables included as predictors (rather than eliminating potentially redundant predictors) because all predictor variables are included in the definition of bullying (besides our control variable) and the goal of the analysis is to examine the validity of the definition. In addition, although it is often practiced that low correlations among predictors is necessary when conducting multiple regression, multicollinearity is not a problem for multiple regression as long as information beyond beta weights is considered when judging variable importance (Henson, 2002; Nathans, Oswald, & Nimon, 2012). Multicollinearity is considered problematic because combined prediction of the dependent variable can lead yield regression weights (and thus significance levels) that are poor reflections of the variable relationship (Kraha, Turner, Nimon, Zientek, & Henson, 2012), making interpretation of the importance of predictors difficult. This problem can be avoided by employing one of a number of other methods for determining variable importance besides regression weights. This analysis follows guidelines for determining variable importance laid out by Kraha, Turner, Nimon, Zientek, and Henson (2012) and Nathans, Oswald, and Nimon, (2012). Variable importance was examined through an analysis of standardized coefficients (Beta weights), structure coefficients, all-possible-

subsets regression, commonality analysis, and dominance weights. As such, multicollinearity should not adversely affect the validity of the results presented.

The high level of correlation between predictor variables led us to exclude potential interaction predictors from our analysis. The level of relationship among predictors, particularly the three definitional criteria, suggests that they are actually different ways of measuring one underlying construct. An interaction between two of these, therefore, would not function as an interaction, but be redundantly measuring the same construct. Therefore, there was no justification for examining combined effects. In addition, the interpretation of variable importance becomes exponentially less feasible for each variable included, making the inclusion of interaction terms impractical. It would be functionally impossible to determine variable importance if all possible interactions among aggression predictors were included.

Each of the 11 regressions to be run were examined for possible violations of the assumptions of regression analysis. For each regression, residuals were plotted and examined for problems with linearity and homoscedasticity. No problems that would invalidate the analyses were found. There were some slight issues with homoscedasticity, which may weaken but not invalidate the analyses (Tabachnick & Fidell, 2007, p. 127). Cook's Distance and leverage were also charted and examined. Leverage values within the "risky" range (0.2-0.49; Norušis, 2012) were rare, but were tolerated as long as Cook's Distance values indicated that no case displayed undue influence on the regression coefficients (Cook's Distance > 1; Tabachnick and Fidel, 2007). No data points in any analysis were found to have Leverage > .5 or Cook's Distance > 1. Therefore, all cases were retained for the analysis and data was not truncated in any way.

Outliers in the solution were found in a number of regressions. Two outliers in the

solution were found when predicting social comparison. In each of these cases, the predicted value of the self-perceived social rank was close to average and the actual ASCS score was 2, the minimum value seen in our sample. Two outliers in the solution were found when predicting submissive behaviour. In both of these cases, the predicted values were close to average, but the ACBS score was 5, the maximum value seen in the sample. Three outliers in the solution were also found when predicting sense of acceptance and security. In both of these cases, the predicted values were close to average, but the ACBS score was 1, the minimum value seen in the sample. Two outliers in the solution were also found when predicting sense of positive affect. In both of these cases, the predicted values were close to average, but the PANAS positive affect score was 1, the minimum value seen in the sample. Six outliers in the solution were found when predicting negative affect. In all of these cases, the predicted values were close to average, but the actual score on the PANAS negative affect scale was 4 or 5, towards the top of the range seen in our sample. Two outliers in the solution were also found when predicting sense of school membership. In both of these cases, the predicted values were close to average, but the PANAS positive affect score was 1, the minimum value seen in the sample. Four outliers in the solution were found when predicting negative beliefs about aggression. In each case of these three cases, the predicted value was much lower than the observe value of the NBAAS, which was at the upper end of the range of responses observed. Twelve outliers in the solution were found when predicting teacher-rated conduct problems (not including bullying). In all cases, the predicted value was lower than the observed value. Our model does not fit well for participants who are outliers in the solution, but these participants make up a small percentage of the overall sample. The cases were retained in the analysis to avoid artificially retrofitting the regression equations to our data.

Overall relationship between aggression and its characteristics and participant outcomes.

Results related to this series of regressions are presented in Table 3. The full model (Step 4) was significantly related to all measures of participant functioning except for submissive behaviour. The adjusted R^2 values reported indicate the proportion of the variance in the outcomes that is accounted for by its relationship with all the predictors in our model. By convention, the amount of variance accounted for in all variables is considered small. The strongest relationships were between the behaviourally measured aggressive behaviour and power, popularity, and conduct problems. The results show that more nominations for aggressive behaviour and the definitional characteristics of bullying were related to greater popularity and power, higher self-perceived social rank, lower felt sense of security and acceptance, more negative affect, lower positive affect, lower sense of school membership, more positive beliefs about aggression, more conduct problems, and lower academic performance. In general, these results could be interpreted to indicate that those who perpetrate aggression have better social functioning than those who do not (as measured here, focusing on social power and dominance), but less sense that they are accepted socially for who they are. At the same time, the perpetration of bullying is linked to undesirable individual outcomes. In terms of interpretation, these results are comparable to much of the existing literature in that they do not distinguish the effects of aggression from bullying. This is because the results discussed thus far consider functioning related to our model as a whole. In the next sections, we will consider the contribution of the definitional characteristics of bullying to our ability to predict participant functioning.

Table 3

Model Level Results of Stepwise Regression Predicting Participant Functioning with Characteristics of Aggression

	F	(df)	R^2	ΔR^2
Popularity				
Step 1	0.0000	(1, 533)	.000	
Step 2	6.400*	(2, 532)	.023	.023***
Step 3	18.305***	(5, 529)	.147	.124***
Step 4	15.433***	(6, 528)	.149	.002
Power				
Step 1	0.0000	(1, 533)	.000	
Step 2	42.000***	(2, 532)	.136	.136***
Step 3	40.870***	(5, 529)	.279	.142***
Step 4	34.172***	(6, 528)	.280	.001
Social Comparison				
Step 1	0.0290	(1, 476)	.000	
Step 2	3.909*	(1, 475)	.016	.016*
Step 3	5.909***	(5, 472)	.059	.043***
Step 4	4.960***	(6, 471)	.059	.001
Submissive Behaviour				
Step 1	4.146*	(1, 474)	.009	
Step 2	2.2670	(2, 472)	.009	.001
Step 3	1.9660	(5, 470)	.020	.011
Step 4	1.8820	(6, 469)	.024	.003
Security and Acceptance				
Step 1	0.3590	(1, 462)	.001	
Step 2	1.6760	(2, 461)	.007	.006
Step 3	2.546*	(5, 458)	.027	.020*
Step 4	2.222*	(6, 457)	.028	.001
Positive Affect				
Step 1	0.0840	(1, 450)	.000	
Step 2	0.0420	(2, 449)	.000	0
Step 3	2.44*	(5, 446)	.027	.026*
Step 4	2.15*	(6, 445)	.028	0.002
Negative Affect				
Step 1	0.0460	(1, 442)	.000	
Step 2	2.2550	(2, 441)	.010	.010*
Step 3	4.328***	(5, 438)	.047	.037***
Step 4	4.041***	(6, 437)	.053	.005
	F	(df)	R^2	ΔR^2

Note: Table continued next page. * $p < .05$ *** $p < .001$

Table 3 continued

	F	(df)	R^2	ΔR^2
School Membership				
Step 1	0.5430	(1, 477)	.001	.001
Step 2	5.399*	(2, 476)	.022	.021***
Step 3	5.212***	(5, 473)	.052	.030*
Step 4	4.485***	(6, 472)	.054	.002
Beliefs about Aggression				
Step 1	0.9430	(1, 469)	.002	.002
Step 2	13.483***	(2, 468)	.054	.052***
Step 3	5.646***	(5, 465)	.057	.003
Step 4	5.456***	(6, 464)	.066	.009*
Conduct Problems				
Step 1	3.1650	(1, 526)	.006	
Step 2	56.082***	(2, 525)	.176	.170***
Step 3	30.967***	(5, 522)	.229	.053***
Step 4	25.773***	(6, 521)	.229	.000
Academic Performance				
Step 1	0.6930	(1, 526)	.001	.001
Step 2	8.791***	(2, 525)	.032	.031***
Step 3	5.457***	(5, 522)	.050	.017*
Step 4	4.643***	(6, 521)	.051	.001

Note: * $p < .05$ *** $p < .001$

Do the definitional characteristics of bullying predict functioning above and beyond the presence of aggressive behaviour by peers?

The significance of individual steps in the model tells us which sets of predictors contribute significantly to our ability to predict participant outcomes. First, we can see that the perpetration of aggression alone significantly predicts many measures of social and interpersonal functioning. The perpetration of aggression is related to all outcomes variables except submissive behaviour, sense of social acceptance, and positive and negative affect. This is important because it implies that aggression does not need to meet the definition of bullying to an important indicator of functioning.

Results associated with step three of the stepwise regression allow for the determination of whether the definitional characteristics of bullying, as a group, add to our

ability to predict participant outcomes. The addition of the three measured behavioural characteristics of aggression adds to our ability to predict almost all examined measures of student functioning, with submissive behaviour and beliefs about aggression being the exceptions. This is important because it indicates that aggression that has at least some of the characteristics of bullying is even more strongly related to functioning than generic aggression. Also interesting to note is the fact that for eight out of eleven outcomes, steps three and four of the model explained more variance than step two. This implies that the amount of predictive power gained by adding the definitional criteria is not superficial. Those who are aggressive in ways that demonstrate at least some of the behavioural criteria for bullying are more popular, have more power, and see themselves as more highly socially ranked than those who are aggressive in other ways. At the same time, they feel less secure and less connected to the school, while experiencing less positive and more negative affect and problems with classroom conduct and academic performance. The relationship between aggression and these outcomes and the relationship between definitional characteristics as a group and these outcomes is in the same direction; however, aggression that has some of the behavioural criteria for bullying seems to identify a more extreme group.

Is meeting full definitional criteria for bullying meaningful in that the presence of all three definitional criteria predicts functioning above and beyond the presence of the individual criteria?

When examining the R^2 change accounted for by different steps of our model, we can see that the fourth step in our regression, which includes the all three characteristics variable, does not add significantly to our ability to predict student functioning except in the case of beliefs about aggression. This makes it appear that there is nothing particularly meaningful about aggression having all three of the defining characteristics of bullying. A potential

explanation for this finding, however, is the high correlation between the all three characteristics variable and the characteristics variables entered in the previous step of the analysis. Any variance that may be explained by the presence of all three characteristics might already be accounted for by other variables. In other words, there may be nothing particularly meaningful about the presence of all three characteristics instead of one or two.

Are each of the three definitional characteristics of bullying independently significant for predicting participant functioning?

To examine the relationship between specific characteristics of aggressive behaviour and participant outcomes, statistics on the significance of individual predictors were examined. As stated, a number of indicators of variable importance were examined, as high levels of multicollinearity makes standardized coefficients unreliable. The purpose of each indicator, as described in Kraha, Turner, Nimon, Zientek, and Henson (2012) and Nathans, Oswald, and Nimon, (2012) is described in Appendix E. Statistics related to individual predictors in step four of the model, in which all predictors are included, are presented for all significantly predicted dependent variables in Table 4. Results will be discussed first separately by variable, and then a summary and general conclusions will be made. Because of the relative infrequency with which this type of follow-up to multiple regression is seen, interpretation will be explained in some detail for the first dependent variable (popularity) before being presented more briefly for subsequent analyses. Full commonality analysis and all-possible-subsets regressions related to popularity are presented in Appendix F.

Individual predictors of aggressor's popularity.

Four predictor variables were identified as being significant predictors of participant popularity based on their beta weight: the presence of aggression, intention to harm, power imbalance, and repetition. A comparison across all statistics presented in Table 4 indicates

that power imbalance was the strongest direct predictor across multiple indices. Power imbalance had the largest beta weight, demonstrating that it made the largest contribution to the regression equation, while holding all other predictor variables constant. The structure coefficient and squared structure coefficient suggest that power imbalance was also the most strongly related to the predicted values of popularity. The correlation coefficient and squared correlation coefficient of power imbalance with popularity show that power imbalance shared the largest amount of its variance with popularity. It also had the highest general dominance and relative importance weight score and accounted for the most unique variance at 41.97% of the total variance explained.

The variables with the next two highest (in absolute value) Beta weights are intention to harm and repetition. These beta weights are negative, despite the fact that the structure coefficient and correlation coefficient of both of these variables with the dependent variable is positive, suggesting the possibility of a suppression effect. A suppressor variable is "a variable which increases the predictive validity of another variable (or set of variables) by its inclusion in the regression equation" (Conger, 1974, pp. 36-37 in Mackinnon et al, 2010). It acts by suppressing variance in other independent variables (IV) that is irrelevant to the dependent variable (DV, Tabachnick and Fidel, 2007). In classical suppression, the suppressor variable itself is not highly correlated to the DV and thus its value in the equation stems only from its suppression of error in other variables (Conen, 2003). Because intention to harm and repetition are correlated with our dependent variable in this case, this appears to be an example not of classic suppression, but net suppression, where the beta weight is influenced both by the suppression effect and the predictive ability of the IV on the DV. We can therefore determine from the beta weights of intention and repetition that both are

Table 4

Individual Predictor Results of Stepwise Regression Predicting Participant Functioning with Characteristics of Aggression.

Dependent Variable	Predictor	β	r_s	r_s^2	r	R^2	Unique	Common	General Dominance Weights	Relative Importance Weights
Popularity	# of Nominators	-0.006	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Aggression	0.271*	0.394	0.155	0.152	0.023	0.010	0.013	0.019	0.025
Total R^2 =	Intention	-0.340*	0.202	0.041	0.078	0.006	0.017	-0.011	0.017	0.020
.149	Power Imbalance	0.733*	0.635	0.403	0.245	0.060	0.063	-0.002	0.076	0.059
	Repeated	-0.386*	0.181	0.033	0.070	0.005	0.021	-0.016	0.020	0.020
	All 3 Characteristics	-0.116	0.410	0.168	0.158	0.025	0.002	0.023	0.018	0.025
Power	# of Nominators	-0.030	0.000	0.000	0.000	0.000	0.001	-0.001	0.001	0.001
	Aggression	0.551*	0.694	0.482	0.367	0.135	0.042	0.093	0.072	0.070
Total R^2 =	Intention	-0.385*	0.484	0.234	0.256	0.066	0.022	0.044	0.032	0.033
.280	Power Imbalance	0.739*	0.806	0.650	0.426	0.181	0.064	0.118	0.119	0.097
	Repeated	-0.443*	0.477	0.228	0.252	0.064	0.028	0.036	0.034	0.033
	All 3 Characteristics	-0.087	0.605	0.366	0.320	0.102	0.001	0.101	0.039	0.045
Social	# of Nominators	0.012	0.032	0.001	0.008	0.000	0.000	0.000	0.000	0.000
Comparison	Aggression	0.122	0.508	0.258	0.127	0.016	0.002	0.014	0.007	0.009
	Intention	-0.336*	0.316	0.100	0.079	0.006	0.017	-0.010	0.012	0.009
Total R^2 =	Power Imbalance	0.374*	0.764	0.584	0.191	0.036	0.016	0.020	0.029	0.023
.059	Repeated	-0.062	0.424	0.180	0.106	0.011	0.001	0.011	0.005	0.007
	All 3 Characteristics	0.055	0.628	0.394	0.157	0.025	0.000	0.024	0.012	0.014
Dependent Variable	Predictor	β	r_s	r_s^2	r	R^2	Unique	Common	General Dominance Weights	Relative Importance Weights

Table 4 continued

Dependent Variable	Predictor	β	r_s	r_s^2	r	R^2	Unique	Common	General Dominance Weights	Relative Importance Weights
Security and Acceptance	# of Nominators	-0.016	-0.160	0.026	-0.028	0.001	0.000	0.001	0.000	0.001
	Aggression	-0.005	-0.469	0.220	-0.082	0.007	0.000	0.007	0.003	0.005
	Intention	-0.163	-0.607	0.368	-0.106	0.011	0.004	0.007	0.008	0.008
Total R^2 =	Power Imbalance	0.334*	-0.166	0.028	-0.029	0.001	0.013	-0.012	0.009	0.006
.028	Repeated	-0.141	-0.601	0.361	-0.105	0.011	0.003	0.008	0.007	0.007
	All 3 Characteristics	-0.113	-0.366	0.134	-0.064	0.004	0.002	0.002	0.003	0.004
Positive Affect	# of Nominators	0.033	0.082	0.007	0.014	0.000	0.001	-0.001	0.000	0.000
	Aggression	-0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.004
	Intention	-0.348*	-0.181	0.033	-0.031	0.001	0.018	-0.017	0.012	0.009
Total R^2 =	Power Imbalance	0.212	0.333	0.111	0.057	0.003	0.005	-0.002	0.008	0.007
.028	Repeated	0.100	0.082	0.007	0.014	0.000	0.001	-0.001	0.002	0.004
	All 3 Characteristics	0.090	0.293	0.086	0.050	0.003	0.001	0.001	0.005	0.006
Negative Affect	# of Nominators	-0.006	0.044	0.002	0.010	0.000	0.000	0.000	0.015	0.017
	Aggression	-0.138	0.444	0.197	0.101	0.010	0.003	0.008	0.223	0.189
	Intention	0.365*	0.633	0.401	0.144	0.021	0.020	0.001	0.147	0.126
Total R^2 =	Power Imbalance	-0.157	0.189	0.036	0.043	0.002	0.003	-0.001	0.141	0.121
.053	Repeated	0.218	0.576	0.332	0.131	0.017	0.007	0.011	0.126	0.105
	All 3 Characteristics	-0.190	0.198	0.039	0.045	0.002	0.005	-0.003	0.154	0.138
School Membership	# of Nominators	0.064	0.146	0.021	0.034	0.001	0.004	-0.003	0.064	0.002
	Aggression	-0.117	-0.605	0.366	-0.141	0.020	0.002	0.018	-0.117	0.011
	Intention	-0.325*	-0.669	0.448	-0.156	0.024	0.016	0.009	-0.325	0.021
Total R^2 =	Power Imbalance	0.245	-0.240	0.058	-0.056	0.003	0.007	-0.004	0.245	0.009
.054	Repeated	-0.035	-0.570	0.325	-0.133	0.018	0.000	0.018	-0.035	0.009
	All 3 Characteristics	0.108	-0.240	0.058	-0.056	0.003	0.002	0.002	0.108	0.006
Dependent Variable	Predictor	β	r_s	r_s^2	r	R^2	Unique	Common	General Dominance Weights	Relative Importance Weights

Table 4 continued

Dependent Variable	Predictor	β	r_s	r_s^2	r	R^2	Unique	Common	General Dominance Weights	Relative Importance Weights
Beliefs about Aggression	# of Nominators	0.016	0.176	0.031	0.045	0.002	0.000	0.002	0.015	0.017
	Aggression	0.301*	0.910	0.828	0.232	0.054	0.012	0.041	0.223	0.189
	Intention	0.045	0.863	0.745	0.220	0.048	0.000	0.048	0.147	0.126
	Power Imbalance	-0.181	0.780	0.608	0.199	0.040	0.004	0.036	0.141	0.121
Total R^2 =	Repeated	-0.162	0.776	0.602	0.198	0.039	0.004	0.036	0.126	0.105
.066	All 3 Characteristics	0.246*	0.839	0.704	0.214	0.046	0.008	0.038	0.154	0.138
Conduct Problems	# of Nominators	-0.125*	-0.161	0.026	-0.077	0.006	0.015	-0.009	0.012	0.011
	Aggression	0.058	0.839	0.704	0.401	0.161	0.001	0.160	0.045	0.042
	Intention	0.512*	0.965	0.931	0.461	0.213	0.039	0.174	0.088	0.072
Total R^2 =	Power Imbalance	-0.063	0.798	0.637	0.381	0.145	0.001	0.145	0.036	0.032
.229	Repeated	-0.066	0.839	0.704	0.401	0.161	0.001	0.160	0.043	0.038
	All 3 Characteristics	0.025	0.800	0.640	0.382	0.146	0.000	0.146	0.038	0.034
Academic Performance	# of Nominators	-.010	-.159	0.025	-.036	0.001	0.000	0.001	0.001	0.001
	Aggression	-.020	-.792	0.627	-.179	0.032	0.000	0.032	0.010	0.010
	Intention	-.330*	-.938	0.880	-.212	0.045	0.016	0.029	0.024	0.019
Total R^2 =	Power Imbalance	-.089	-.770	0.593	-.174	0.030	0.001	0.029	0.009	0.009
.051	Repeated	.150	-.695	0.483	-.157	0.025	0.003	0.022	0.008	0.007
	All 3 Characteristics	.094	-.690	0.476	-.156	0.024	0.001	0.023	0.007	0.007
Dependent Variable	Predictor	β	r_s	r_s^2	r	R^2	Unique	Common	General Dominance Weights	Relative Importance Weights

Note: Total R^2= the total amount of variance predicted by the full model. The primary predictor suggested by a method is underlined. β = the standardized correlation coefficient (Beta weight). * = significant at α - .05. r_s = structure coefficient. r_s^2 = squared structure coefficient r = bivariate correlation coefficient. R^2 = squared bivariate correlation coefficient. Unique = proportion of criterion variance explained uniquely by the predictor. Common = proportion of criterion variance explained by the predictor that is also explained by one or more other predictors. Unique + Common = R^2.

important predictors of popularity, but we cannot determine from beta weights alone how much of the effect is explained by the relationship of these predictors to the DV, and how much is explained by their relationship to other predictors. We can see, however, that both repetition and intention to harm are correlated with popularity (although not as strongly as other variables) and do contribute some unique variance to the total variance explained (11.47% and 14.97% respectively). Relative importance weights do not take suppression effects into account and these indicate that both repetition and intention are only slightly less important as individual predictors than the presence of aggression. An examination of the full commonality analysis (see Appendix F) indicates that both variables do indeed have a suppression effect, as demonstrated by the negative coefficients when paired with other

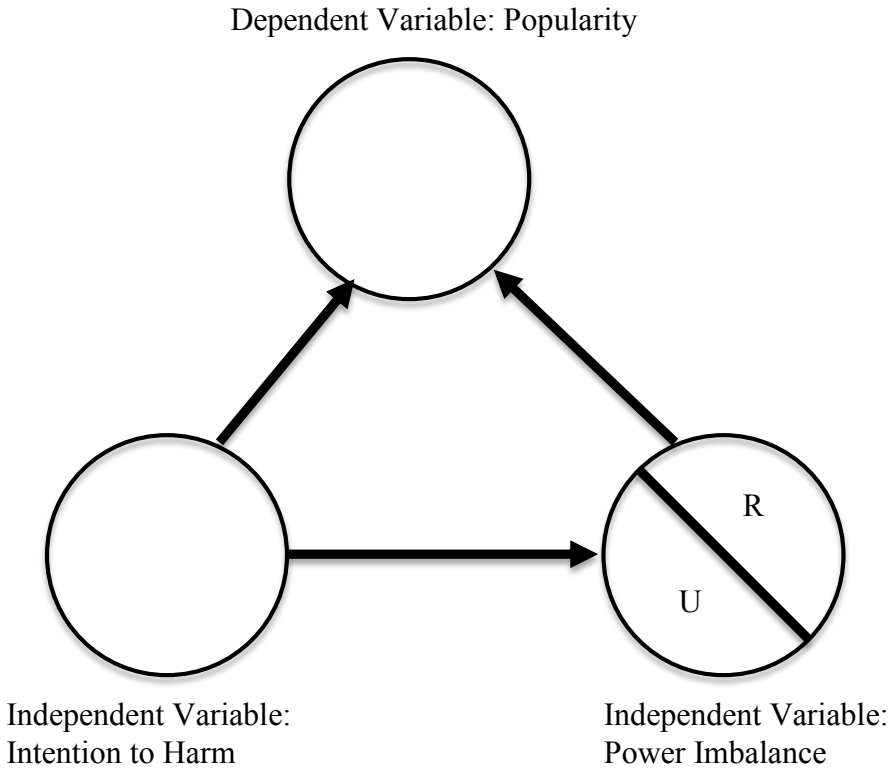

Figure 2. Illustration of direct and suppression effects of intention to harm and power imbalance on popularity. U = Variance in power imbalance that is unrelated to popularity. R = Variance in power imbalance that is related to popularity.

variables. Repetition suppresses variance in the aggression and power imbalance predictors, plus other combinations of variables containing these two variables. Intention to harm acts as a suppressor for the aggression variable and combinations of variables that contain aggression. A final analysis would be that both repetition and intention to harm are important for predicting popularity both because of a direct effect (although the significance of this effect cannot be quantified) and because they help suppress variance in other predictors that is not related to popularity. Their importance as predictors is greater in combination with other variables than it would be on its own. The suppression can be thought of as illustrated in Figure 2 (for both repetition and intention to harm, although only intention to harm is illustrated).

A fourth predictor of popularity that appears to be individually significant is the aggression variable. It has a statistically significant beta weight (and therefore significantly contributes to the regression equation), has a higher structural coefficient than the intention and repetition to harm variables, is correlated with the dependent variable, and has dominance and important weights in line with other variables believed to contribute to the regression equation. What weakens the strength of aggression as a predictor is that it shares a large amount of its relevant variance with the power imbalance variable. Together these two variables contribute 26.19% of the variance, whereas aggression contributes only 6.75% of the prediction of the variance in popularity on its own.

The all three characteristics predictor did not emerge as a significant predictor of participant popularity, accounting for only .002% of the variance on its own. This variable was highly correlated with other variables, particularly the power imbalance predictor. Together, the all three characteristics and the power imbalance predictor accounted for 24.42% of the variance in popularity.

An examination of the importance of individual variables versus shared effects in this analysis can initially be misleading. The sum of the unique variance accounted for by unique variables is equivalent to 75.5% of the total variance in popularity explained. However, common effects explain 120.41% of the variance. The total amount of variance explained sums to almost 200% because suppression effects sum to -95.70% of the variance explained. Given that suppression effects rely on the relationship between predictors, we can conclude from these statistics that the combination of predictors as a group contributes more to our ability to predict student popularity than the individual variables on their own. Despite this conclusion, aggression, intention to harm, repetition, and especially power imbalance also emerge as important individual predictors.

The results of the all-possible-subsets regression analysis support this conclusion. The highest possible amount of variance explained is explained by a regression model that includes all of the predictors, (or all predictors except our control variable, number of possible nominators – the results in terms of R^2 are identical). This model has an R^2 value of .149, which seems meaningfully greater than the most significant individual predictor, power imbalance, which has an R^2 value of .06. However, including five variables in all regressions meant to measure bullying may not be practical, so it is important to note that the power imbalance variable makes a better single predictor than the all three characteristics variable, which may be assumed to be best summary (R^2 value = .025), at least for predicting student popularity. The best two variable combinations would be either power imbalance and repetition or power imbalance and intention to harm, which have almost identical R^2 values and account for almost 80% of the R^2 value of the total model (R^2 value = .118).

Individual predictors of aggressor's power.

The results for individual predictors related to power are very similar to the results for popularity, although more total variance is explained. Power imbalance emerges as the most important predictor across all measures. It accounts for 22.76% of the variance predicted in the outcome on its own. The presence of aggression emerges as the next most important predictor, accounting for 14.87% of the variance in power. However, the amount of common variance that both of these predictors share with other predictors is roughly double their unique contribution. Intention to harm and repetition again emerge in this regression as net suppressor variables, increasing the predictive power of other variables (specifically, per the commonality analysis, aggression and power imbalance plus associated combinations for intention to harm, and aggression for repetition) while still contributing some unique variance to the analysis (7.84% and 9.89% respectively). The all three characteristics predictor is correlated with power and has a relatively high structural coefficient and dominance and importance weights, but contributes almost no unique variance. Most of the variance explained in this regression comes from combined variables (unique effects = 56.04%, Shared Effects = 110.02%, Suppression effects = -66.06%). The all-possible-subsets results are consistent with this finding. The most powerful model for explaining variance in participant power includes all variables (R^2 = .279), although the control variable (number of nominators) and the all three characteristics predictor can be excluded with very little loss of predictive power (R^2 = .278). The most powerful individual predictor, power imbalance, on its own has an R^2 value of .182. The all three characteristics variable, which might be assumed to be the best summary variable, explains less variance (R^2 = .10). The best two variable combinations would be either power imbalance and intention to harm (R^2 = .219) or power imbalance and repetition (R^2 = .215).

Individual predictors of aggressor's social comparison score.

The results for individual predictors related to social comparison or self-perceived social rank are very similar to the results for popularity and power, although less total variance is explained and fewer individual predictors emerge as significant. Power imbalance emerges as the most important predictor by most measures. It accounts for slightly less unique variance than intention to harm (26.07% versus 26.70%) but is overall the strongest predictor. Intention to harm acts as a net suppressor variable, suppressing unrelated variance in the aggression and power imbalance predictors plus other combinations that include these variables. None of the other variables emerge as important individual predictors in this regression. Nevertheless, overall, more variance is accounted for by shared effects than individual predictors (unique effects = 57.72%, shared effects = 103.82%, suppression effects = -61.54%). The all-possible-subsets results are consistent with this finding. The most powerful model for explaining variance in participant power includes all variables (R^2 = .0625) although all of the non-significant predictors can be excluded from the analysis with very little loss of predictive power (R^2 = .0602). The most powerful individual predictor, power imbalance, on its own has an R^2 value of .0365. The all three characteristics variable, which might be assumed to be the best summary variable, explains less variance (R^2 = .0246). The best two variable combinations would be power imbalance and intention to harm (R^2 = .0602).

Individual predictors of aggressor's security and acceptance.

The results for individual predictors related to sense of security and acceptance are somewhat different. Two variables emerge as potentially important individual predictors. Intention to harm, though its beta weight is not significant, is by most other measures the most important individual predictor. It does not predict much actual variance, however, as it

predicts only 12.96% of the total variance explained. In real terms, this is less than half a percent of the overall variance in participants' security and acceptance. Power imbalance is a significant unique predictor, and accounts for more unique variance (42.66% of 1.3% of the variance in participants' security and acceptance). The change in sign between its beta weight and its structural and bivariate correlation coefficients suggest that power imbalance also acts as a net suppressor variable in this case. Indeed, commonality analysis results suggest that it suppresses variance in intention to harm and several larger-order combinations containing this variable. None of the other variables emerge as important individual predictors in this regression. Nevertheless, overall, more variance is accounted for by shared effects than individual predictors (unique effects = 71.11%, shared effects = 96.34%, suppression effects = -83.23%). The all-possible-subsets results are consistent with this finding. The most powerful model for explaining variance in participant power includes all variables or all variables except for aggression ($R^2 = .0305$). The model with the two individual important predictors, intention and power imbalance has an R^2 value of .0241. The most powerful individual predictor in terms of R^2 value is either intention to harm or power imbalance (both $R^2 = .011$). The all three characteristics variable, which might be assumed to be the best summary variable, explains less variance ($R^2 = .0041$).

Individual predictors of aggressor's positive affect.

Two variables emerge as potentially important predictors – intention to harm and power imbalance. This time, although the signs of the two predictors vary, it does not appear to be caused by a suppression effect. Intention to harm is negatively correlated with positive affect, whereas power imbalance is positively correlated with negative affect. We would not expect two of the defining characteristics of bullying to be related to outcomes in an opposite direct, but that appears to happen in this case. However, the overall amount of variance

predicted in positive affect is small and the power imbalance variable accounts for, on its own, less than half a percent in the total variance in positive affect. Intention has a stronger unique relationship, accounting for 61.29% of the variance in positive affect, and is statistically significant where power imbalance is not. No other individual variable emerges as an important predictor. Nevertheless, overall, more variance is accounted for by shared effects than individual predictors (unique effects = 91.71%, Shared Effects = 110.14%, Suppression effects = -101.85%). The all-possible-subsets results are consistent with this finding. The most powerful model for explaining variance in participant power includes all variables (R^2 = .0292). The model with the two important individual important predictors, intention and power imbalance, accounts for almost all of that variance on its own (R^2 = .0252). The most powerful individual predictor, in terms of R^2 value is power imbalance (R^2 = .0032). The all three characteristics variable, which might be assumed to be the best summary variable, explains less variance (R^2 = .0025).

Individual predictors of aggressor's negative affect.

One variable emerges as an important unique predictor of negative affect: intention to harm. It is identified as the most important predictor by most methods and accounts for by far the most unique variance (38.02%). No other individual variables emerge as important. Despite the high general dominance and relative importance weights assigned to aggression, it accounts for very little unique variance and is not a strong predictor of negative affect on its own in the all-possible-subsets regression analysis (R^2 = .0102). More of the variance explained by our model is attributable to shared effects than individual predictors (unique effects = 70.82%, shared Effects = 111.23%, suppression effects = -82.05%). The all-possible-subsets results are consistent with this finding. The most powerful model for explaining variance in participant power includes all variables or all variables except for our

control variable, number of nominators (R^2 = .0518). The most powerful individual predictor in terms of R^2 value is intention to harm (R^2 = .0207). The all three characteristics variable, which might be assumed to be the best summary variable, explains less variance (R^2 = .0020). The best two variable combinations would be power imbalance and intention to harm (R^2 = .0424).

Individual predictors of aggressor's sense of school membership.

The pattern of results for school membership is very similar to the pattern for negative affect. One variable emerges as an important, unique predictor of sense of school membership: intention to harm. It is identified as the most important predictor by most methods and accounts for by far the most unique variance (28.08%). No other individual variable emerges as an important predictor. More of the variance explained by our model is attributable to shared effects than individual predictors (unique effects = 55.53%, shared effects = 103.83%, suppression effects = -59.36%). The all-possible-subsets results are consistent with this finding. The most powerful model for explaining variance in participant power includes all variables (R^2 = .0544). The most powerful individual predictor in terms of R^2 value is intention to harm (R^2 = .0243). The all three characteristics variable, which might be assumed to be the best summary variable, explains less variance (R^2 = .0031). The best two variable combinations would be power imbalance and intention to harm (R^2 = .0444).

Individual predictors of aggressor's beliefs about aggression.

The presence of aggression itself emerged as the most important individual predictor of conduct problems. It is identified as the most important predictor by all methods and accounts for the most unique variance (19.10%). The all three characteristics variable also emerged as significant on this occasion, accounting for 12.26% of the variance accounted for. No other individual variable emerges as an important predictor. More of the variance

explained by our model is attributable to shared effects than individual predictors (unique effects = 43.82%, shared effects = 89.62%, suppression effects = -33.44%). The all-possible-subsets results are consistent with this finding. The most powerful model for explaining variance in participant power includes all variables (R^2 = .0650). The most powerful individual predictor in terms of R^2 value is aggression (R^2 = .0538). In this case, there is only a small difference between the full model and the single variable in terms of predictive ability. The all three characteristics variable, which might be assumed to be the best summary variable, explains less variance (R^2 = .0458).

Individual predictors of aggressor's conduct problems.

The pattern of results for conduct problems is very similar to the pattern for negative affect and sense of school membership. Intention to harm emerges as the most important individual predictor of conduct problems. It is identified as the most important predictor by all methods and accounts for by far the most unique variance (17.03%). Our control variable also emerges as a significant predictor. Since conduct problems are teacher rated and the number of nominators varies by classroom, this is likely an artefact of rater or classroom differences. Number of nominators accounts for 6.67 percent of the variance explained. No other individual variable emerges as an important predictor. More of the variance explained by our model is attributable to shared effects than individual predictors (unique effects = 24.41%, shared effects = 83.45%, suppression effects = -7.86%). The all-possible-subsets results are consistent with this finding. The most powerful model for explaining variance in participant power includes all variables (R^2 = .2282). The most powerful individual predictor in terms of R^2 value is intention to harm (R^2 = .2125). In this case, there is very little difference between the full model and the single variable in terms of predictive ability. The

all three characteristics variable, which might be assumed to be the best summary variable, explains less variance (R^2 = .1459).

Individual predictors of aggressor's academic performance.

The pattern of results for academic performance is very similar to the pattern for negative affect, sense of school membership, and conduct problems. Intention to harm emerges as the most important individual predictor of academic performance. It is identified as the most important predictor by all methods and accounts for by far the most unique variance (31.58%). No other individual variable emerges as an important predictor. More of the variance explained by our model is attributable to shared effects than individual predictors (unique effects = 42.13%, shared effects = 77.25%, suppression effects = -19.38%). The all-possible-subsets results are consistent with this finding. The most powerful model for explaining variance in participant power includes all variables (R^2 = .0511). The most powerful individual predictor in terms of R^2 value is intention to harm (R^2 = .0449). In this case, there is only a small difference between these the full model and the single variable in terms of predictive ability. The all three characteristics variable, which might be assumed to be the best summary variable, explains less variance (R^2 = .0243).

Summary

A number of general patterns can be drawn from the results regarding the relationship between individual aggression characteristics and their social and emotional functioning. First, in all cases, the predicting variables are highly related and the majority of the variance accounted for by the overall model is in fact variance that was shared between the dependent variable and multiple dependent variables. This suggests that all of the aggression-related variables are related to a single underlying construct. While this convergent validity is good in that it suggests something important is being captured, it does not support the idea that

each of the definitional characteristics of bullying makes a unique contribution to predicting participant outcomes, nor that outcomes associated with bullying are unique from (and not simply more severe than) outcomes associated with aggression. Indeed, all three characteristics do not emerge as uniquely important for any of the eleven measured outcomes. Including all the variables related to aggression in our analysis often led to the strongest predictive power. This is perhaps largely because of reduced error in measurement of an underlying construct. It is not, however, efficient to have redundant predictors in a definition, as these add complexity to measurement and analyses for relatively little gain in predictive power.

The overall finding that much of the variance among predictors was shared and that our predictors likely reflect different operationalizations of a single underlying construct is very strong. However, there is some variation in the amount of predictive power displayed by individual predictors. This data could help to determine which variables to retain as important if our conceptualization of bullying was to be simplified. The general finding, for which beliefs about aggression is the main exception, is that power imbalance and intention to harm are the most important individual predictors of aggressor's social and emotional functioning. The repetition and generic aggression variables were occasionally significant as well (i.e., for power and popularity, which were the variables with the most variance explained), but were less independently useful overall. The all three characteristics variable was independently useful for predicting only beliefs about aggression. This suggests that the fact that aggression has all three characteristics of aggression tells us little on its own. Therefore, the results suggest that the best measures of the underlying construct in a simplified model would be that aggression occurs in the context of a power imbalance and that it is perceived to be done with the intention to harm.

An interesting pattern of results occurred in which power imbalance was clearly the most important predictor for popularity, power, and social comparison – areas in which perpetrators of aggression had good outcomes – whereas intention to harm was the most important predictor for negative affect, sense of school membership, conduct problems and academic performance – areas where aggressors had negative outcomes. This suggests that the power imbalance criterion is more strongly related to whether aggressors will be successful in achieving social dominance goals, whereas peers' perception that aggressors intend to harm people is more associated with unpleasant emotions and problems at school.

Another interesting pattern is that, besides when suppression effects were at play, the definitional criteria of bullying had, with one exception, the same type of relationship to participant outcomes as the presence of aggression. This result makes it appear that the outcomes associated with bullying, at least in the areas of functioning measured, are not different than the outcomes associated with generic aggression, they are simply often more severe.

Characteristics related to victimization by aggression.

In order to determine whether behaviourally measured characteristics of victimization are related to participant outcomes, I conducted a series of sequential regression analyses, one for each outcome measured. As in the analyses related to the perpetration of aggression, eleven outcomes were measured for each participant: Popularity ($M = 0$, $SD = 1$), Power ($M = 0$, $SD = 1$), Social Comparison ($M = 6.51$, $SD = 1.48$), Submissive Behaviour ($M = 2.57$, $SD = 0.67$), Security and Sense of Acceptance ($M = 3.77$, $SD = 0.86$), Positive Affect ($M = 3.47$, $SD = 0.77$), Negative Affect ($M = 1.67$, $SD = 0.70$), Sense of School Membership ($M = 3.71$, $SD = 0.79$), Positive Beliefs about Aggression ($M =$

Table 5

Model for Sequential Regression Examining the Relationship between the Characteristics of Victimization and Participant Outcomes

Step	Predictor	Description
1	Class size	Number of people in the participant's classroom who participated in the study (i.e., number of potential nominators)
2	Victimization	Number of nominations for victimization
	Intention	Number of nominations for victimization being intentionally harmful
3	Power Imbalance	Number of nominations for victimization occurring in the context of a power imbalance
	Repetition	Number of nominations for victimization being repeated
4	All Three Characteristics	Number of nominators that victimization and endorsed all three definitional characteristics measured

1.41, *SD* = 0.44), Conduct Problems (*M* = 1.21, *SD* = 0.35), and Academic Performance (*M* = 4.57, *SD* = 1.66).

In each regression, predictors were entered in four steps, which are outlined in Table 5. The first step contained a control variable: classroom size (*M* = 19.14, *SD* = 3.79) to control for differences in number of aggressive nominations caused by differences in the size of the nominating pool. The second step also contained one variable: the number of times an individual was nominated to be a victim of aggression (*M* = 2.43, *SD* = 2.88). This step indicates to what extent victimization alone predicts student outcomes. The third step contained three variables, corresponding to the number of nominations for each of the three characteristics of victimization that indicate the victimization meets the criteria for bullying: repetition (*M* = 1.24, *SD* = 1.99), intention to harm (*M* = 1.28, *SD* = 1.98), and power imbalance (*M* = 1.12, *SD* = 1.85). This step indicates to what extent the specific characteristics of aggressive behaviour contained within the definition of bullying add to our ability to predict student functioning related to victimization. The fourth and final set of

Table 6

Correlation among Predictors in Regression Predicting Participant Functioning from Characteristics of Victimization

	Number of Students in Classroom	Victimization	Repeated	Intentionally Hurtful	Power Imbalance
Victimization	.169***	–	–	–	–
Repeated	.041	.840***	–	–	–
Intentionally Hurtful	.159***	.828***	.846***	–	–
Power Imbalance	.071	.790***	.806***	.840***	–
All Three Characteristics	.065	.706***	.797***	.825***	.897***

Note: * $p < .05$ *** $p < .001$

predictors includes one variable: the number of times that a participant's victimization was nominated as having all three characteristics that define bullying ($M = 0.65$, $SD = 1.29$). According to the definition, all three characteristics must be present in order for the behaviour to be bullying, so this step indicates to what extent, if any, the fact that the victimization is victimization by bullying adds to our ability to predict student functioning.

An initial examination of the predictors in this model indicated that the predictors related to victimization were highly related (see Table 6 for details). The size of these correlations indicated that multicollinearity among predictors is present. To ensure that there was some differentiation among characteristics of victimization, the endorsement of differential power, repetition, and frequency nomination combinations was examined. Figure 3 illustrates the percentage of participants who were victimized with each possible combination of variable characteristics. Roughly one quarter of participants displayed aggression only, one third displayed aggression with some characteristics that did not meet the definition of bullying, and two fifths met the full criteria for bullying. Despite the high

correlation, this was considered to be enough differentiation to consider the characteristics variables not completely identical.

In order to avoid problems in interpretation associated with multicollinearity, in this analysis we followed guidelines for interpreting the results of multiple regression laid out by Kraha, Turner, Nimon, Zientek, and Henson (2012) and Nathans, Oswald, and Nimon, (2012). Variable importance was examined through an analysis of standardized coefficients (beta weights), structure coefficients, all-possible-subsets regression, commonality analysis, and dominance weights. The analyses are believed to capable of providing valid information using these techniques. For the same reasons as described in the regression series related to the perpetration of aggression, none of the definitional criteria were excluded from the analysis, but interaction terms were not investigated because of the redundancy.

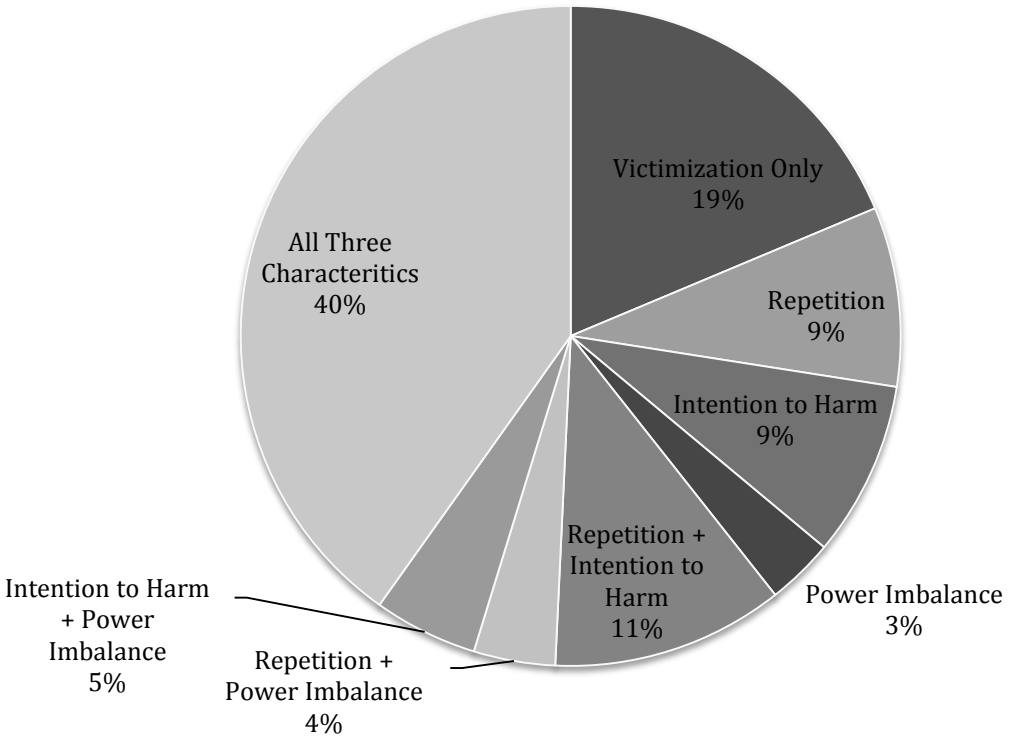

Figure 3. Overlap Between Characteristics of Victimization.

Each of the 11 regressions run was examined for possible violations of the assumptions of regression analysis. For each regression, residuals were plotted and examined for problems with linearity and homoscedasticity. For most variables, no problems were found. For the negative beliefs about aggression scores, the assumption of homoscedasticity appears to have been violated. This weakens the analysis but does not invalidate it (Tabachnick and Fidel, 2004, p. 127). Cook's Distance and leverage were also charted and examined. Leverage values within the "risky" range (0.2-0.49; Norušis, 2012) were rare, but were tolerated long as Cook's Distance values indicated that no case displayed undue influence on the regression coefficients (Cook's Distance > 1; Tabachnick and Fidel, 2004). No data points in any analysis were found to have Leverage > .5 or Cook's Distance > 1. Therefore, all cases were retained for the analysis and data was not truncated in any way.

Outliers in the solution were found in a number of regressions. Three outliers in the solution were found when predicting social comparison. In each case of these cases, the predicted value of the self-perceived social rank was close to average and the actual ASCS score was 2, the minimum value seen in our sample. Two outliers in the solution were found when predicting submissive behaviour. In both of these cases, the predicted values were close to average, but the ACBS score was 5, the maximum value seen in the sample. Two outliers in the solution were also found when predicting sense of acceptance and security. In both of these cases, the predicted values were close to average, but the ACBS score was 1, the minimum value seen in the sample. Five outliers in the solution were found when predicting negative affect. In all of these cases, the predicted values were close to average, but the actual score on the PANAS negative affect scale was four or five, towards the top of the range seen in our sample. Two outliers in the solution were also found when predicting sense of positive affect. In both of these cases, the predicted values were close to average,

but the PANAS positive affect score was 1, the minimum value seen in the sample. Two outliers in the solution were also found when predicting sense of school membership. In both of these cases, the predicted values were close to average, but the PANAS positive affect score was 1, the minimum value seen in the sample. Three outliers in the solution were found when predicting negative beliefs about aggression. In each case of these cases, the predicted value was much lower than the observed value of the NBAAS, which was at the upper end of the range of responses observed. Eleven outliers in the solution were found when predicting teacher-rated conduct problems (not including bullying). In 10 cases, the predicted value was lower than the observed value, but in one case this pattern was reversed. The model does not fit well for participants with this pattern of responses, but they make up a small percentage of the overall sample. The cases were retained in the analysis to avoid artificially retrofitting the regression equations to our data.

Overall relationship between aggression and its characteristics and participant outcomes.

Results related to this series of regressions are presented in Table 7. The full model (step 4) was significantly related to all measures of participant functioning except beliefs about aggression, which is significant only after Model 3. The R^2 values reported indicate the proportion of the variance in the outcomes that is accounted for by its relationship with all the predictors in our model. By convention, the amount of variance accounted for in all variables would be considered small. The strongest relationships were between the behaviourally measured victimization and power, popularity, and self perceived social rank. Being the victim of aggression that meets the criteria for bullying was related to having less power, lower popularity, and lower self-perceived social rank. It was also associated with more submissive behaviour, lower felt sense of security and acceptance, more negative

Table 7

Model Level Results of Stepwise Regression Predicting Participant Functioning with Characteristics of Victimization

Outcome	*F*	(df)	R^2	ΔR^2
Popularity				
Step 1	0.0000	(1, 533)	.000	.000
Step 2	29.857***	(2, 532)	.101	.101***
Step 3	35.594***	(5, 529)	.252	.151***
Step 4	30.843***	(6, 528)	.260	.008*
Power				
Step 1	0.0000	(1, 533)	.000	.000
Step 2	19.854***	(2, 532)	.066	.069***
Step 3	24.651***	(5, 529)	.181	.120***
Step 4	21.214***	(6, 528)	.185	.005
Social Comparison				
Step 1	0.0290	(1, 476)	.000	.000
Step 2	9.833***	(1, 475)	.040	.040***
Step 3	10.270***	(5, 472)	.098	.058***
Step 4	9.022***	(6, 471)	.103	.005
Submissive Behaviour				
Step 1	4.146*	(1, 474)	.009	.009*
Step 2	5.387*	(2, 472)	.022	.014*
Step 3	4.312***	(5, 470)	.044	.022*
Step 4	3.596*	(6, 469)	.044	.000
Security and Acceptance				
Step 1	0.3590	(1, 462)	.001	.001
Step 2	9.276***	(2, 461)	.039	.038***
Step 3	5.620***	(5, 458)	.058	.019*
Step 4	5.042***	(6, 457)	.062	.004
Positive Affect				
Step 1	0.0840	(1, 450)	.000	.000
Step 2	4.579*	(2, 449)	.020	.020*
Step 3	3.558*	(5, 446)	.038	.018*
Step 4	3.115*	(6, 445)	.040	.002
Negative Affect				
Step 1	0.0630	(1, 442)	.000	.000
Step 2	13.619***	(2, 441)	.058	.058***
Step 3	8.469***	(5, 438)	.088	.030*
Step 4	7.161***	(6, 437)	.090	.001
	F	(df)	R^2	ΔR^2

Note: Table continued next page. * $p < .05$ *** $p < .001$

Table 7 continued

	F	(df)	R^2	ΔR^2
School Membership				
Step 1	0.5430	(1, 477)	.001	.001
Step 2	13.853***	(2, 476)	.055	.054***
Step 3	6.617***	(5, 473)	.065	.010
Step 4	5.608***	(6, 472)	.067	.001
Beliefs about Aggression				
Step 1	0.9430	(1, 469)	.002	.002
Step 2	0.4710	(2, 468)	.002	.000
Step 3	2.328*	(5, 465)	.024	.022*
Step 4	1.9500	(6, 464)	.025	.000
Conduct Problems				
Step 1	3.1650	(1, 526)	.006	.006
Step 2	12.464***	(2, 525)	.045	.039***
Step 3	5.510***	(5, 522)	.050	.005
Step 4	5.211***	(6, 521)	.057	.006
Academic Performance				
Step 1	0.6930	(1, 526)	.001	.001
Step 2	9.552***	(2, 525)	.035	.034***
Step 3	4.143***	(5, 522)	.038	.003
Step 4	3.479*	(6, 521)	.039	.000
	F	(df)	R^2	ΔR^2

Note: * $p < .05$ *** $p < .001$

affect, less positive affect, lower sense of school membership, more conduct problems and lower academic performance. In general, these results indicate that being victimized by bullying is associated with poor social and individual functioning.

Do the definitional characteristics of bullying predict functioning above and beyond the presence of victimization by peers?

The significance of individual steps in the model tells us which sets of predictors contribute significantly to our ability to predict participant outcomes. First, the results indicate that the victimization of aggression alone significantly predicts many measures of social and interpersonal functioning. The victimization by aggression variable is related to all outcomes variables except beliefs about aggression. This is important because it implies that

victimization does not need to be victimization by bullying to an important indicator of functioning.

Second, the results indicate that the addition of the three measured behavioural characteristics of aggression does add to our ability to predict most, but not all, examined measures of student functioning. There are three outcome variables measured – sense of school membership, conduct problems, and academic performance – that are not related to whether victimization is repeated, intentionally harmful, or occurs in the context of a power imbalance in our sample. For the remaining eight variables, however, information about whether the characteristics of bullying are present increases our ability to predict student functioning. This is important because it indicates that victimization that meets some of the criteria for victimization by bullying has a stronger relationship to several measures of functioning than victimization that does not meet the criteria for bullying. Specifically, those who are victimized by aggression with some characteristics of bullying are less popular, see themselves as having a lower social rank, and show more submissive behaviour than those who are victimized in other ways. At the same time, they feel less secure, experience less positive and more negative affect and have more positive views about the acceptability of aggression.

There are two other points of interest when examining the contributions of victimization and the definitional criteria of bullying in these regressions. First, the relationship between victimization and these outcomes and the definitional characteristics as a group and these outcomes is in the same direction. That implies that the outcomes associated with victimization by bullying are not unique, but simply more severe than the outcomes associated with generic victimization. Second, in general, the definitional characteristics of bullying appear to contribute much more to our understanding of variables

related to social dominance (popularity 61%, power 68%, self-perceived social rank 61%, submissive behaviour 50% of total variance explained) than they do for the individual variables (security and acceptance 37%, positive affect 50%, negative affect 34%, school membership 16%, conduct problems 19%, academic performance 8%). This implies that the behavioural criteria that define bullying are especially predictive of the effects of victimization on social status. Beliefs about aggression is an exception to this trend, as it is not related to generic victimization at all.

Is meeting full definitional criteria for bullying meaningful in that the presence of all three definitional criteria predicts functioning above and beyond the presence of the individual critieria?

The fourth step in our regression model, which includes the all three characteristics variable, does not add significantly to our ability to predict student functioning for any variable except participants' popularity. As such, it does not appear to be especially meaningful in terms of predicting student outcomes to have all three defining characteristics of victimization by bullying present. As with our examination related to the perpetration of bullying, a potential explanation for this finding is the high correlation between the all three characteristics variable and the characteristics variables entered in the previous step of the analysis.

Are each of the three definitional characteristics of bullying independently significant for predicting participant functioning?

To examine the nature of the relationship between specific characteristics of victimization and participant outcomes, we turn to statistics on the significance of individual predictors. As stated, we examined a number of indicators of importance, as high levels of multicollinearity make standardized coefficients unreliable. Statistics related to individual

predictors in step four of the model, in which all predictors are included, are presented for all significantly predicted dependent variables in Table 8. Results for individual outcomes will be summarized briefly and then a general summary and conclusion will follow.

Individual predictors of popularity for people who are victimized.

Two variables emerge in this regression as important individual predictors of participant popularity: power imbalance and repetition. On the whole, power imbalance appears to be the more important predictor. Though it doesn't explain as much of the unique variance as repetition does (14.11% versus 15.90%), it is the more prominent predictor by most measures. Repetition also appears to be acting as a net suppressor variable, as its beta weight has an opposite sign to its overall relationship with the variable. It appears to suppress error especially in larger combinations of variables that contain victimization, intention and power imbalance. It is also individually predicts popularity, accounting for quite a large unique effect. Intention to harm, victimization and the all three characteristics predictor do not emerge as strong individual predictors here, though they do share quite a large amount of common variance with other predictors. Overall, most of the variance explained in this regression comes from combined variables (unique effects = 56.04%, Shared Effects = 110.02%, Suppression effects = -66.06%). The all-possible-subsets regression analysis is consistent with this finding. The most powerful model for explaining variance in participant power includes all variables (R^2 = .2600) although multiple variables can be excluded with only a small loss of power. The best two-variable combination, power imbalance and repetition (R^2 = .247) accounts for almost as much variance. The most powerful individual predictor, power imbalance, on its own has an R^2 value of .210.

Table 8

Individual Predictor Results of Stepwise Regression Predicting Participant Functioning with Characteristics of Victimization

Dependent Variable	Predictor	β	r_s	r_s^2	r	R^2	Unique	Common	General Dominance Weights	Relative Importance Weights
Popularity	# of nominators	0.059	0.000	0.000	0.000	0.000	0.003	-0.003	0.002	0.001
	Victimization	-0.123	-0.614	0.377	-0.313	0.098	0.003	0.095	0.029	0.033
Total R^2 =	Intention	-0.045	-0.685	0.469	-0.349	0.122	0.000	0.121	0.038	0.039
.260	Power Imbalance	-0.507*	-0.898	0.806	-0.458	0.210	0.037	0.173	0.108	0.090
	Repeated	0.468*	-0.500	0.250	-0.255	0.065	0.041	0.024	0.038	0.028
	All 3 Characteristics	-0.220*	-0.828	0.686	-0.422	0.178	0.008	0.170	0.073	0.070
Power	# of nominators	0.050	0.000	0.000	0.000	0.000	0.002	-0.002	0.001	0.001
	Victimization	-0.075	-0.591	0.349	-0.260	0.068	0.001	0.066	0.020	0.023
Total R^2 =	Intention	-0.062	-0.682	0.465	-0.300	0.090	0.001	0.089	0.028	0.030
.185	Power Imbalance	-0.451*	-0.893	0.797	-0.393	0.154	0.029	0.126	0.082	0.068
	Repeated	0.407*	-0.486	0.236	-0.214	0.046	0.031	0.015	0.029	0.021
	All 3 Characteristics	-0.181	-0.822	0.676	-0.362	0.131	0.005	0.126	0.054	0.052
Social Comparison	# of nominators	0.054	0.025	0.001	0.008	0.000	0.003	-0.003	0.001	0.001
	Victimization	-0.092	-0.605	0.366	-0.195	0.038	0.002	0.036	0.011	0.013
	Intention	-0.093	-0.705	0.497	-0.227	0.052	0.002	0.050	0.017	0.018
Total R^2 =	Power Imbalance	-0.254*	-0.869	0.755	-0.280	0.078	0.009	0.069	0.038	0.032
.103	Repeated	0.342*	-0.478	0.228	-0.154	0.024	0.022	0.002	0.017	0.012
	All 3 Characteristics	-0.173	-0.826	0.682	-0.266	0.071	0.005	0.066	0.030	0.028
Dependent Variable	Predictor	β	r_s	r_s^2	r	R^2	Unique	Common	General Dominance Weights	Relative Importance Weights

Table 8 continued

Dependent Variable	Predictor	β	r_s	r_s^2	r	R^2	Unique	Common	General Dominance Weights	Relative Importance Weights
Submissive Behaviour	# of nominators	0.109*	0.440	0.194	0.093	0.009	0.011	-0.002	0.008	0.008
	Victimization	-0.044	0.620	0.384	0.131	0.017	0.000	0.017	0.005	0.005
	Intention	-0.197	0.535	0.286	0.113	0.013	0.007	0.006	0.006	0.004
Total R^2 =	Power Imbalance	0.236*	0.781	0.610	0.165	0.027	0.008	0.019	0.013	0.011
.044	Repeated	0.184	0.715	0.511	0.151	0.023	0.006	0.016	0.009	0.009
	All 3 Characteristics	-0.029	0.672	0.452	0.142	0.020	0.000	0.020	0.006	0.007
Security and Acceptance	# of nominators	0.009	-0.133	0.018	-0.028	0.001	0.000	0.001	0.019	0.021
	Victimization	-0.134	-0.792	0.627	-0.197	0.039	0.004	0.035	0.131	0.114
	Intention	-0.021	-0.808	0.653	-0.201	0.040	0.000	0.040	0.141	0.115
Total R^2 =	Power Imbalance	-0.112	-0.924	0.854	-0.230	0.053	0.002	0.051	0.134	0.112
.062	Repeated	0.183	-0.659	0.434	-0.164	0.027	0.006	0.021	0.103	0.085
	All 3 Characteristics	-0.159	-0.904	0.817	-0.225	0.051	0.004	0.047	0.174	0.153
Positive Affect	# of nominators	0.052	0.070	0.005	0.014	0.000	0.002	-0.002	0.001	0.001
	Victimization	-0.162	-0.679	0.461	-0.136	0.018	0.006	0.013	0.008	0.009
	Intention	-0.034	-0.654	0.428	-0.131	0.017	0.000	0.017	0.005	0.006
Total R^2 =	Power Imbalance	-0.130	-0.809	0.654	-0.162	0.026	0.002	0.024	0.012	0.011
.040	Repeated	0.257*	-0.435	0.189	-0.087	0.008	0.013	-0.005	0.008	0.005
	All 3 Characteristics	-0.096	-0.734	0.539	-0.147	0.022	0.002	0.020	0.008	0.009
Negative Affect	# of nominators	-0.012	0.040	0.002	0.012	0.000	0.000	0.000	0.000	0.000
	Victimization	-0.031	0.798	0.637	0.238	0.057	0.000	0.056	0.014	0.012
	Intention	0.085	0.922	0.850	0.275	0.076	0.001	0.074	0.022	0.018
Total R^2 =	Power Imbalance	0.105	0.956	0.914	0.285	0.081	0.002	0.080	0.025	0.021
.09	Repeated	0.060	0.892	0.796	0.266	0.071	0.001	0.070	0.019	0.016
	All 3 Characteristics	0.096	0.956	0.914	0.285	0.081	0.002	0.080	0.025	0.022
Dependent Variable	Predictor	β	r_s	r_s^2	r	R^2	Unique	Common	General Dominance Weights	Relative Importance Weights

Table 8 continued

Dependent Variable	Predictor	β	r_s	r_s^2	r	R^2	Unique	Common	General Dominance Weights	Relative Importance Weights
School Membership	# of nominators	0.074	0.132	0.017	0.034	0.001	0.005	-0.004	0.003	0.003
	Victimization	-0.159	-0.863	0.745	-0.222	0.049	0.005	0.044	0.016	0.015
	Intention	-0.034	-0.839	0.704	-0.216	0.047	0.000	0.046	0.013	0.011
Total R^2 = .067	Power Imbalance	-0.095	-0.921	0.848	-0.237	0.056	0.001	0.055	0.018	0.015
	Repeated	0.108	-0.773	0.598	-0.199	0.040	0.002	0.037	0.010	0.008
	All 3 Characteristics	-0.090	-0.870	0.757	-0.224	0.050	0.001	0.049	0.015	0.013
Conduct Problems	# of nominators	-0.115*	-0.325	0.106	-0.077	0.006	0.012	-0.006	0.010	0.009
	Victimization	0.067	0.765	0.585	0.181	0.033	0.001	0.032	0.011	0.011
	Intention	0.186	0.786	0.618	0.186	0.035	0.006	0.029	0.014	0.013
Total R^2 = .057	Power Imbalance	0.142	0.727	0.529	0.172	0.030	0.003	0.027	0.010	0.009
	Repeated	0.009	0.735	0.540	0.174	0.030	0.000	0.030	0.009	0.008
	All 3 Characteristics	-0.199	0.545	0.297	0.129	0.017	0.006	0.010	0.007	0.005
Academic Performance	# of nominators	-0.014	-0.184	0.034	-0.036	0.001	0.000	0.001	0.000	0.001
	Victimization	-0.096	-0.954	0.910	-0.187	0.035	0.002	0.033	0.012	0.010
	Intention	0.011	-0.862	0.743	-0.169	0.029	0.000	0.029	0.007	0.006
Total R^2 = .039	Power Imbalance	-0.122	-0.918	0.843	-0.180	0.032	0.002	0.030	0.010	0.009
	Repeated	-0.043	-0.887	0.787	-0.174	0.030	0.000	0.030	0.009	0.007
	All 3 Characteristics	0.049	-0.785	0.616	-0.154	0.024	0.000	0.023	0.006	0.005
Dependent Variable	Predictor	β	r_s	r_s^2	r	R^2	Unique	Common	General Dominance Weights	Relative Importance Weights

Note: Total R^2= the total amount of variance predicted by the full model. The primary predictor suggested by a method is underlined. β = the standardized correlation coefficient (Beta weight). r_s = structure coefficient. r_s^2 = something***. r = the correlation between the predictor and the outcome variable. R^2 = amount of variance in the outcome variable explained by the predictor when it is entered as the only predictor. Unique = proportion of criterion variance explained uniquely by the predictor. Common = proportion of criterion variance explained by the predictor that is also explained by one or more other predictors. Unique + Common = r^2. General Dominance Weights = something***. Relative Importance weights = something***. Σ General Dominance Weights = Σ Relative Importance Weights = R^2.

Individual predictors of power for people who are victimized.

The pattern of results for power is very similar to the pattern for popularity. Two variables emerge in this regression as important individual predictors: power imbalance and repetition. On the whole, power imbalance appears to be the more important predictor. Though it doesn't explain as much of the unique variance as repetition does (14.94% versus 16.17%) it is the more prominent predictor by most measures. Repetition appears to be acting as a net suppressor variable, as its beta weight has an opposite sign to its overall relationship with the variable. It appears to suppress error, especially in larger combinations of variables that contain victimization, intention and power imbalance. It is also a predictor of popularity independently, accounting for quite a large unique effect. Intention, victimization and the all three characteristics predictor do not emerge as strong individual predictors here, though they do share quite a large amount of common variance with other predictors. Overall, most of the variance explained in this regression comes from combined variables (unique effects = 35.95%, shared effects = 99.21%, suppression effects = -35.17%). The all-possible-subsets regression analysis is consistent with this finding. The most powerful model for explaining variance in participant power includes all variables (R^2 = .1938), although multiple variables can be excluded with only a small loss of power. The best two-variable combination, power imbalance and repetition (R^2 = .1846) accounts for almost as much variance. The most powerful individual predictor, power imbalance, on its own has an R^2 value of .1544.

Individual predictors of social comparison for people who are victimized.

The pattern of results for social comparison is very similar to the results for popularity and power. Two variables emerge in this regression as important individual predictors: power imbalance and repetition. On the whole, power imbalance appears to be the

more important predictor. Though it doesn't explain as much of the unique variance as repetition does (8.74% versus 21.40%), it is the more prominent predictor by most measures. Repetition appears to be acting as a net suppressor variable, as its beta weight has an opposite sign to its overall relationship with the variable. It appears to suppress error, especially in larger combinations of variables that contain victimization, intention and power imbalance. It is also a predictor of popularity independently, accounting for quite a large unique effect. Intention, victimization and the all three characteristics predictor do not emerge as strong individual predictors here, though they do share quite a large amount of common variance with other predictors. Overall, most of the variance explained in this regression comes from combined variables (unique effects = 35.95%, shared effects = 99.21%, suppression effects = -35.17%). The all-possible-subsets regression analysis is consistent with this finding. The most powerful model for explaining variance in participant power includes all variables (R^2 = .1938), although multiple variables can be excluded with only a small loss of power. The best two-variable combination, power imbalance and repetition (R^2 = .1846) accounts for almost as much variance. The most powerful individual predictor, power imbalance, on its own has an R^2 value of .1544.

Individual predictors of submissive behaviour for people who are victimized.

Two individual variables are identified as statistically significant predictors of submissive behaviour: power imbalance and our control variable, number of potential nominators. On the whole, power imbalance appears to be the most important predictor. Though it doesn't explain as much of the unique variance as the number of nominators variable does (15.51% versus 24.34%), it is the more prominent predictor by most measures. The number of nominators variable does not share much variance with other predictors. Its significance here is probably to do with classroom features. Two other variables contribute

almost as much unique variance as power imbalance, although their beta weights are not statistically significant. Intention to harm appears to be acting as a net suppressor variable, as its beta weight has an opposite sign to its overall relationship with the variable. Repetition also appears to have some amount of utility as a unique predictor. It appears to suppress error, especially in larger combinations of variables that contain victimization, intention and power imbalance. It is also a predictor of submissive behaviour independently, accounting for quite a large unique effect. Victimization and the all three characteristics predictor do not emerge as strong individual predictors here, though they do share quite a large amount of common variance with other predictors. Overall, most of the variance explained in this regression comes from combined variables (unique effects = 73.19%, shared effects = 87.17%, suppression effects = -60.36%). The all-possible-subsets regression analysis is consistent with this finding. The most powerful model for explaining variance in participant power includes all variables (R^2 = .0446), although multiple variables can be excluded with only a small loss of power. A model containing number of nominators, intention to harm, power imbalance and repetition has an R^2 value of .0442. The best two-variable combination, power imbalance and number of nominators (R^2 = .0339) accounts for almost as much variance. The most powerful individual predictor, power imbalance, on its own has an R^2 value of .0272.

Individual predictors of security and acceptance for people who are victimized.

No individual variables are identified as statistically significant predictors of security and acceptance. Two variables appear to be more prominent than others. Repetition appears to be acting as a net suppressor variable, as its beta weight has an opposite sign to its overall relationship with the variable. It appears to suppress error, especially in larger combinations of variables that contain victimization, intention and power imbalance. Power imbalance is

also identified as an important predictor by a number of methods, despite the fact that it explains very little unique variance (< 3% of the variance explained). Most of the variance accounted for in this regression is in fact shared (unique effects = 26.04%, shared effects = 98.37%, suppression effects = -24.40%). The all-possible-subsets regression analysis is consistent with this finding. The most powerful model for explaining variance in participant power includes all variables (R^2 = .0619), although multiple variables can be excluded with only a small loss of power. The four-variable combination of victimization, power imbalance, repetition, and all three characteristics has an R^2 value of .0618. The best two-variable combination, power imbalance and all three characteristics has an R^2 value of .0547. The commonly efficient combination of power imbalance and repetition (R^2 = .0542) accounts for almost as much variance. The most powerful individual predictor, power imbalance, on its own has an R^2 value of .0529.

Individual predictors of positive affect for people who are victimized.

Only repetition is a statistically significant individual predictor of positive affect. It appears to be acting as a net suppressor variable, as its beta weight has an opposite sign to its overall relationship with the variable. It appears to suppress error, especially in the victimization variable and larger combinations of variables. Power imbalance is also identified as an important predictor by a number of methods, despite the fact that it accounts for only 6.05% of the variance explained. Overall, most of the variance explained in this regression comes from combined variables (unique effects = 61.33%, shared effects = 106.91%, suppression effects = -68.23%). The all-possible-subsets regression analysis is consistent with this finding. The most powerful model for explaining variance in participant power includes all variables (R^2 = .0401), although multiple variables can be excluded with only a small loss of power. The four-variable combination of number of nominators,

victimization, power imbalance, and all three characteristics has an R^2 value of .0618. The best two-variable combination, power imbalance and repetition has an R^2 value of .0317. The most powerful individual predictor, power imbalance, on its own has an R^2 value of .0262.

Individual predictors of negative affect for people who are victimized.

No individual predictor has a statistically significant beta weight in our prediction of negative affect. Power imbalance and the all three characteristics predictor appear to be the two most important predictors: they are tied in strength in almost all indicators. Overall, most of the variance explained in this regression comes from combined variables (unique effects = 6.00%, shared effects = 97.33%, suppression effects = -3.33%). The all-possible-subsets regression analysis is consistent with this finding. The most powerful model for explaining variance in participant power includes all variables (R^2 = .0889), although multiple variables can be excluded with only a small loss of power. The best two-variable combination, intention to harm and all three characteristics has an R^2 value of .0862. The commonly efficient combination of power imbalance and repetition (R^2 = .0850) accounts for almost as much variance. The most powerful individual predictors, power imbalance and the all three characteristics variable, both have an R^2 value of .0812 on their own.

Individual predictors of sense of school membership for people who are victimized.

No individual predictor has a statistically significant beta weight in our prediction of sense of school membership. Though victimization has the highest beta weight, power imbalance appears to be the most important predictor. Though it doesn't explain as much of the unique variance as the number of victimization does (8.08% versus 1.94%), it is the more prominent predictor by most measures. Overall, most of the variance explained in this regression comes from combined variables (unique effects = 23.21%, shared effects = 98.29%, suppression effects = -21.49%). The all-possible-subsets regression analysis is

consistent with this finding. The most powerful model for explaining variance in participant power includes all variables (R^2 = .0662), although multiple variables can be excluded with only a small loss of power. The best two-variable combination, victimization and power imbalance, has an R^2 value of .0594. The commonly efficient combination of power imbalance and repetition (R^2 = .0564) accounts for almost as much variance. The most powerful individual predictor, power imbalance, has an R^2 value of .0562 on its own.

Individual predictors of conduct problems for people who are victimized.

Only our control variable, number of nominators, is a statistically significant individual predictor of positive affect. It accounts for more unique variance than other predictors. Intention to harm is also identified as an important predictor by a number of methods. Overall, most of the variance explained in this regression comes from combined variables (unique effects = 50.74%, shared effects = 93.34%, suppression effects = -44.08%). The all-possible-subsets regression analysis is consistent with this finding. The most powerful model for explaining variance in participant power includes all variables (R^2 = .0560), although multiple variables can be excluded with only a small loss of power. The four-variable combination of number of nominators, intention to harm, power imbalance, and all three characteristics has an R^2 value of .0618. The best two-variable combination, number of nominators and intention to harm has an R^2 value of .0462. The most powerful individual predictor, intention to harm, on its own has an R^2 value of .0346.

Individual predictors of academic performance for people who are victimized.

No individual predictor has a statistically significant beta weight in our prediction of sense of academic performance. Though power imbalance has the highest beta weight, victimization appears to be the most important predictor. Neither variable explains very much unique variance. Most of the variance explained in this regression comes from

combined variables (unique effects = 13.01%, shared effects = 94.02%, suppression effects = -7.03%). The all-possible-subsets regression analysis is consistent with this finding. The most powerful model for explaining variance in participant power includes all variables (R^2 = .0395), although multiple variables can be excluded with only a small loss of power. Five two-variable combinations had an R^2 values between .0350 and .0359. The commonly efficient combination of power imbalance and repetition (R^2 = .0348) accounts for almost as much variance. The most powerful individual predictor, victimization, has an R^2 value of .0350 on its own.

Summary and conclusions regarding the importance of individual victimization characteristics in predicting participant social and emotional functioning.

A number of general patterns can be drawn from the results regarding the relationship between individual victimization characteristics and participants' social and emotional functioning. First, as was the case for variables related to aggression, in all analyses here, the predicting variables are highly related and the majority of the variance accounted for by the overall model is in fact variance that was shared between the dependent variable and multiple dependent variables. This suggests that all predictors related to victimization are related to a single underlying construct. While convergent validity is good in that it suggests something important is being captured, it does not support the idea that each of the definitional characteristics of bullying makes a unique contribution to predicting participant outcomes, nor that outcomes associated with victimization by bullying are unique from (and not simply more severe than) outcomes associated with victimization in general. None of the three definitional characteristics, nor the combination of the three, emerges as a strongly unique and important predictor for any of the eleven measured outcomes, despite the fact that including all the variables related to aggression in our analysis often lead to the strongest

predictive power. The later finding is perhaps because of reduced error in measurement of an underlying construct that occurs with multiple measures. It is not, however, efficient to have redundant predictors in a definition, as these add complexity to measurement and analyses for relatively little gain in predictive power.

This brings us to our second trend in the results, which concerns which predictors emerged as more important than the others. The results for victimization are not as clear cut as the results for aggression. However, the general trend is that power imbalance is the most important predictor of the social and emotional functioning of people who are victimized. Academic performance and conduct problems are exceptions to this trend. Also, in general, repetition, not intention to harm is the second most important predictor. Though there are exceptions to this, repetition contributes significant variance, and in combination with power imbalance best predicts the outcomes measured here in a two-predictor model, more often than any other variable. This is not to say that intention to harm, victimization, or the presence of all three predictors is not related to the outcomes measured here – they are and can often explain significant variance in the outcome; it's just that power imbalance and repetition share much of this variance with the outcome and make more efficient predictors of functioning as it is related to victimization. Therefore, power imbalance and repetition of victimization variables should be retained in models that measure bullying.

Summary.

This analysis addressed four questions of interest:

1. Do the definitional characteristics of bullying predict functioning above and beyond the presence of aggressive behaviour or victimization by peers?

2. Is meeting full definitional criteria for bullying meaningful in that the presence of all three definitional criteria predicts functioning above and beyond the presence of the individual critieria?
3. Are each of the three definitional characteristics of bullying independently significant for predicting participant functioning?
4. What type of functioning is associated with the definitional criteria of bullying?

I addressed these questions through two series of multiple regression analyses, one which addressed characteristics of aggression and the other which addressed characteristics of victimization. For both aggression and victimization, our models generally were related to participant functioning (the two exceptions were that submissive behaviour was not related to aggression and beliefs about aggression was not significantly related to victimization). The amount of variance in functioning explained varied across our dependent variables. Both our model of bullying and our model of victimization by bullying were most strongly related to popularity and power. Regarding our fourth question, perpetrating aggression that has some of the criteria of bullying was associated with better social functioning (in terms of social dominance), but less sense that they are accepted socially for who they are. It was also associated with undesirable individual outcomes. Being the victim of aggression that had some of the criteria of bullying predicted worse social and individual functioning overall.

In terms of our first questions of interest, we can state that information about the definitional characteristics of bullying improves our ability to predict functioning above and beyond the presence of aggressive behaviour or victimization by peers. The amount of variance accounted for by information about the definitional criteria of bullying varied by dependent variable. Bullying criteria accounted for between 18% and 100% of the variance

accounted for in our ability to predict outcomes associated with aggression. For participants who were victimized, information about definitional criteria accounted for ranged between 10% and 64%. In general, the definitional characteristics of bullying accounted for more variance in variables related to social dominance than variables related to individual functioning for people who were victimized. It is also interesting to note that, with one exception, the definitional criteria for bullying were related to outcomes in the same way (i.e., the same direction) as aggression or victimization on their own. This suggests that the outcomes associated with bullying are not different, but simply more severe than those associated with aggressive involvement in general.

Regarding our second question, the presence of all three characteristics of bullying did not emerge as an important predictor of participant outcomes in our analyses. This was not because the presence of all three characteristics was unrelated to participant outcomes, but because the presence of all three characteristics was not uniquely related to participant outcomes. Most of the variance in outcomes that could be accounted for by the presence of all three characteristics (or meeting full definitional criteria) was already accounted for before the all three characteristics variable entered our model. Therefore, it appears that the requirement that aggression or victimization have all three definitional characteristics of bullying is redundant.

In terms of our third question, about which characteristics of aggression or victimization, in particular, were significant for predicting participant functioning, the results show that for both bullying and victimization our predictors were highly related and most of the variance explained was shared among multiple variables. This suggests that, to some extent, all three definitional criteria for bullying are ways of quantifying a single underlying construct. If one predictor was to be chosen as the most important, however, it appears that

the correct choice would be the presence of a power imbalance in aggressive relationships. It often accounts for the most unique variance and is also generally the single variable able to capture the most overall variance. Intention to harm also appeared to be important for predicting the outcomes of aggressors, particularly the indicators of poor individual functioning. With regards to victimization, however, the second most important predictor of functioning appeared to be repetition, although this pattern was less clear across predictors.

Study 2 Results

The purpose of Study Two is to examine the construct validity of bullying in terms of whether common measurement strategies accurately assess the construct as defined. Evidence that we are accurately able to discriminate between bullying as defined and other forms of aggression will add to confidence in both the utility of the definition and that research which intends to capture bullying specifically accurately does so. This study consists of three sets of analyses examining the relationship between participant nominations for behaviours that meet the criteria for bullying (i.e., the explicit behavioural method of assessing bullying with questions about definitional characteristics) with participant identifications of who is involved in bullying (i.e., the definitional method of assessing bullying). The first set of analyses examines the amount of agreement between categorical identifications of bullying involvement made using the explicit behavioural and definitional methods. The second set examines which behaviourally measured characteristics predict bullying status as identified through definitional methods. The third set of analyses examines the extent to which differences between who is identified as involved in bullying by different methods are meaningful in terms of participant outcomes.

Table 9

Participant Bullying Involvement Status Across Methods of Measurement

Method	# of individuals with specified bullying status			
	Bully	Victim	Bully/Victim	Uninvolved
Definitional	77 (14%)	162 (30%)	66 (12%)	230 (43%)
Behavioural	68 (13%)	92 (17%)	17 (3%)	358 (67%)

Agreement between definitional and explicit behavioural methods of measuring bullying involvement.

This section of our results contains analyses investigating to what extent our definitional measure of bullying involvement agrees with the explicit behavioural method of measuring bullying involvement. The purpose of this analysis is to examine whether definitional measurement techniques identify participants who perpetrate or are victimized by aggression that is repeated, intentionally harmful, and occurs in the context of a power imbalance (i.e., aggression that meets the definition of bullying). Because bullying involvement it typically measured categorically, our raw number of nominations data was made categorical using a two-nomination cut-off as detailed in our section on the operationalization of peer nominations. Bullying involvement was divided into four widely recognized categories: the Bully group, the Victim group, the Bully/Victim group, and an Uninvolved group.

For ease of comparison, the number of participants with different bullying involvement statuses calculated using the definitional method (based on responses to questions about bullying involvement) and the explicit behavioural method (based on peer reports of behaviour and behavioural characteristics) are presented in Table 9. Table 10

shows the relationship between the two measurement methods. To test the statistical relationship between the definitional and explicit behavioural methods of measuring bullying at the group level, I completed a two-way contingency table analysis. The two variables were definitional bullying status with four levels (Bully, Victim, Bully/Victim, Uninvolved) and explicit behavioural bullying status with four levels (Bully, Victim, Bully/Victim, Uninvolved). The assumptions underlying the contingency table analysis were met. The two methods of measuring bullying status were found to be significantly related, Pearsons χ^2 (9, $N = 535$) = 259.71, $p > .001$. The effect size was medium, Cramér's $V = .40$. Therefore, the two methods of measurement are related, but the strength of the relationship is not high considering the two methods are intended to measure the same construct.

Overall, 57.76% of participants were classified in the same bullying status group using both methods. Though this number is above 50%, Cohen's Kappa's coefficient, which measures agreement with chance taken into account, was only .35. According to Fleiss's guidelines (1981, in Norušis, 2012), this indicates poor agreement above chance. Examination of the correct classifications by group indicates that much of the agreement in the sample is accounted for by students who are classified as uninvolved. Ninety percent of participants who had a definitional bullying status of uninvolved also had an explicit behavioural bullying status of uninvolved. The numbers were much lower for the groups of

Table 10

Correspondence between Definitional and Explicit Behavioural Bullying Status

		Explicit Behavioural Bullying Status			
		Uninvolved	Bully	Victim	Bully/Victim
Definitional Bullying Status	Uninvolved	**207**	10	13	0
	Bully	33	**38**	3	3
	Victim	101	5	**53**	3
	Bully/Victim	17	15	23	**11**

Note: Numbers in bold represent agreement between the two methods.

participants involved in bullying. Agreement with explicit behavioural status was only 49.35% for the definitional bully group, 32.72% for the definitional victim group, and 16.67% for the definitional bully/victim group. This indicates that a large number of those identified as being involved in bullying using the definitional method are not classified the same way based on observations of their behaviour and its specific characteristics.

We next examined the types of classification errors that were made when there was disagreement between definitional and explicit behavioural methods of determining bullying status. A one-sample chi-square goodness of fit analysis showed that the proportion of participants with each bullying status using the explicit behavioural method differed significantly from the proportion found using the definitional method, χ^2 (3, N = 535) = 138.91, p > .001. A series of follow-up one-sample goodness of fit tests was conducted to examine how the proportions differed for each bullying status (Uninvolved, Bully, Victim of Bullying, and Bully/Victim). These tests showed that compared to the definitional method, in the explicit behavioural method, more participants were identified as uninvolved in bullying, χ^2 (1, N = 535) = 124.95, p > .001, and fewer participants were identified as part of the Victim of Bullying, χ^2 (1, N = 535) = 43.38, p > .001, or Bully/Victim group, χ^2 (1, N = 535) = 41.50, p > .001. The difference in the overall number of participants identified in the Bully group was not significant.

The nature of disagreements between definitional and explicit behavioural methods of determining bullying status can be further understood by examining the sensitivity and specificity of the definitional method for predicting explicit behavioural status. Sensitivity (SN), also called the true positive rate, measures the proportion of total positives identified that are correctly classified. For this study, SN was defined as the proportion of those identified as being involved in bullying through definitional questions (as part of the Bully,

Victim, or Bully/Victim group) who were also observed to be involved in bullying using the explicit behavioural method. Given this, the sensitivity was 50.5%. This indicates that approximately 50.5% of participants that were classified as involved in bullying through the definitional method were categorized as involved in bullying through the explicit behavioural method and 49.5% of those classified as involved in bullying through the definitional method were not categorized as involved in bullying with the explicit behavioural method. This indicates low sensitivity.

Specificity (SP) is also called the true negative rate. It is a measure of the proportion of all negatives that are true negatives. For this study, SP was defined as the proportion of the definitional Uninvolved group also classified as uninvolved with the explicit behavioural method of measurement. The SP rate was 90.00%. In other words, 90% of those who were identified as uninvolved in bullying using the definitional method were also classified as uninvolved using the explicit behavioural method. Therefore, overall, the definitional method of measuring bullying involvement could be said to have low sensitivity but high specificity. This indicates that the low agreement observed between the two methods is a result of a high number of false positives rather than false negatives. Many people classified as being involved in bullying with the definitional method were not classified as being involved in bullying with the explicit behavioural method.

To examine whether sensitivity and specificity varied across bullying status group, these rates were calculated separately for each bullying status group. Findings indicate that there are important differences. For the Bully group (SN= 49.35%, SP= 93.45%), Victim of Bullying group (SN= 32.72%, SP= 89.54%) and the Bully/Victim group (SN= 16.67%, SP= 98.72%), SN and SP results follow the same pattern as seen in the overall results: low sensitivity and high specificity. Again, this means that many people who were classified as

part of these groups using the definitional method were not classified as part of these groups using the explicit behavioural method. This seems to be particularly true for groups that are defined by the experience of victimization (i.e., the Victim of Bullying and the Bully/Victim group). The results for the uninvolved group, however, are reversed. For this group, SN was 90.00%, while SP was 49.51%. This indicates a relatively low number of false negatives, and a relatively high number of false positives. More people were classified as uninvolved in bullying using the explicit behavioural method. About half of people who were classified as uninvolved in bullying using the explicit behavioural method were not classified that way using the definitional method. Overall, the definitional method appears to be far more inclusive than the explicit behavioural method.

Summary and conclusions.

In this section, the agreement between two ways of measuring bullying involvement was examined. Definitional bullying status was calculated based on reporter's nominations of classmates who bully and are victimized by bullying, and explicit behavioural bullying status was calculated based on ratings of aggressive behaviour and its characteristics. The explicit behavioural and definitional methods of measuring bullying status were significantly related; however, the strength of this relationship is not as strong as expected. Only 58% of participants were categorized the same way using each method and the Kappa statistic of .35 indicates weak agreement. More bullying involvement was captured using the definitional measurement style, especially more victimization by bullying. When assessing the sensitivity and specificity of the definitional method of bullying for replicating the categorical labels assigned using the explicit behavioural method, we found that the definitional method generally had poor sensitivity but good specificity. This means that there were many false positives, but few false negatives. When we examined sensitivity and specificity within

specific definitional bullying statuses, this pattern held true for the Bully group, the Victim of Bullying Group, and the Bully/Victim group, but was reversed for the Uninvolved group. Definitional measures of the Uninvolved group showed poor sensitivity but good specificity. This was because fewer people were categorized as uninvolved using the definitional method than the explicit behavioural method.

One important implication of these results is that they show that many people who were identified as being involved in bullying through definitional methods, especially many people who were identified as being victimized by bullying, do not have explicit behavioural bullying results that indicate that their aggression or victimization actually meets the criteria for bullying involvement.

Predicting definitional bullying status with behaviourally measured definitional characteristics.

The goal of this set of analyses was to examine the extent to which behaviourally measured characteristics of aggression included in the definition of bullying predict definitional bullying status. This showed us to what extent, if any, the definitional characteristics were taken into account by participants when completing the definitional measure. Analyses consisted of two sequential regression analyses, one examining whether characteristics of aggression predicted identification as a perpetrator of bullying and the other examining whether characteristics of victimization predicted identification as a victim of bullying. Three questions of interest were addressed in these analyses: 1. Do the definitional characteristics of bullying predict definitional bully or victim of bullying status above and beyond the presence of aggressive behaviour or victimization by peers? 2. Is meeting full definitional criteria for bullying meaningful in that the presence of all three definitional criteria predicts definitional bullying status above and beyond the presence of the

individual criteria? 3. Are each of the three definitional characteristics of bullying independently significant for predicting definitional bullying status?

Predicting definitional bully status.

The first sequential multiple regression examined how well behavioural characteristics of aggression predicted whether participants were identified as perpetrating bullying using the definitional method. The outcome variable was the number of nominations for bullying behaviour that an individual received from their classmates using the definitional method (i.e., how many people identified this person as a bully when given a definition of bullying and asked to identify bullying relationships). In each regression, predictors were entered in four steps (see Table 11). The first step contained a control variable: classroom size (M = 19.14, SD = 3.79). This is to control for differences in number of nominations for aggression caused by differences in the size of the nominating pool. The second step also contained one variable: the number of times an individual was nominated as being the perpetrator of aggression (M = 2.42, SD = 2.90). This step indicated

Table 11

Model for Sequential Regression Examining the Relationship between the Characteristics of Aggressive Behaviour and Definitional Bully Status

Step	Predictor	Description
1	Class size	Number of people in the participant's classroom who participated in the study (i.e., number of potential nominators)
2	Aggression	Number of nominations for aggressive behaviour
3	Intention	Number of nominations for aggressive behaviour being intentionally harmful
	Power Imbalance	Number of nominations for aggressive behaviour occurring in the context of a power imbalance
	Repetition	Number of nominations for aggressive behaviour being repeated
4	All Three Characteristics	Number of nominators that observed aggression and all three definitional characteristics measured

to what extent aggression alone is related to definitional bullying status. The third step contained three variables, corresponding to the number of nominations for each of the three characteristics of aggressive behaviour that would indicate that aggression meets the criteria for bullying: aggression is repeated (M = 1.25, SD = 1.97), aggression is intentionally hurtful (M = 1.3, SD = 2.00), and aggression occurs in the context of a power imbalance (M = 1.25, SD = 1.86). This step indicated to what extent the specific characteristics of aggressive behaviour contained within the definition of bullying add to our ability to predict the outcomes of those participants who are aggressive. The fourth and final set of predictors included one variable: the number of times that a participant's aggression was nominated as having all three characteristics that define bullying (M = 0.86, SD = 1.44). According to the definition, all three characteristics must be present in order for the behaviour to be bullying, so this step indicates to what extent, if any, the fact that the an individual is aggressive in a way that meets the criteria for bullying adds to our ability to predict student outcomes.

Table 12

Correlation among Predictors for the Regression Predicting Definitional Bully Status from Characteristics of Aggression

	Number of Students in Classroom	Aggression	Repeated	Intentionally Hurtful	Power Imbalance	All Three Characteristics
Aggression	.11*	–	–	–	–	–
Repeated	.05	.89***	–	–	–	–
Intentionally Hurtful	.09*	.88***	.88***	–	–	–
Power Imbalance	.04	.84***	.83***	.85***	–	–
All Three Characteristics	.033	.76***	.81***	.84***	.91***	–
Nominations for Bully Status	.17***	.79***	.70***	.74***	.73***	.73***

Note: * $p < .05$ *** $p < .001$

Our predictors related to the perpetration aggression were significantly related (see Table 12 for details). The size of these correlations indicates that multicollinearity among predictors is present. To avoid problems in interpretation associated with multicollinearity, we followed guidelines for interpreting the results of multiple regression laid out by Kraha, Turner, Nimon, Zientek, and Henson (2012) and Nathans, Oswald, and Nimon, (2012). Variable importance was examined through an analysis of standardized coefficients (Beta weights), structure coefficients, all-possible-subsets regression, commonality analysis, and dominance weights. We felt it was important to do this in order to be able to examine the importance of all variables included as predictors (rather than eliminating potentially redundant predictors) because all predictor variables are included in the definition of bullying (besides our control variable).

Before interpretation, this regression was examined for possible violations of the assumptions of regression analysis. For each regression, residuals were plotted and examined for problems with normality, linearity, and homoscedasticity. The residuals were positively skewed; however, variables were not transformed given that they are meaningful in their current state and multiple regression is robust to moderate departures from normality (Norušis, 2012). No problems with linearity and homoscedasticity that would invalidate the analyses were found. Cook's Distance and leverage were charted and examined. The maximum leverage value observed was 0.17 and the largest value of Cook's Distance observed was 0.32. This indicates that no one data point was overly influential (Norušis, 2012; Tabachnick and Fidel, 2004). Therefore, all cases were retained for the analysis and data was not truncated in any way.

Thirteen outliers in the solution were observed in this regression. In each case, the number of nominations for bullying observed was higher than the predicted value. Our

Table 13

Model Level Results of Stepwise Regression Predicting Definitional Bully Status with Characteristics of Aggression

Outcome	*F*	(df)	R^2	*F* change	ΔR^2
Step 1	15.89***	1, 533	.03		
Step 2	463.61***	2, 532	.64	885.00***	.61
Step 3	207.52***	5, 529	.66	14.05***	.03
Step 4	201.86***	6, 528	.70	59.27***	.03

Note: * $p < .05$ *** $p < .001$

model does not fit well for these 13 participants who were nominated for bullying more often than would be expected, but these participants make up a small percentage of the overall sample (2.5%). The cases were retained in the analysis to avoid artificially retrofitting the regression equations to our data.

Statistics related to this analysis can be seen in Table 13. The model was significant at each level, and each step in the model significantly added to our predictive ability. With all variables entered, we are able to account for 70% of the variance in number of nominations for definitional bully status. With respect to our specific questions of interest, we see that adding information about the characteristics of aggression does significantly increase our ability to predict nominations for definitional bully status, and so does the presence of all three characteristics. However, taken together, these four variables only account for 6% of the variance in number of nominations, whereas nominations for aggressive status, on its own, accounts for 61% of the variance. So while the behavioral characteristics of aggression contained within the definition of bullying do predict nominations for definitional bullying status, their addition in this regression accounts for only a small proportion of the total amount of variance in aggression predicted.

We now turn our attention to statistics on the importance of individual predictors. Statistics related to the importance of individual predictors can be seen in Table 14.

Table 14

Individual Predictor Results of Stepwise Regression Predicting Definitional Bully Status with Characteristics of Aggression

	Number of Nominators	Aggression	Intentional to Harmful	Power Imbalance	Repeated	All Three Characteristics
β	.088*	0.796*	.041	-.152*	-.345*	.511*
r_s	0.204	0.949	0.888	0.876	0.836	0.873
r_s^2	0.042	0.901	0.789	0.767	0.699	0.762
r	0.170	0.792	0.741	0.731	0.698	0.729
R^2	0.029	0.627	0.549	0.534	0.487	0.531
Unique	0.008	0.087	0.000	0.003	0.017	0.034
Common	0.029	0.627	0.549	0.534	0.487	0.531
General Dominance Weights	0.015	0.223	0.147	0.141	0.126	0.154
Relative Importance Weights	0.017	0.189	0.126	0.121	0.105	0.138

Note: Total R^2= the total amount of variance predicted by the full model. The primary predictor suggested by a method is underlined. β = the standardized correlation coefficient (Beta weight). * = significant at α - .05. r_s = structure coefficient. r_s^2 = squared structure coefficient r = bivariate correlation coefficient. R^2 = squared bivariate correlation coefficient. Unique = proportion of criterion variance explained uniquely by the predictor. Common = proportion of criterion variance explained by the predictor that is also explained by one or more other predictors. Unique + Common = R^2.

A clear finding from these results is that much of the variance explained in our regression is shared among multiple predictors. Given the high correlations observed among predictors, this is not unexpected. This finding can be seen most clearly in the results of the commonality analysis performed. The largest chunk of the variance explained (61.76%) is common to all of the predictor variables except our control variable. The four predictors related to the characteristics of bullying (intention, power imbalance, repetition, and all three characteristics) uniquely share 3.99% of the variance explained. This means that most of the variance accounted for in this regression could be explained by any one of our aggression-related variables, including the presence of aggression alone. Also, the small amount of variance that is accounted for uniquely by the characteristics of bullying is mostly shared

among all of these characteristics variables. The all-subsets-regression supports these conclusions as well. The absolute best model we have for predicting nominations as a bully using the definitional method includes all six of the predictors in our analysis (R^2 value = .696). Because of shared variance, we could leave any one individual predictor out of the analysis except for aggression and still predict an R^2 value >.662. The model could also be reduced to a single predictor (aggression) with a loss of only less than 10% of the total variance accounted for (R^2 value = .627).

Because of the large amount of shared variance among predictors, none of the individual definitional characteristics of bullying emerge as individually important, despite being highly correlated with the outcome variable. The all three characteristics variable uniquely accounts for roughly 3% of the variance in definitional bullying nominations, and repetition almost 2%, whereas power imbalance and intention to harm account for almost no unique variance. The most important individual predictor is the presence of aggression. It is the most closely related to actual and predicted bullying status, has high general dominance and relative importance weights, and accounts for more variance (both unique and combined) than any other variable.

Predicting definitional victim of bullying status.

The second sequential multiple regression examines how well behavioural characteristics of aggression predict whether participants are identified as victims of bullying using the definitional method. The outcome variable is the number of nominations for being a victim of bullying that an individual received from their classmates using the definitional method (i.e., how many people identified this person as a victim of bullying when given a definition of bullying and asked to identify bullying relationships). In each regression, predictors were entered in four steps (see Table 15). The first step contained a control

Table 15

Model for Sequential Regression Examining the Relationship between the Characteristics of Victimization and Definitional Victim of Bullying Status

Step	Predictor	Description
1	Class size	Number of people in the participant's classroom who participated in the study (i.e., number of potential nominators)
2	Victimization	Number of nominations for victimization
3	Intention	Number of nominations for victimization being intentionally harmful
	Power Imbalance	Number of nominations for victimization occurring in the context of a power imbalance
	Repetition	Number of nominations for victimization being repeated
4	All Three Characteristics	Number of nominators that victimization and endorsed all three definitional characteristics measured

variable: classroom size (M = 19.14, SD = 3.79). This is to control for differences in number of nominations for victimization caused by differences in the size of the nominating pool. The second step also contained one variable: the number of times an individual was nominated as being the victim of aggression (M = 2.42, SD = 2.90). This step indicated to what extent victimization alone is related to student functioning. The third step contained three variables, corresponding to the number of nominations for each of the three characteristics of victimization that would indicate that aggression meets the criteria for victimization by bullying: aggression is repeated (M = 1.25, SD = 1.97), aggression is intentionally hurtful (M = 1.3, SD = 2.00), and aggression occurs in the context of a power imbalance (M = 1.25, SD = 1.86). This step indicated to what extent the specific characteristics of aggressive behaviour contained within the definition of bullying add to our ability to predict the outcomes of those participants who experience victimization. The fourth and final set of predictors includes one variable: the number of times that a participant's victimization was nominated as having all three characteristics that define bullying (M = 0.86, SD = 1.44). According to the definition, all three characteristics must be present in

order for the behaviour to be bullying, so this step indicated to what extent, if any, the fact that the an individual is victimized in a way that meets the criteria for bullying adds to our ability to predict student outcomes.

Our predictors related to victimization by aggression were significantly related (see Table 16 for details). As before, because of the likelihood of multicollinearity, we interpreted results related to the importance of individual predictors thorough a number of different methods. Before interpretation, this regression was examined for possible violations of the assumptions of regression analysis. For each regression, residuals were plotted and examined for problems with normality, linearity, and homoscedasticity. The residuals were positively skewed; however, variables were not transformed given that they are meaningful in their current state and multiple regression is robust to moderate departures from normality (Norušis, 2012). No problems with linearity and homoscedasticity that would invalidate the

Table 16

Correlation among Predictors for the Regression Predicting Definitional Victim of Bullying Status from Characteristics of Victimization

	Number of Students in Classroom	Nomination as victim of aggression	Nominations indicating that aggression is Repeated	Nominations indicating that aggression is Intentionally Hurtful	Nominations indicating that aggression is Power Imbalance	Nominations indicating that aggression has All Three Characteristics
Victimization	.17***	–	–	–	–	–
Repeated	.04	.84***	–	–	–	–
Intentionally Hurtful	.16***	.83***	.85***	–	–	–
Power Imbalance	.07*	.84***	.81***	.84***	–	–
All Three Characteristics	.07	.71***	.80***	.83***	.90***	–
Nominations as definitional victim of bullying	.20***	.68***	.64***	.71***	.70***	.73***

Note: * $p < .05$ *** $p < .001$

analyses were found. Cook's Distance and Leverage were also charted and examined. The maximum Leverage value observed was 0.10 and the largest value of Cook's Distance observed was 0.18. This indicated that no one data point was overly influential (Norušis, 2012; Tabachnick and Fidel, 2007). Therefore, all cases were retained for the analysis and data was not truncated in any way.

Three outliers in the solution were observed in this regression. In each case, the number of nominations for victimization observed was higher than the predicted value. Our model does not fit well for these three participants who were nominated for bullying more often than would be expected, but these participants make up a small percentage of the overall sample (0.6%). The cases were retained in the analysis to avoid artificially retrofitting the regression equations to our data.

Model-level statistics related to this analysis (or these analyses) can be seen in Table 17. The model was significant at each level, and each step in the model significantly added to our predictive ability. With all variables entered, we are able to account for 60% of the variance in number of nominations for definitional victim status. With respect to our specific questions of interest, we see that adding information about the characteristics of aggression does significantly increase our ability to predict nominations for definitional victim status, and so does the presence of all three characteristics. Taken together, these four variables

Table 17

Model Level Results of Stepwise Regression Predicting Definitional Victim of Bullying Status with Characteristics of Victimization

	F	(df)	R^2	*F* change	ΔR^2
Step 1	22.25***	1, 533	.04		
Step 2	231.75***	2, 532	.47	423.61***	.46
Step 3	133.49***	5, 529	.56	36.80***	.09
Step 4	132.30***	6, 528	.60	56.42***	.04

Note: * $p < .05$ *** $p < .001$

account for 13% of the variance in number of nominations. Nominations for victimization by aggression, on its own, accounts for 46% of the variance.

We now turn our attention to statistics related to the importance of individual predictors, which can be seen in Table 18. As with the results related to definitional measures of bullying, most of the variance explained in definitional nominations for victimization by bullying is shared across variables. Given the high correlations observed among predictors, this is not unexpected. This can be seen most clearly in the results of the commonality analysis performed. The largest chunk of the variance explained (57.90%) is common to all of the predictor variables except our control variable. The four predictors related to the characteristics of bullying (intention, power imbalance, repetition, and all three characteristics) uniquely explain 3.49% of the variance. Power imbalance and the all three

Table 18

Individual Predictor Results of Stepwise Regression Predicting Definitional Victim of Bullying Status with Characteristics of Victimization

	Number of Nominators	Victimization	Intentional to Harmful	Power Imbalance	Repeated	All Three Characteristics
β	0.097*	0.316*	0.158*	-0.042	-0.137*	0.514*
r_s	0.258	0.874	0.912	0.900	0.827	0.937
r_s^2	0.067	0.764	0.832	0.810	0.684	0.878
R	0.200	0.677	0.707	0.697	0.641	0.726
R^2	0.040	0.458	0.500	0.486	0.411	0.527
Unique	0.009	0.021	0.004	0.000	0.004	0.042
Common	0.032	0.437	0.495	0.486	0.407	0.485
General Dominance Weights	0.019	0.131	0.141	0.134	0.103	0.174
Relative Importance Weights	0.021	0.114	0.115	0.112	0.085	0.153

Note: Total R^2= the total amount of variance predicted by the full model. The primary predictor suggested by a method is underlined. β = the standardized correlation coefficient (Beta weight). * = significant at α - .05. r_s = structure coefficient. r_s^2 = squared structure coefficient r = bivariate correlation coefficient. R^2 = squared bivariate correlation coefficient. Unique = proportion of criterion variance explained uniquely by the predictor. Common = proportion of criterion variance explained by the predictor that is also explained by one or more other predictors. Unique + Common = R^2.

characteristics variable share a significant chunk of the variance (4.11%) as do intention to harm, power imbalance, and the all three characteristics variable (4.89%). There is also some significant suppression of error variance happening, particularly between victimization and the all three characteristics variable (-2.11%).

The all-subsets regression supports these conclusions as well. The absolute best model we have for predicting nominations as a bully using the definitional method includes all six of the predictors in our analysis (R^2 value = .600). Because of shared variance, we could leave any one individual predictor out of the analysis except for the all three characteristics variable and still predict an R^2 value >.57. The model could be reduced to a single predictor (all three characteristics) with a loss of only 12% of the total variance explained (R^2 value = .527).

Because of the large amount of shared variance among predictors, it is difficult to conclude that any of the individual predictors is individually important. The most individually important variable, however, is the all three characteristics variable. It is the most closely related to actual and predicted bullying status, has high general dominance and relative importance weights, and accounts for more unique variance than any other variable. In terms of total variance accounted for (unique + common variance), however, victimization, intention to harm, power imbalance and repetition are remarkably similar. None emerges clearly more important than the others. It does appear, however, that unlike in our results related to nominations for bullying, something shared about the characteristics of bullying is significantly useful for improving our ability to predict definitional victim of bullying status.

Summary

This analysis addressed three questions of interest:

1. Does information about the characteristics of the aggression or victimization predict nominations for definitional bully or victim status above and beyond the presence of aggressive behaviour or victimization by peers?
2. Does the presence of all three characteristics contained within the definition of bullying have additional significance in predicting nominations for involvement in bullying?
3. Are each of the three characteristics of bullying contained within the definition of bullying independently significant for predicting definitional bullying status?

I addressed these questions through two multiple regression analyses, one which addressed aggression and bullying nominations and the other which addressed victimization and victim nominations. For both aggression and victimization, our models predicted definitional nominations. Behavioural information about aggression predicted 70% of the variance in definitional bullying nominations, and behavioural information about victimization predicted 60% of definitional victimization nominations. In both analyses, the presence of aggression or victimization alone accounted for a large proportion of the variance predicted (64% and 47% respectively). Also, in both regressions, the predictor variables were highly correlated, so much of the total variance predicted in the final model was shared among multiple predictors. This meant that the individual characteristics variables did not emerge as uniquely important predictors, despite being highly related to the outcome variable. The three behavioural characteristics accounted for 2% of the variance in bully nominations and 9% of the variance in victim of bullying nominations. The all three characteristics variable accounted for 4% of the variance in bully nominations and 4% of the variance in victim of bullying nominations. This implies that definitional measures of bullying do measure bullying as defined to some extent, although bullying measured this

way may not be as distinct from other aggression as would be hoped. The amount of variance that is uniquely explained by definitional characteristics is much less than could be explained by the presence of aggression or victimization alone. This result is more pronounced for perpetrators of bullying. For victims of bullying, the amount of variance explained by behavioural characteristics is proportionally larger, even though, again, much of the variance explained by these characteristics is shared among them. This pattern can be seen in which variable emerged as the most important individual predictor in each analysis. The presence of aggression was the best individual predictor for definitional bullying status, whereas the all three characteristics variable was the best individual predictor for definitional victim of bullying status. These variables could be interpreted as the best summary description of what participants are taking into account when making definitional nominations, at least terms of the amount of what participants are taking into account that we can quantify. Remember that 30% of the variance in definitional bully nominations and 40% of the variance in definitional victim of bullying nominations is unexplained by the behaviourally measured definitional characteristics of aggression.

Differences between participants identified as involved in bullying using different methods.

This section of our results contains analyses investigating to what extent people who are identified as being involved in bullying using the explicit behavioural method differ from participants identified as being involved in bullying using the definitional method in terms of social and individual functioning. The purpose of this analysis was to determine whether the differences in who is identified are meaningful in that the participants identified represent different groups. Potential areas of differenced explored include participant social functioning (popularity, power, self-perceived social rank, submissive behaviour, sense of

acceptance) and individual functioning (negative affect, positive affect, sense of connection to school, beliefs about aggression, conduct problems and academic functioning). These analyses were run separately for participants who bullied others, and participants who were victimized by bullying. The number of participants identified with Bully and Victim of Bullying status using different measurement approaches is shown in Table 19.

Participants who perpetrate bullying.

Social functioning.

To determine whether students identified as part of the bullying group using the definitional method only were different from the students identified as part of the bullying group using the explicit behavioural method in terms of their social functioning, I completed a one-way MANOVA analysis. I chose to use a MANOVA analysis, despite the fact that the dependent variables were not highly correlated (absolute values of *r* ranged from -.14 to .85), as a way of controlling for Type 1 error. The grouping variable for this analysis is based on how an individual was nominated to be a bully. Those whose behaviour was nominated as meeting the criteria for bullying using our explicit behavioural measurement method are one group, and those who were identified as part of the bully group using the definitional method, but did not meet behavioural criteria for bullying, were in the other group. The purpose of this analysis was to determine whether meaningful differences exist between

Table 19

Number of Participants Identified as Involved in Bullying with Different Measurement Strategies

		Bully	Victim of Bullying
Method Identified	Behavioural Involvement	85	109
	Definitional Involvement only	76	138

these two groups in terms of social functioning. The outcome variables were popularity, power, self-perceived social rank, submissive behaviour, sense of acceptance. No significant differences were found between the two bullying involvement groups on the dependent measures, Pillai's Trace = .035, $F(5, 132) = 0.97$, $p = .44$, partial $\eta^2 = .035$. Therefore, it does not appear that there are significant differences between those who meet the behavioural criteria for bullying and those who are identified as involved in bullying despite not meeting the behavioural criteria in terms of social functioning.

Individual functioning.

To determine whether students identified as part of the bullying group using the definitional method only were different from the students identified as part of the bullying group using the explicit behavioural method in terms of their individual functioning, I completed a one-way MANOVA analysis. I chose to use a MANOVA analysis, despite the fact that the dependent variables were not highly correlated (absolute values of *r* ranged from .08 to .63), as a way of controlling for Type 1 error. The grouping variable for this analysis is based on how an individual was nominated to be a bully. Those whose behaviour was nominated as meeting the criteria for bullying using our explicit behavioural measurement method were one group, and those who were identified as part of the bully group using the definitional method, but did not meet behavioural criteria for bullying according to the explicit behavioural method, were in the other group. The purpose of this analysis was to determine whether meaningful differences exist between these two groups in terms of individual functioning. The outcome variables are negative affect, positive affect, sense of connection to school, beliefs about aggression, conduct problems and academic functioning. No significant differences were found between the two bullying involvement groups on the dependent measures, Pillai's Trace = .091, $F(6, 118) = 1.98$, $p = .074$, partial $\eta^2 = .09$.

Therefore, it does not appear that there are significant differences between those who meet the behavioural criteria for bullying and those who are identified as involved in bullying, despite not meeting the behavioural criteria in terms of individual functioning.

Participants who are victims of bullying.

Social functioning.

To determine whether students identified as part of the victim of bullying group using the definitional method only were different from the students identified as part of the victim of bullying group using the explicit behavioural method in terms of their social functioning, I completed a one-way MANOVA analysis. I chose to use a MANOVA analysis as a way of controlling for Type 1 error, despite the fact that the dependent variables were not highly correlated (absolute values of *r* ranged from -.14 to .85). The grouping variable for this analysis is based on how an individual was nominated to be a victim of bullying. Those whose behaviour was nominated as meeting the criteria for victimization by bullying using our explicit behavioural measurement method were one group, and those who were identified as part of the victim of bullying group using the definitional method, but did not meet behavioural criteria for victimization by bullying according to the explicit behavioural method, were in the other group. The purpose of this analysis was to determine whether meaningful differences exist between these two groups in terms of social functioning. The outcome variables were popularity, power, self-perceived social rank, submissive behaviour, sense of acceptance.

Significant differences were found between the two bullying involvement groups on the dependent measures, Pillai's Trace = .056, $F(5,212) = 2.52$, $p = .031$. The effect size was small, partial $\eta^2 = .06$. Given this significant overall finding showing group differences, a series of post-hoc tests (one-way AVOVAs) were performed to understand the source of this

difference. No correction was made to the alpha level of these tests given the overall significant result. As displayed in Table 20, there were significant group differences on only one variable, popularity. The group who met the behaviour criteria for victimization by bullying had an average popularity score of -0.55 (*SD* = .95) and those who were identified as victims of bullying but did not meet the criteria for bullying had an average popularity score of -0.16 (*SD* = .92). Therefore, those students whose victimization met the criteria for bullying are less popular than other students who were nominated to be victims of bullying using the definitional method, but do not meet the behavioural criteria. There are no other significant differences between groups.

Individual functioning.

To determine whether students identified as part of the victim of bullying group using the definitional method only were different from the students identified as part of the victim of bullying group using the explicit behavioural method in terms of their individual functioning, I completed a one-way MANOVA analysis. I chose to use a MANOVA analysis

Table 20

Follow-up One-Way ANOVAs Examining Differences in Social Functioning between Victims of Bullying Identified through Different Measurement Strategies

Variable	*F*	Df	*P*	Partial Eta Squared
Power	2.74	1, 216	.10	.01
Popularity	9.46	1, 216	.002	.04
Social Rank (ASCS)*	1.20	1, 216	.27	.01
Submissive Behaviour (ASBS)	.93	1, 216	.34	.004
Social Acceptance (SAIS)	1.34	1, 216	.25	.01

* Indicates that Levene's test of Equality of Error Variance was significant, indicating significant variation in variance scores.

as a way of controlling for Type 1 error, despite the fact that the dependent variables were not highly correlated (absolute values of *r* ranged from .08 to .63). The grouping variable for this analysis was based on how an individual was nominated to be a victim of bullying. Those whose behaviour was nominated as meeting the criteria for victims of bullying using our explicit behavioural measurement method were one group, and those who were identified as part of the victim of bullying group using the definitional method, but did not meet behavioural criteria for victimization by bullying according to the explicit behavioural method, were in the other group. The purpose of this analysis was to determine whether meaningful differences exist between these two groups in terms of individual functioning. The outcome variables are negative affect, positive affect, sense of connection to school, beliefs about aggression, conduct problems and academic functioning.

Significant differences were found between the two bullying involvement groups on the dependent measures, Pillai's Trace = .063, $F(6,190) = 2.13$, $p = .05$. The effect size was small, partial $\eta^2 = .06$. Given this significant overall finding showing group differences, a series of post-hoc tests (one way AVOVAs) were performed to understand the source of this difference. No correction was made to the alpha level of these tests given the overall significant result. As displayed in Table 21, there were significant group differences on only one variable, negative affect. The group who met the behaviour criteria for victimization by bullying had an average negative affect score of 1.92 (*SD* = .93) and those who were identified as victims of bullying but did not meet the criteria for bullying had an average negative affect score of 1.62 (*SD* = .67). Therefore, those students whose victimization met the criteria for bullying have more negative affect than other students who were identified as victims of bullying. There are no other significant differences between groups, but the difference between groups in sense of school membership approaches significance.

Table 21

Follow-up One-Way ANOVAs Examining Differences in Individual Functioning between Victims of Bullying Identified through Different Measurement Strategies

Variable	*F*	Df	*P*	Partial Eta Squared
Negative Affect	6.87	1, 197	.01	.03
Positive Affect	0.00	1, 197	.953	.00
Sense of School Membership	3.66	1, 197	.057	.02
Beliefs about Aggression	0.07	1, 197	.80	.000
Conduct Problems	2.21	1, 197	.14	.01

* Indicates that Levene's test of Equality of Error Variance was significant, indicating significant variation in variance scores.

The group who met the behaviour criteria for victimization by bullying had an average school membership score of 3.46 (*SD* = .84) and those who were identified as victims of bullying but did not meet the criteria for bullying had an average popularity score of 3.69 (*SD* = .84). Those students whose victimization met the criteria for bullying have a trend towards less sense of connection to the school than other students who are identified as victims of bullying.

Summary

In this section, I examined whether participants who were identified as being involved in bullying by different measurement methods differ in terms of social and individual functioning. Participants who were identified through the explicit behavioural method explicitly met criteria for bullying involvement (in either a bully or victim role) according to peer ratings of their behaviour. These participants were compared to participants identified as being involved in bullying (in either a bully or victim role) through the definitional method, but whose scores of the explicit behavioural method indicated that their behaviour did not meet the full criteria for bullying. Few differences between groups

were found. For students who were identified as bullies, there were no significant differences between groups. For students identified as victims of bullies, there were two significant differences. Participants identified as victims of bullying using the explicit behavioural method were less popular and had more negative affect than students identified as being victimized by bullying using the definitional method only.

Chapter 6. Discussion

The purpose of the present research project is to empirically explore the validity of the bullying construct. Bullying is defined as a subset of aggressive behaviour characterised by repetition, intention to harm, and a power imbalance between the person who perpetrates the aggression and the person victimized (Olweus, 2010). This conceptualization of bullying, though widely used and accepted, has been criticised on the grounds that it lacks construct validity. This paper attempts to address this issue by providing an empirical investigation into the usefulness of the definition of bullying and our ability to capture behaviour that meets that definition though common measurement techniques. This paper has two major purposes:

3. To examine the validity of the definition of bullying in terms of its ability to predict student functioning and identify a distinct group.
4. To examine the validity of widely used measurement techniques in identifying aggression that meets this definition of bullying.

For ease of understanding, results related to these two goals are presented as separate studies, although they are all part of the same larger project. This discussion will begin by addressing each study separately, and then end with a general discussion on the project as a whole.

Study One

Review of study results.

This study addresses the validity of the bullying construct by examining whether the current, widely used definition of bullying is useful for predicting student functioning and identifying a meaningful subset of aggressive behaviour. It also addresses whether each of the three behavioural characteristics contained within the definition contributes individually to the overall finding.

Overall, the results of this study are mixed in terms of their support for the construct validity of the current definition of bullying. Two findings support the construct validity of bullying as defined. Firstly, consistent with Hypothesis 1a, the results of this study show that information about the characteristics of aggressive behaviour that define bullying allows for stronger prediction of student outcomes in many areas than information about the presence of aggression or victimization alone. In addition, the amount of variance accounted for by the definitional characteristics of bullying was not insignificant relative to the total amount of variance predicted. This is evidence for discriminative validity. Secondly, consistent with Hypothesis 1d, the characteristics of bullying predicted worse individual and social functioning for those students who were victimized. This demonstrates predictive validity, in that this is what would be expected if bullying is a particularly harmful form of aggression. With regards to the perpetration of aggressive behaviour, the results of this study indicate an interesting pattern. Participants' whose aggressive behaviour had at least some of the characteristics of bullying reported more positive indicators for social dominance than their peers, but at the same time, they reported worse indicators of individual functioning than those who were not aggressive or whose aggression was not repeated, intentionally harmful, or occurring in the context of a power imbalance. This is again evidence of a distinct group.

Other results of this study were not supportive of the construct validity of bullying. Contrary to Hypothesis 1b, the results of this study suggest that there is no additional predictive ability gained by requiring that all three characteristics that define bullying be present. This suggests that meeting the full definitional criteria for bullying is not particularly meaningful. In addition, contrary to Hypothesis 1c, each of the three behavioural criteria contained within the definition of bullying was not independently significant for predicting participant functioning. In this study, the three defining characteristics of bullying were very highly correlated and appeared to be somewhat redundant measures of a single underlying factor rather than independently important characteristics with an additive effect. This suggests that the current definition of bullying consists of indicators that bullying has occurred rather than essential components of the construct.

The relationship between the characteristics of bullying and individual functioning.

This study focused in particular on examining the relationship between bullying involvement and variables related to social dominance. This focus was chosen because social dominance theories of bullying posit that bullying is dominance-related behaviour (Salmivalli, 2010; Volk et al., 2012). The findings of this study support this theory. In general, the definitional characteristics of bullying accounted for more variance in variables related to social status than individual functioning. This suggests that bullying may have an especially strong relationship with these types of outcomes. Even if this finding on the relative strength of the relationship is an artefact of our use of peer-report methods (since social outcomes are more strongly related to peer- than self-reported bullying; Bouman et al., 2012; Scholte et al., 2013), our study does demonstrate a clear link between the definitional

characteristics of bullying and victimization by bullying and participant dominance status. This supports the possibility that bullying could be dominance related behaviour.

The types of functioning that are associated with the definitional characteristics of bullying also support the conceptualization of bullying as dominance related behaviour. The definitional characteristics of bullying predict indicators of low social status among participants who are victimized and indicators of high social status among those who bully. While the direction of the relationship between being the behavioural characteristics of victimization and low social status cannot be determined by the results of this study, the possibility of a bidirectional relationship should be explored. Low social status may lead to an individual being bullying, as per social dominance theory, which suggests that adaptive or dominance-related aggression is typically aimed at lower-status individuals against whom the aggressor is confident that they can prevail (Hawley, 1999, 2003). The possibility that victimization by bullying could lead to low social status also seems reasonable, and suggests a way that bullying may be unique from other forms of aggression. If a unique causal relationship between victimization by bullying and low social status can be demonstrated, it would suggest that the social consequences of victimization by bullying might be the mechanism though which bullying becomes particularly harmful. Low-status individuals have less access to valued resources (Hawley, 1999) fewer friends, less social support (K. K. Thomas & Bowker, 2013), and less agency in relationships (Sercombe & Donnelly, 2013). Social outcomes such as these would likely have an impact on individual's emotional and educational functioning. If so an understanding of bullying as a form of aggression that is particularly harmful to the victim's social standing may also be able to explain the other, non-social outcomes that have been shown to be associated with victimization by bullying.

For this reason, exploring the social outcomes of victims of bullying, and the uniqueness of this relationship, should become a future research priority.

The finding that the definitional characteristics of bullying predict indicators of high social status among those who are aggressive is also consistent with a social dominance theory of bullying. It suggests that the characteristics that define bullying are associated with the type of aggression that is effectively used to achieve social dominance. This finding is unique in specifically linking repetition, intention to harm, and power imbalance in aggression to dominant social status. Other studies looking at the different outcomes for bullying and peer aggression focused on victims of bullying only and did not examine the functioning of those who bully (Felix et al., 2011; Hunter et al., 2007; Ybarra et al., 2014; You et al., 2008). The finding is consistent, however, with evidence that individuals identified as bullies have higher than average social standing (de Bruyn et al., 2009; Peeters et al., 2009; Pellegrini & Long, 2002; Rodkin & Berger, 2008; Salmivalli, 2010; Vaillancourt et al., 2003) and the conceptualization of individuals who bully as part of the socially skilled subgroup of aggressors. At the same time, however, in this study the definitional characteristics of bullying also predict indications of problematic individual functioning. This is not consistent with previous research indicating that the social advantages of bullying behaviour for some individuals comes with few costs (Reijntjes, Vermande, Olthof, et al., 2013; Volk et al., 2012). Our results suggests that even when aggressors are successful at achieving dominance goals through aggression, there may be a cost to pay in terms of one's perception of their social acceptance, one's overall experience at school, and one's emotional experiences. This is not consistent with social dominance theories of bullying, unless individuals are motivated enough to acquire dominance that they are willing to sacrifice individual functioning. A more likely explanation for this contradictory seeming set of

results may be variability within the bullying group. As discussed in the literature review, there is generally variability in the popularity of adolescents identified as bullies (Rodkin & Berger, 2008; Vaillancourt et al., 2003). That is true in this study as well. This pattern of findings could either be a result of true variability in the social outcomes experienced by those who bully, or of measurement error which causes true bullies (assumed to be those who successfully achieve social dominance and experiences few negative individual consequences) and other types of aggressors (including those who use aggression unskilfully and primarily reactively) to be grouped together. This second interpretation would be more consistent with social dominance theory and is plausible considering the difficulties in measuring bullying (see literature review). Though the behavioural characteristics contained within the definition of bullying do predict indicators of social dominance, they do not appear to be sufficient, as currently measured, to distinguish between low status and high status groups of aggressors. The question now becomes whether this distinction should be made. Unfortunately there exists confusion on this point. While Olweus intended for the bullying group to be made up of the "winners" of aggressive episodes (Olweus, 2010), ensuring that this has occurred has not been a strong focus of research to this point. Bullying has been defined by form rather than function or results. This can be seen in the existence of the bully/victim group, who are considered bullies despite the fact that they are by definition sometimes the "losers" of aggressive incidents and that they have outcomes that would be more consistent with the low status subgroup of aggressors. Conceptual clarity in the bullying construct is needed to determine whether the definitional criteria need to be refined in an attempt to capture only socially dominant aggressors.

A final note related to the types of functioning predicted by the definitional characteristics of bullying in this study. Overall, the results are clear that the characteristics

of bullying are, as a group, related to participant functioning in a way that is distinct from the relationship between these measures of functioning and the simple presence of aggression or victimization. In this case distinct means reliably distinguishable. There is far less evidence, however, that the relationships found are unique, meaning singular or unlike anything else. This is important because bullying is studied separately from other forms of aggression because it is believed to be associated with worse outcomes (Olweus, 2013), but it is also studied separately because it is believed to be conceptually unique. There are differences of opinion among researchers about what makes bullying unique. Some would argue that bullying is unique because it is difficult to escape or defend oneself (Olweus, 2010), others that it is unique because of the relationship context in which it occurs (Pepler et al., 2005), and still others that it is unique because it is a proactive, goal-direct form of aggression aimed at increasing social status (Salmivalli & Peets, 2009). Whatever the source of conceptual uniqueness is believed to be, the idea that it is unique is taken as a given. One may expect that if bullying is conceptually unique from other forms of aggression, involvement in bullying should predict functioning in a way that is different, not just more severe, than other forms of aggression. The definitional characteristic of repetition, in particular, has been criticised for being a marker of only severity rather than a qualitative difference (Hanish et al., 2013; Ringrose & Renold, 2010). To this point, only one of the studies comparing the outcomes of bullying and peer victimization has found unique outcomes associated with bullying. That is Hunter, Boyle and Warden's (2007) study, which demonstrated that the perceptions of the aggressive situation and coping strategy used by people who are victimized varies depending on whether or not the victimization meets the criteria for bullying. The results of the present study are similar to other results from Hunter Boyle and Warden's study and other studies investigating the differences between bullying

and peer aggression (Felix et al., 2011; Hunter et al., 2007; Ybarra et al., 2014; You et al., 2008): the relationship between bullying or victimization by bullying and functioning is the same, but stronger, than the relationship between aggression or victimization by aggression and functioning for most measures. The exceptions to this general finding in this study are measures of functioning that are not significantly related to aggression or victimization by aggression alone, but that are significantly related to the characteristics of bullying. In this study, those variables are security and acceptance, positive affect and negative affect for aggression, and beliefs about aggression for victimization. These results tentatively suggest that bullying is unique in that bullying, but not aggression, is associated with feeling less socially secure, less positive affect, and more negative affect, and that victimization by bullying is unique in that only people victimized by bullying tend to endorse higher than typical beliefs that the use of aggression is acceptable. These results are presented tentatively as they are highly dependent on statistical artefacts, have low effect sizes, and because we were not able to test interactions directly in this study due to the high correlation among variables. These results need to be replicated, but nevertheless are potentially important given the theoretical importance of discovering what aspects of bullying make it unique. Without replication or expansion of these findings, however, the current balance of evidence suggests that bullying may be associated with more severe, but not unique, outcomes for those who are involved. This suggests the bullying construct has discriminative validity, but an understanding of the any truly unique implications of bullying for individual functioning is still lacking.

This study is unique among the existing research literature in that it is the first to show differences between bullying and aggression and individual functioning using peer-report data. This is important because it suggests that it is not simply the subjective

experience of victimization by bullying that leads to worse outcomes. Classmates were able to observe the circumstances that were related to participants' private reports of worse individual functioning. Classmates also showed agreement about the social standing of participants who experience this type of aggression. In other words, there is something about the type of victimization characterized by repetition, intention to harm and power imbalance that not only feels different, but is also observably different to classmates. This reinforces the conceptual validity of the bullying construct, as victimization by bullying can be observed and recognized by others, as well as subjectively experienced.

The importance of repetition, intention to harm, power imbalance, and meeting full definitional criteria for bullying in predicting functioning.

Results of this study related to the importance of individual definitional characteristics and meeting full definitional criteria for bullying in predicting participant outcomes were, in general, not supportive of the construct validity of the bullying construct as currently defined. In this study, the presence of all three characteristic of bullying did not emerge as an important predictor of participant outcomes, nor did each of the three characteristics specified in the definition contribute uniquely to our ability to predict student outcomes. These findings suggest that the definitional requirement that bullying is repeated, intentionally harmful, and occurs in the context of a power imbalance is not especially meaningful. They also suggest that rather than being independently important, the behavioural characteristics contained within the definition are redundant.

In this study the three characteristics specified in the definition of bullying were very highly correlated. Correlations between characteristics were between .83 and .89 for aggression and .79 and .85 for victimization. While these correlations are high, it is difficult to tell if they are uniquely so compared to other studies. Correlations between the

characteristics are not reported in any of the studies reviewed that measure them directly. We can extrapolate from the data reported by Hunter, Boyle and Warden (2007) that 79.4% of those who experienced repeated aggression also reported at least one form of power imbalance, and only 13% were clear that the aggressor did not intend to hurt them (46.2% thought that they did and 40.8% didn't know). In Felix and colleagues (2011), they conducted a chi-squared analysis and found that frequency of victimization and power imbalance were significantly related, but no effect size estimation was made and not enough data to calculate one was reported. In their study, 75% (i.e., 12 of 16) of a frequently victimized group did report power imbalance, but three of the four most frequently victimized did not report a power imbalance, suggesting to the authors that they were separate constructs. In Ybarra and colleagues (2014), half of participants who were bullied by someone with greater power also reported that the bullying was repeated. In another study using the CBVS, but not comparing outcomes of bullying with peer victimization directly, Green and colleagues (2013) reported that half of their participants who were victimized indicated that it was repeated, whereas three quarters indicated some form of power imbalance existed. Again, the overall correlation between repetition and power imbalance is not reported, but these results are interpreted to mean that repetition and power imbalance in aggressive relationships are distinct experiences, suggesting that each may have unique meaning and predictive power in terms of identifying an at-risk group. That was not the case in this study. Though the amount of overlap among characteristics in the current study is similar to what was reported elsewhere – 65% of participants identified as aggressive in the context of a power imbalance and 52% of participants identified as victims in the context of a power imbalance met the full criteria for bullying, indicating that repetition and intention to harm were also present – each definitional characteristic was not uniquely predictive.

The high correlation among the characteristic variables seen in this study explains why the requirement that all three be present and each individual predictor were not significant on their own. Correlations among repetition, intention to harm, power imbalance and meeting the full criteria for bullying on one hand, and measures of functioning on the other, suggest that for many outcome variables, if entered individually, all of the characteristic variables would be significantly predictive. When they are entered together, however, so much variance is shared among them that individual significance is difficult to achieve (and statistically unreliable – see discussion of multicollinearity).

What is the implication of such a close relationship among variables? On the one hand, it's possible that the correlation is an artefact of measurement – that even when asked explicitly to differentiate among the characteristics, reporters have a difficult time doing so. Criticisms of the definitional criteria of intention to harm and power imbalance on the basis of their subjectivity and measurability have already been discussed (Bauman, 2013; Carrera et al., 2011; Cascardi et al., 2014; Hellström et al., 2013). In addition, while adolescents are able to demonstrate a differentiated understanding of bullying characteristics in a recognition task (Smith & Monks, 2008), it may be more difficult in a recall task such as completing measures like the one used in this study. The measure used in this study is also new. Although similar to other explicit behavioural methods, it is possible that there is something specific about the way that the characteristics were assessed in this study that made them difficult to distinguish. Given that a relatively high degree of overlap among measures of the characteristics seems to have occurred in all of the existing studies that measure the characteristics individually, however, this would appear to be a structural, rather than measure-specific problem. If so, this represents a real weakness in the usefulness of the

current definition of bullying. The operationalization is of limited use if the target subject pool has difficulty separating its components.

On the other hand, rather than being an artefact of measurement, the observed overlap among the characteristics of aggression could be real. It has been argued that in the absence of direct information about an aggressor's intentions, repetition of actions may be the basis on which individuals make their judgements about intention to harm (Olweus, 2013). In addition, if power imbalance makes it difficult to defend oneself, it makes sense that the aggression is likely to be repeated (as it hasn't been halted by a defence). Both repetition and power imbalance have also been theoretically linked to feelings of helplessness and lack of control (Sercombe & Donnelly, 2013). Therefore, the three definitional characteristics of bullying contained within the definition may simply be strongly related.

If we accept that repetition, intention to harm, and the context of a power imbalance are highly related constructs, the implication may be that instead of representing separate aspects of aggressive behaviour that additively define bullying behaviour, they are instead indicators of a single underlying construct. It may be helpful to think of them as indicators or probable characteristics of bullying behaviour, rather than as the fundamental characteristics embodying the bullying construct. In other words, the characteristics contained within the definition of bullying may be useful indicators that bullying has likely occurred (which was, of course, its original purpose), rather than a definitive statement of what bullying is and what makes it unique.

This way of thinking about the definition of bullying is suggested by Smith, del Barrio and Tokunaga (2013) with reference to the classic cognitive psychology work of Rosch and Mervis (1975). Rosch and Mervis propose that cognitively, categories are formed in terms of "family resemblance" or probably indicators rather than definitional attributes

which must apply to every example. As was famously declared by US Supreme Court Justice Potter Stewart, sometimes a category lacks a clearly defined border and one simply "know[s] it when [one] see[s] it" (p. 378; Jacobellis v. Ohio, 378 U.S. 184, 1964). This seems to make intuitive sense when talking about bullying. For example, bullying is sometimes referred to as a natural category or intuitively understandable concept because of the existence of words that identify it in many languages, despite the fact that the meaning of these words varies slightly across cultures (e.g., in terms of the type of aggression emphasized; Monks & Smith, 2006; Smith et al., 2013). Consider also Olweus' (1999) concession that not every instance of bullying need be repeated. In addition, currently many of the different theoretical conceptions of bullying (that it is difficult to defend against or escape from, that it results in a loss of agency, that it is a proactive form of aggression aimed at achieving social dominance) are united in being based on the function or results, rather than the forms, of the behaviour. This suggests that perhaps defining bullying based on form is misguided. As theory develops it may be more precise to define bullying in terms of its unique functional attributes. Behavioural operationalizations of bullying may be necessary for measurement, but may not embody what is essential about the construct. Perhaps we must accept that any attempt to characterise bullying in terms of the form of behaviour is an operationalization of the construct that is bound to be imperfect.

Even if, however, it is more realistic and accurate to think of bullying behaviour as a category with "diffuse borders" (Smith et al., 2013) this is, in general, an impractical position for researchers and policy makers. Both groups need to be able to identify and exclude aggressive incidents from the category of bullying in a systematic way. Given this, the question becomes, of all the operationalizations of bullying possible, is the currently used definition of bullying the best one? The results of this study would indicate no. The presence

of all three specified behavioural characteristics was not helpful in terms of predicting participant outcomes. This indicates that the definition is unnecessarily restrictive. In addition, the high correlation among the characteristics when measured individually indicates either that they are difficult to distinguish for the study population or that they are redundant. Given the difficulty in measuring all three characteristics, the statistical complexity introduced by the number of closely related predictors, and the lack of empirical evidence that all of the components are necessary, the definition seems inefficient. In addition, though there is now building proof that the current definition can identify a distinct group, evidence that this group is truly unique is not robust. The current definition is also rather atheoretical and lacks a clear conceptual statement about what bullying is beyond how it is operationalized. As such, there are legitimate reasons to question the construct validity of bullying as currently defined.

Though this study alone cannot determine what a better way to operationalize bullying would be, the results of this study suggest that the most promising aspect of the current definition to focus on moving forward is the power imbalance characteristic. Though much of the variance in predicting outcomes was shared among predictors in this study, power imbalance was often the defining variable that accounted for the most unique variance and was also generally the single variable able to capture the most overall variance. This result is consistent with previous research, and the general perception that power imbalance is the hallmark feature of bullying (Hanish et al., 2013; Ybarra et al., 2014). Power imbalance also has the strongest theoretical backing of any of the three characteristics included in the current definition. Power imbalance is strongly tied to social dominance theories about the possible function and unique nature of bullying. Pursuing a deeper understanding of the types of power imbalance that are important, the mechanisms by which

power imbalance affects the outcomes' victims, and the potential benefits of the aggression for the perpetrators could allow for a more streamlined definition of bullying that could also be more precise in operationalizing the underlying features that make bullying unique. Perhaps a focus on the power imbalance construct could also resolve the open question about whether variability in dominance outcomes observed in individuals who bully is a real phenomenon or due to imprecision in measurement. If bullying is defined by function, or outcomes, operationalization of bullying into any behavioural characteristic will continue to be imperfect, but increased precision and clarity of construct could reduce measurement error.

Study Two

This study addresses the validity of the bullying construct by examining whether a widely used measurement strategy captures bullying as defined. In the definitional strategy for assessing bullying, participants are presented a definition of bullying before being asked to identify whether they or others have been involved in bullying. In this way, definitional measures are unambiguously designed to capture bully defined as aggression that is repeated, intentionally harmful, and occurs in the context of a power imbalance. In this study, we examined the extent to which a definitional measurement strategy actually captured aggression that meets these criteria, as measured through an explicit behavioural strategy where the presence of aggression and each of the definitional criteria are measured separately. We found limited evidence for the construct validity of the definitional measurement strategy. There was poor agreement between the two measurement strategies examined, and most of the variance predicted in definitional bullying status could be accounted for by the presence of aggression or victimization alone. This lack of agreement between measurement strategies may not be that meaningful, however, as there is little

evidence of differences in functioning between those identified as involved in bullying using the different methods.

Agreement between measurement strategies.

The first goal of Study Two was to measure the agreement between bullying measured through the definitional measurement strategy with bullying measured through an explicit behavioural strategy. The hypothesis related to this goal (Hypothesis 2a) was supported. The explicit behavioural and definitional methods of assessing bullying were significantly related, but the concordance between measures was found to be only poor, at the low end of concordance levels reported in previous studies (Felix et al., 2011; Green et al., 2013). This suggests a meaningful difference in what is being captured by each measurement strategy. Almost half of participants (49.5%) were definitionally identified as having a bullying status (i.e., bully, victim of bullying, or bully/victim) that was not justified by their measured behaviour.

As in the previous research on this topic, the concordance rates found between two measures that were designed to capture identical constructs was lower than may be expected. The concordance rate observed in this study is at the low end of reported concordance rates, but that can perhaps be explained by that fact that in this study, concordance rates were examined for all types of bullying involvement, whereas the previous studies examined concordance among identification of victims of bullying only. Concordance measured across four categories is more difficult to achieve than concordance across two.

Examining concordance across four types of bullying involvement (bullying, victimization by bullying, bullying and being victimized by bullying, and being uninvolved in bullying) allows our study to compare rates of concordance for different bullying roles, something that other studies were unable to do. The highest level of agreement found was for

the uninvolved group, with 90% of participants that had a definitional categorization of uninvolved also having uninvolved status according to the explicit behavioural methods. Agreement rates were much lower for the bully group (49%), the victim of bullying group (33%) and the bully/victim group (17%). We could expect that the bully/victim group would have the lowest level of agreement overall, as classification into this group depends on agreement on both whether the person bullies others and whether they are bullied. Nevertheless, the bully/victim group has distinct qualities and expected to have the worst outcome of any group involved in bullying (Volk et al., 2012), so it is important to be able to identify this group accurately. Our results suggest that the definitional method of measurement do not accurately classify these individuals based on their behaviour.

The levels of agreement between measurement methods observed in our study suggests that agreement between measures was weaker for victimization by bullying than it was for the perpetration of bullying. This can be accounted for by the fact that many more participants were identified as victims of bullying using the definitional method than the explicit behavioural method. Forty-two percent of the total sample was identified as experiencing victimization by bullying using the definitional method, compared with 20% using the explicit behavioural method. This over-inclusive reporting using the definitional method was not seen for the bully group. The higher prevalence of victimization by bullying reported using the definitional method is seen as over-reporting due to the fact that the explicit behavioural method enforces definitional compliance in identifying bullying involvement, in that participants are asked explicitly about each definitional component of bullying, whereas The definitional method asks definitional compliance, it does not demand it and could be influenced by other factors.

The over reporting of victimization by bullying through the definitional method observed in this study is unusual compared to the two previous studies that compared the definitional and explicit behavioural methods. Felix and colleagues (2011) found that the two methods identified roughly the same number of victims overall and Green and colleagues (2013) found that the explicit behavioural method identified more victimization. Given this conflict between our findings and previous studies, it must first be considered whether this finding could be simply an artefact of the specific measurement strategy used in this study. Being classified as involved in bullying using the explicit behavioural method is based on four questions (presence of aggression, intention, repetition, and power imbalance), as opposed to one question for the definitional method. Being identified as a victim of bullying through the explicit behavioural method could therefore be less common simply because it is a higher methodological bar to cross. If this were true, however, we would expect to see the same pattern hold across all categories of bullying involvement, and the definitional method does not identify significantly more participants in the bullying group than the explicit behavioural method. In addition, the explicit behavioural measures in the other studies were also made up of multiple questions, as they also explicitly assessed definitional characteristic separately. Unlike the other two studies, however, our definitional method did not include a measure of frequency to be used for forming the victimization group. Instead, participants were simply asked to identify bullying that had occurred. This is likely a lower threshold for group entry (Solberg & Olweus, 2003). Therefore, our version of the definitional method may be slightly more inclusive than in other studies, which could partially account for the difference. There is no frequency specification put forth in the definition of bullying, however, so our method still measures bullying as defined.

In addition, our study uses peer-report nominations to collect information on bullying involvement, whereas the other two studies use self-report. This could affect the amount of victimization by bullying reported in several ways. If there is indeed a stigma associated with identifying oneself as a victim of bullying (Espelage & Swearer, 2003), peer reports would seem to avoid this problem, as identifying others as victims of bullying would not have the same implications for one's self-image. Peer reports also measure a different aspect of bullying involvement than self reports (Branson & Cornell, 2009; Juvonen & Graham, 2001). Social reputations for victimization by bullying may be more likely to be over reported using the definitional method than self perceptions of victimization.

What is clear is that, in this study, the only known study to examine this question using peer report and only the third known study to examine this question overall, many people who were identified as being involved in bullying through definitional methods, especially many people who were identified as being victimized by bullying, do not have explicit behavioural bullying nominations that indicate that their aggression or victimization actually meets the criteria for bullying involvement. This is true both because of over reporting of victimization, and because of lack of agreement among all groups involved in bullying.

One interpretation of these findings is that there is a difference in what is evaluated using the definitional and explicit behavioural methods of assessing bullying involvement. If our assumption is true, and the explicit behavioural method closely captures bullying as defined, then the disagreement between methods indicates that the responses to definitional measure are strongly influenced factors other than the behavioural criteria specified in the definition. Perhaps, as has been feared, reporters using the definitional method do not distinguish between bullying and aggression, and their reports are strongly influenced merely

by the presence of any type of aggression or victimization (Finkelhor et al., 2012; Hanish et al., 2013; Rigby, 2012). Counter to that theory however, is that fact that in this study, more aggressive involvement than bullying involvement was identified. Specifically 83% of the total sample was identified as having some type of aggressive involvement whereas only 57% of the total sample was identified as having some type of definitionally identified bullying involvement. Therefore, reporters do appear to make some sort of distinction, it is just not clear that this distinction is always based on definitional criteria. The basis upon which reporters identify bullying involvement using definitional methods is the subject of the next section.

Predicting definitional bullying status with behaviourally measured definitional characteristics.

The second goal of Study Two was to examine which components of aggressive behaviour (as measured through the explicit behavioural method) were associated with being identified as a bully or victim of bullying using the behavioural method. The results of this study support some of our study hypotheses but not others. First of all, information about aggression and victimization gathered through the explicit behavioural method was able to predict definitional identification as a bully or victim of bullying. Seventy percent of the variance in definitional bully nominations and 60% of the variance in definitional victim of bullying nominations was accounted for by this information. Of interest is the breakdown in how much of that variance is explained. Of the 70% of variance explained in nominations for definitional bully status, 64% could be accounted for by the presence of aggression alone, 2% was accounted for by information about the three definitional characteristics, and 4% was accounted for by whether or not individual met full definitional criteria. This means that hypotheses 2b and 2d were technically supported with regards to bullying, but the amount of

variance in definitional identifications for bullying accounted for by definitional characterises of bullying was proportionally very small. Of the 60% of variance accounted for in nominations for definitional victim of bullying status, 47% could be explained by the presence of victimization alone, 9% was accounted for by information about the three definitional characteristics, and 4% was accounted for by whether or not an individual met full definitional criteria for victimization by bullying. Again, hypotheses 2b and 2d are supported with regards to victimization by bullying. The amount of variance accounted for by definitional characteristics of bullying was still proportionally small, but it was more than double the amount of variance that the characteristics of bullying explained with regards to definitional bully nominations.

The finding that the majority of variance in definitional nominations for bully or victim of bullying status can be accounted for by the presence of aggression or victimization alone does not necessarily mean that participants are not distinguishing between bullying and other forms of aggressive behaviour on the definitional measure. Bullying is a subtype of aggression, meaning that for bullying to have occurred, an aggressive act is necessarily also present. We would therefore expect that participants who are not nominated as aggressive or victims of aggression would not be identified as part of the bully or victim group using the definitional method. Some relationship between the presence of aggression and definitional bullying status is expected, and even a sign of convergent validity. The finding is also consistent with the existing research, as Green and colleagues (2013) also found that the presence of various forms of victimisation predicted definitional victim status. The question of interest is, in reality, not whether the presence of aggression predicts definitional bullying status, but how much variance is left over to be predicted by other factors (i.e., the things which make bullying a unique subset of aggression). In this study, the amount of variance

not explained by aggression is sufficient to suggest that reporters are taking other factors into account when making definitional bullying identifications.

The amount of variance accounted for by definitional characteristics of bullying in our study suggests that repetition, intention to harm, power imbalance and whether all three characteristics are present simultaneously are only a small part of what participants are considering when reporting on bullying involvement. For bullying in particular, the results suggest that definitionally identified bullying may not be as distinct from other forms of aggression in terms of these features as one would hope, given that the features are meant to be defining. To the extent that a distinction between bullying and aggression is made, it appears to be largely based on other unknown factors.

The fact that definitional features of bullying are more important in predicting identification as a victim of bullying than identification as a perpetrator of bullying is something of a puzzle, given the previous finding that there is less agreement between definitional and explicit behavioural methods in the identification of people who are victimized by bullying than there is in the identification of people who are bullied. We can't say that participants seemed to be making more of a differentiation between victimization by peer aggression and victimization by bullying than they were between aggression and bullying overall, because the definitional measure actually appears to be over-inclusive when measuring victimization by bullying. Perhaps the answer lies in the fact that the mere presence of victimization accounts for less of the variation in whether a person is identified as being involved in bullying and the fact that there is less variation in definitional victim status accounted for overall. If the occurrence of victimization accounts for less variance, then there is more variance left to be explained by other factors. The definitional criteria for bullying would appear to be only one of the other factors that are considered.

With regards to identifying which characteristics of bullying are particularly important for predicting definitional bullying status, our hypothesis that each would be individually significant (Hypothesis 2c) was not supported. As in Study One, the characteristic of aggression predictors were found to be highly correlated, meaning that while they were significantly related to definitional bullying status overall, no individual characteristics contributed much individual variance prediction. For victimization by bullying, the all three characteristics variable did emerge as the strongest predictor, and the best summary of what can be accounted for in the variance in definitional nominations for victim of bullying. This study did not find, however, that for either bullying or victimization by bullying the definitional characteristics were independent predictors that additively predicted bullying or victim of bullying nominations. Rather, as discussed in Study One, it is perhaps better to think of the characteristics as indicators of a common underlying construct rather than separate defining characteristics. This result is contrary to Green and colleagues (2013), who found that repetition of victimization independently predicted identification as a victim of bullying. More replication of this type of analysis is required to determine the source of this disparity. Since this study uses peer reporters and their study is based on self-report, one explanation could be that peer reporters differentiate less among the definitional characteristics of bullying than self-reporters do when using definitional methods. This would be consistent with the idea that peer reports are based on social reputation as a victim (Branson & Cornell, 2009; Juvonen & Graham, 2001), which is a less differentiated construct than personal experiences. Whatever the reason, while Grief Green and colleagues (2013) conclude that definitional measures are sensitive to repetition but not power imbalance, the conclusion of this study is that using peer report, definitional measures were

somewhat sensitive to an underlying bullying construct, particularly definitional measures of victimization, but not to any to definitional characteristic in particular.

A final point to consider is that, in this study, only a proportion of the total variance in definitional bully and victim nominations could be accounted for through the definitional characteristics of bullying as measured through explicit behavioural methods. Consistent with hypothesis 2e, 30% of the variance in bullying nominations and 40% of the variance in victim of bullying nominations was unexplained. This is a relatively large amount of unexplained variance given that the definitional method of measuring bullying is intended to measure the definitional criteria of bullying and only these criteria. Still, it was hypothesized that such a gap between what was intended to be measured and what was measured might occur for two reasons that come out of the literature on bullying measurement: the definition of bullying may be difficult for reporters to comprehend and apply (Bieber, 2013) and children and adolescent's own personal understanding of bullying seems to differ from researchers' definition (Frisén et al., 2008; Monks & Smith, 2006; Smith et al., 2002; Vaillancourt et al., 2008). These reasons would appear to be complementary, as difficulty understanding the provided definition would seem to contribute to a tendency to augment it with one's own personal understanding.

While it is beyond the scope of this study to make any conclusions with regards to the source of variance not explained in our prediction model of definitional bullying status, it seems likely that variance in nominations for definitional bullying status that are not predicted by definitional characteristics of bullying are likely related to our participants' own understanding of the bullying construct. Research has shown that adolescents define bullying based on the presence of various forms of aggressive behaviour, personality characteristics of the bully, function of the behaviour, and who is carrying out the aggression and who is

victimized, among other factors (Frisén, Hasselblad, & Holmqvist, 2012; Land, 2003; Vaillancourt et al., 2008). Future research should explore the extent to which these factors influence nominations for bullying involvement using the definitional method. Research in this area could provide vital information on what factors in addition to definitional criteria influence peer nominations for bullying involvement using definitional methods, allowing for increased understanding of what is being measured. Adolescents' understanding of bullying is also of interest as it may provide ideas for characteristics of aggressive behaviour that are not currently encoded in the definition of bullying but could help to increase the clarity and distinctiveness of the way that bullying is conceptualized and measured (Carrera et al., 2011).

Comparing indicators of social and individual functioning of participants identified as involved in bullying through the definitional and explicit methods.

The third goal of Study Two was to examine whether the method by which an individual was identified as being involved in bullying was related to indicators of social and individual functioning. Specifically, those who were identified as involved in bullying through the explicit behavioural method (and therefore were presumed to meet full definitional criteria for bullying involvement), were compared to those who were identified as involved in bullying through the definitional method only (and thus presumed not to meet full criteria for bullying involvement). Very few differences between groups were found to be significant. Those who met the full criteria for victimization by bullying were less popular and had more negative affect than students who were identified as victims of bullying through the definitional method only. These differences were small, and the only differences found among eleven measures of functioning. The results overall do not suggest that meeting full definitional criteria for victimization by bullying is as meaningful as expected. No

differences in functioning were found between those identified as perpetrating bullying using different methods.

These results were counter to hypotheses 2f and 2g and somewhat surprising, given meeting the full definitional criteria for bullying should be meaningful in terms of measurable differences in outcomes if the definition itself is valid. The most likely reason for the lack of significant findings is that those identified as involved in bullying through the definitional and explicit behavioural methods have more in common than they have distinguishing them. The results of Study Two to this point have shown that that definitional bullying involvement is highly related to aggressive involvement, and that definitional identification for victimization in particular is also predicted by definitional characteristics. Taken together, these results suggest that those identified as being involved in bullying through the definitional method only are likely involved in aggressive behaviour or victimization and that this aggressive behaviour may have some of the characteristics of bullying. Given that the results of Study One suggested that meeting full criteria for bullying is not as meaningful as expected in terms of participant functioning and, as a cut-off for identifying bullying involvement, may be too restrictive, the differences aggressive involvement between groups may simply not be meaningful. Therefore, the group identified as involved in bullying through the definitional method only may have functioning that is similar to those whose bullying involvement meets the criteria for bullying, though perhaps not as strongly impacted. These similarities make it difficult to find statistically significant differences between groups. The similarities may also mean that though definitional and explicit behavioural methods for identifying bullying involvement show poor agreement and the definitional method does not necessarily strongly reflect the presence of definitional criteria for bullying, in practical terms, these distinctions may not be that important.

Conclusions and Implications for Future Research

Taken together, the results of Studies One and Two suggest that the current definition of bullying lacks construct validity. This conclusion is based on four major findings:

1. While the behavioural markers contained within the definition do seem to identify particularly severe aggressive involvement, there is little evidence that this aggression is qualitatively different in terms of its relationship to functioning.
2. There is little evidence for the integrity of the definition itself, in that there is little evidence that each characteristic is necessary for capturing a larger construct or that meeting full definitional criteria is meaningful.
3. Results indicate that definitional measurement strategies do not capture bullying as defined
4. Few differences were found between individuals who did and did not meet the definitional criteria for bullying.

Given this conclusion, two important foci for future research into bullying are suggested. First, the results suggest that it is time to reopen the discussion on what bullying is, what makes it unique, and how it should be defined. The definition of bullying generated by Dr. Olweus in the 1970s has absolutely served a valuable purpose and been crucial to moving the field forward (Finkelhor et al., 2012). Its widespread adoption has, however, also unintentionally limited the discussion and research on the conceptual and objective factors that delineate what is theorized to be a unique subset of aggressive involvement. The definition is often treated as fact, despite not being conceptually or empirically well validated, and that this results in a lack of work assessing the importance of various aspects of the bullying experience (Bauman et al., 2013; Finkelhor et al., 2012).

Dr. Olweus recently gave the following analysis, not of bullying research, but of peer rejection research, whose popularity he feels delayed widespread interest in studying bullying in North America:

> "It appears to me that the availability of a convenient measurement technique and classification system did take precedence over conceptualization and problem formulation. In my view, this also contributed to preventing new ideas, approaches and themes such as bullying from coming into focus for quite some time." (Olweus, 2013, p. 773)

Ironically, the same statements could be applied to how bullying research itself has progressed. Overall, there is a growing sense that conceptual differences between bullying and other forms of victimization need to be better understood before there is sufficient justification for studying the construct separately (Bauman et al., 2013; Finkelhor et al., 2012; Hellström et al., 2013). The results of this study suggest that the behavioural characteristics contained within the current definition of bullying do not adequately describe a unique form of aggression. It is time that the field seriously addresses this issue.

Finkelhor and colleagues (2012) suggests that one way forward is to make bullying less of a central concept and instead focus on the broader categories of peer victimization and peer aggression.

> "A more empirical and scientific approach often adopted in aggression research has been to define acts and behaviours in a broad fashion, and then to study the various contextual features that make such acts in some situations and relationships more harmful than others (p. 272)".

In other words, he suggests letting research findings guide the definition of especially harmful kinds of aggressive involvement (which may still be referred to as bullying) rather than defining bullying and setting out to prove that it is more harmful. He cites as a potential example the way that the study of rape gave way to the broader study of

sexual assault and sexual violence. This, he argued, allowed research to continue studying the effects of both penetrative and non-penetrative sexual assault. As a result, the research has since shown empirically that penetration is indeed associated with increased seriousness, while also exploring the harmful effects of other types of sexual assault. This example shows that increasing the scope of enquiry doesn't necessarily mean that researchers must or should abandon efforts to identify specific contexts in which peer victimization has a particular meaning or is particularly harmful, but that it does help to ensure that "whole categories of behaviours have [not] been excluded a priori and not even studied (IBID)."

Taking the approach advocated by Finkelhor would allow researchers more freedom to investigate the current definition and also to investigate other contexts and variations in the experience of peer aggression and victimization which may ultimately allow us to improve upon current understanding. These types of studies, however, could also take place without abandoning current terminology. The term bullying does give a central focus and common language to a large body of research work. It has also already been encoded into law and educational policy, meaning that there is perhaps a responsibility to clarify how the term should be best measured and understood. Whether or not the field ultimately responds positively to Finkelhor's suggestion of expanded focus, it is clear that increased freedom and creativity in attempts to understand the bullying construct are needed.

A second focus of future research should be to closely tie the definition and conceptualization of bullying to a strong theoretical viewpoint. The highly correlated nature of the definitional criteria observed in this study supports the intuition that there is a natural category of behaviour that we have labelled bullying (Smith et al., 2013), while at the same time suggesting that the behaviours contained within the definition are more indicators of this category than the embodiment of it. The question now becomes what theoretically

defines this natural category. A theoretical explanation of bullying will generate hypotheses about the unique etiology, function, consequences of and intervention strategies for bullying which can then be tested empirically. It is ultimately only a strong theoretical understanding of bullying which will lead to conceptual clarity and thus construct validity.

In the opinion of this author, the most promising theoretical conceptualization of bullying to examine in future research is the theory of bullying as proactive, aggressive behaviour used to achieve and maintain social status (Pellegrini & Long, 2002; Salmivalli & Peets, 2009; Salmivalli, 2010; Veenstra et al., 2007). There are several reasons for this. First, the power imbalance criteria already contained within the definition of bullying is thought to be the most important and distinctive feature of current conceptualizations of the meaning of bullying (Hanish et al., 2013; Ybarra et al., 2014). Second, it connects theories of bullying to existing well developed theories such as social dominance theory and evolutionary theory (Hawley, 1999; Volk et al., 2012). Thirdly, there is already an existing and growing body of research conceptualizing bullying from this perspective that has shown promising results. Research has shown that some individuals who bully appear to fit the profiles of bistrategic controllers (known to be the profile most associated with social dominance; Book, Volk, & Hosker, 2012; Olthof et al., 2011) and successfully use bullying behaviour as a method of achieving social dominance without high costs (Dijkstra, Lindenberg, & Veenstra, 2008; Juvonen & Graham, 2014; Olthof et al., 2011; Reijntjes, Vermande, Goossens, et al., 2013; Volk et al., 2012). Bullying has also been shown to be related to status goals (Houghton et al., 2012; Thornberg, 2013), and those goals appear to drive bullying behaviour (Dawes & Xie, 2014; Li & Wright, 2014). This work all suggests that bullying is related to social dominance. Future work should focus on determining whether this relationship between social dominance and aggression is unique to the bullying subtype of aggressive behaviour,

and on translating increased theoretical understanding to increased clarity in the way that bullying is defined and measured.

There is already some suggestion from within this study and other existing research about how our understanding of bullying behaviour may be refined by social dominance theories of the behaviour: it may help to resolve the issue of variability in the bully group. Current methods of defining and measuring bullying appear to capture two subtypes of bullies: those who are successful at using proactive aggression to achieve social status, and those whose aggression appears less skilled, more reactive, and less successful at achieving social dominance (Carrera et al., 2011; Hart & Ostrov, 2013; Olthof et al., 2011; Volk et al., 2012). Many of those in the second group may fall into the bully/victim category. Besides the differences in the skill and forethought that goes into their aggressive behaviour, individuals in these groups also have different social consequences for their aggression (Dijkstra et al., 2008), respond differently to intervention (Garandeau, Lee, & Salmivalli, 2014), and seem to aggress against different people (Rodkin & Berger, 2008). There is also good reason to suppose that being bullied by members of these different subgroups would lead to different outcomes for the victims of the bullying, as a socially dominant peer would seem to have more ability to affect a person's social status than an unpopular one. This last assertion has not yet been examined empirically. Nevertheless, it seems that identifying both of these classes of aggressors as bullies introduces variability into the meaning of the bullying construct. If bullying is conceptualized as proactive, aggressive behaviour used to achieve and maintain social status, perhaps the bullying behaviour of the bully/victim group is not really bullying at all, despite similarities in behavioural characteristics. Reactive unskilled aggression can be repeated, intentionally harmful, and in the context of some type of power imbalance, without necessarily being the same in terms of function and effects as

the aggression of socially dominate individuals. If social dominance theory can provide an explanation as to why the function or outcomes of bullying behaviour are more important than the form, our knowledge of the differences between these two subgroups could be used to increase the precision and clarity of the bullying construct by being incorporated into how bullying is measured and defined. Perhaps the power imbalance criterion could be refined to specify imbalances in social status, or bullying could be limited to only aggression that results in increased social status for the aggressor and decreased social status for the victim. These possibilities and their implications should be investigated empirically, but illustrate how a strong theoretical conceptualization of bullying could result in a more precise bullying construct and the possibility of more specific measurement.

Implications for research procedure.

While work on clarifying the bullying construct is conducted, researchers will continue to have to make choices about how to measure bullying. Somewhat surprisingly, the results of this study suggest that the definitional method of assessing bullying involvement, though imperfect, is the appropriate choice until improvements in how bullying is conceptualized are realized. This is despite the findings that identification by the definitional method is only somewhat predicted by the definitional criteria of bullying and shows poor agreement with measurement methods that explicitly measure these criteria. After all, what these results really show is that the definitional measurement strategy shows weak to poor fidelity to Olweus' definition of bullying, and the usefulness of this definition is also an open question. The results of this study suggest that this definition is in fact not particularly meaningful. It appears to be overly restrictive and there is not compelling evidence of large differences in functioning between those who meet full criteria for bullying and those who are only definitionally identified as involved. Given this, the advantages of continuing to use

the definitional method appear to outweigh the weakness of not being strongly related to definitional criteria. The definitional method of measuring bullying has already been widely adopted and thus many studies using this technique are comparable to one another even if they don't measure the bullying construct exactly as defined. Given their longer history, definitional strategies of measuring bullying also have more evidence of validity than the newer, explicit behavioural approach. While the explicit behavioural measurement approach may indeed be more effective at measuring bullying as defined, and is important for investigating the definitional characteristics of bullying, it does not appear to be worth the cost of moving the field towards a more definitionally bound measurement strategy until the definition that is being measured is better validated.

Implications for policy.

Those who are in the position of crafting policy based on the current definition of bullying are in a difficult position. While evidence for the construct validity of bullying is weak, some objective way of defining the construct is needed while definitional refinements are made. The finding from this study most of interest to policy makers is the finding that the presence of all three characteristics contained within the definition of bullying do not add to our ability to predict the functioning of those who perpetrate or are victimized by bullying. As such, the requirement that all three criteria be present may be too restrictive an operationalization of bullying.

One possible way forward would be to de-emphasize the bullying construct in policy and focus on the broader category of peer victimization, especially since the results of this study suggest that the terms may be used almost interchangeably. However, doing so may mean that the range of behaviour captured under peer victimization would be quite broad and risks the advances that have been made in increasing awareness and responsiveness to

particularly harmful and severe types of peer victimization. In addition it remains intuitively likely that a unique and especially harmful form of aggression (i.e., bullying) exists, despite our difficulty in measuring and defining the construct. It may therefore be advisable to reserve the term bullying for describing this special subtype of behaviour and retain the distinction between bullying and peer victimization in policy.

Another possibility, then, may be to work from a definition of bullying that is less restrictive than Olweus' definition. In this study, while the definitional characteristics were redundant and no one predictor stood out as uniquely important, the power imbalance characteristic stood out as the best summary variable to capture the variance largely shared among predictors. It is also the behavioural characteristics with the most empirical and theoretical support. Therefore, an appropriate way of operationalizing bullying in legal or educational policy may be aggression that occurs in the context of a power imbalance. This definition would make bullying distinct from harassment only in terms of the fact that the type of power imbalance was not limited to certain categories, but it would retain the special emphasis on bullying prevention that has been achieved.

Limitations

As with all research, this study has several limitations. First, it assesses the behavioural characteristics of aggression using an implementation of the explicit behavioural measurement strategy that is unique to this study. Little evidence for the validity or reliability of this measurement strategy, or this particular implementation of it, exists. It is possible that this measure was confusing to participants, or that for some other reason it does not accurately assess the behavioural characteristics of bullying as has been assumed. If this were true, it would change the interpretation of many of the results observed in this study. It was necessary, however, to adopt this measurement approach, as there is no existing measure

that aims to explicitly capture the definitional characteristics of bullying through peer report. Future replication of these results using different implementations of this measurement strategy, or validation of the current measurement approach would increase confidence in the results of this study.

Another limitation of this study is that conclusions were drawn about the relationship between the definitional characteristics of bullying and adolescent social and individual functioning although the number of areas of student functioning investigated was limited by the data collected in this study. Bullying characteristics may be related more strongly or in different ways to types of functioning not measured. Results related to individual functioning should be treated as especially preliminary, as bullying involvement was assessed through peer report in this study and peer-reported bullying involvement is known to be less strongly related to individual functioning (Bouman et al., 2012; Scholte et al., 2013).

Summary

This study assessed the construct validity of bullying as it is currently measured and defined. Evidence for construct validity in both areas is weak. The definitional criteria of bullying are associated with worse, but not unique, functioning in participants, and the definition overall was shown to be redundant and overly restrictive. Theoretically based empirical work to increase clarity of the bullying construct is identified as a research priority for the field. In the meantime, it may be more useful to think of repetition, intention to harm, and power imbalance as indicators that bullying is likely to have occurred, rather than as defining characteristics. In terms of measurement, the results of the study suggest that the definitional measurement strategy, which is widely used, does not strictly measure bullying as defined. Definitional measurement strategies are still recommended, however, because of problems with the current definition. Increasing the clarity of the bullying construct and the

way it is measured are fundamental to increased understanding of what makes bullying unique.

Bibliography

#UnitedAgainstBullying. (2014). Retrieved September 07, 2014, from http://www.ellentv.com/tags/UnitedAgainstBullying/

Adler, P., & Adler, P. (1995). Dynamics of Inclusion and Exclusion in Preadolescent Cliques. *Social Psychology Quarterly*, *58*(3), 145–162. Retrieved from http://www.jstor.org/stable/2787039

Allan, S., & Gilbert, P. (1995). A social comparison scale: Psychometric properties and relationship to psychopathology. *Personality and Individual Differences*, *19*(3), 293–299. doi:16/0191-8869(95)00086-L

Allan, S., & Gilbert, P. (1997). Submissive behaviour and psychopathology. *The British Journal of Clinical Psychology*, *36*(4), 467–88. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9403141

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th Editio.). Washington, DC: American Psychiatric Publishing.

Anderson, C. A., & Bushman, B. J. (2002). Human aggression. *Psychology*, *53*(1), 27.

Archer, J. (2004). Sex Differences in Aggression in Real-World Settings: A Meta-Analytic Review. *Review of General Psychology*. *Vol 8(4)*, *8*(4), 291–322.

Atlas, R. S., & Pepler, D. J. (1998). Observations of Bullying in the Classroom. *The Journal of Educational Research*, *92*(2), 86–99.

Babad, E. (2001). On the conception and measurement of popularity: more facts and some straight conclusions. *Social Psychology of Education*, *5*(1), 3–29. Retrieved from http://resolver.scholarsportal.info/resolve/13812890/v05i0001/3_otcamomfassc.xml

Bagwell, C. L., Coie, J. D., Terry, R., & Lochman, J. (2000). Peer clique participation and social status in preadolescence. *Merrill-Palmer Quarterly*, *46*(2), 280–305. Retrieved from file:///Users/maggiemcgugan/Library/Application Support/Firefox/Profiles/ufj7taq9.default/zotero/storage/GPS5EMTA/n7.xml(1).bib

Baly, M. W., & Cornell, D. G. (2011). Effects of an educational video on the measurement of bullying by self-report. *Journal of School Violence*, *10*, 221–238.

Bandura, A. (1977). *Social learning theory* (p. 264). Prentice Hall. Retrieved from http://books.google.ca/books?id=IXvuAAAAMAAJ

Baron, R. A., & Richardson, D. R. (2004). *Human Aggression* (p. 444). Springer. Retrieved from http://books.google.ca/books?id=MOljg8xXI1oC

Bauman, S. (2013). Why It Matters. In S. Bauman, D. Cross, & J. Walker (Eds.), *Principles of Cyberbullying Research: Definitions, Measures, and Methodology* (pp. 23–25).

Bauman, S., Underwood, M. K., & Card, N. A. (2013). Definitions: Another Perspective and a Proposal for Beginning with Cyberaggression. In *Principles of Cyberbullying Research: Definitions, Measures, and Methodology*.

Bieber, K. E. (2013). *Do Students Understand What Researchers Mean by Bullying ? by*. University of Nebraska-Lincoln.

Biggs, B. K., Vernberg, E. M., Little, T. D., Dill, E. J., Fonagy, P., & Twemlow, S. W. (2010). Peer victimization trajectories and their association with children's affect in late elementary school. *International Journal of Behavioral Development*, *34*(2), 136 –146. doi:10.1177/0165025409348560

Bjorkqvist, K., & Osterman, K. (1998). *Scales for Research on Interpersonal Relations*. Vasa, Finland.

Book, A. S., Volk, A. a., & Hosker, A. (2012). Adolescent bullying and personality: An adaptive approach. *Personality and Individual Differences*, *52*(2), 218–223. doi:10.1016/j.paid.2011.10.028

Bouman, T., van der Meulen, M., Goossens, F. A., Olthof, T., Vermande, M. M., & Aleva, E. A. (2012). Peer and self-reports of victimization and bullying: Their differential association with internalizing problems and social adjustment. [References]. *Journal of School Psychology*, *50*(6).

Bovaird, J. A. (2010). Scales and Surveys: Some Problems with Measuring Bullying Behavior. In S. R. Jimerson, S. M. Swearer, & D. L. Espelage (Eds.), *Handbook of bullying in schools: An international perspective*. (pp. 277–292). New York: Routledge.

Brank, E. M., Hoetger, L. a., & Hazen, K. P. (2012). Bullying. *Annual Review of Law and Social Science*, *8*(1), 213–230. doi:10.1146/annurev-lawsocsci-102811-173820

Branson, C. E., & Cornell, D. G. (2009). A Comparison of Self and Peer Reports in the Assessment of Middle School Bullying. *Journal of Applied School Psychology*, *25*(1), 5–27. doi:10.1080/15377900802484133

Bukowski, W. M., & Abecassis, M. (2007). Self, Other, and Aggression: The Never-Ending Search for the Roots of Adaptation. In P. Hawley, T. D. Little, & P. C. Rodkin (Eds.), (pp. 185–209). Mahwah N.J.: L. Erlbaum Associates.

Bullying Policy & Legislation. (2014). Retrieved September 08, 2014, from http://www.prevnet.ca/resources/policy-and-legislation

Bushman, B. J., & Anderson, C. A. (2001). Is it time to pull the plug on hostile versus instrumental aggression dichotomy? *Psychological Review*, *108*(1), 273–279. doi:10.1037/0033-295X.108.1.273

Bushman, B. J., & Huesmann, L. R. (2010). Aggression. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), (5th ed., pp. 833–863). Wiley. Retrieved from http://www.amazon.com/dp/0470137487

Butovskaya, M. L., Timentschik, V. M., & Burkova, V. N. (2007). Aggression, conflict resolution, popularity, and attitude to school in Russian adolescents. *Aggressive Behaviour*, *33*(July 2004), 170–183. doi:10.1002/ab

Cairns, R. B., Cairns, B. D., Neckerman, H. J., & Gest, S. D. (1988). Social networks and aggressive behavior: Peer support or peer rejection? *Developmental Psychology*, *24*, 815–823. doi:10.1037/0012-1649.24.6.815

Card, N. A., & Hodges, E. V. E. (2005). Power differential in aggressor-victim relationships. In N. A. Card & E. V. E. Hodges (Eds.), . Atlanta, GA.

Card, N. A., & Hodges, E. V. E. (2008). Peer victimization among schoolchildren: Correlations, causes, consequences, and considerations in assessment and intervention. *School Psychology Quarterly*, *23*(4), 451.

Card, N. A., Isaacs, J., & Hodges, E. V. E. (2009). Aggression and victimization in children's peer groups: A relationship perspective. In A. Vangelisti (Ed.), (pp. 235–259). New York: Cambridge University Press.

Card, N. A., & Little, T. D. (2006). Proactive and reactive aggression in childhood and adolescence: A meta-analysis of differential relations with psychosocial adjustment. *International Journal of Behavioral Development*, *30*(5), 466 –480. doi:10.1177/0165025406071904

Card, N. A., & Little, T. D. (2007). Differential relations of instrumental and reactive aggression with maladjustment: Does adaptivity depend on function? In P. Hawley, T. D. Little, & P. C. Rodkin (Eds.), (pp. 107–134). Mahwah N.J.: L. Erlbaum Associates.

Card, N. A., Stucky, B. D., Sawalani, G. M., & Little, T. D. (2008). Direct and indirect aggression during childhood and adolescence: A meta-analytic review of gender differences, intercorrelations, and relations to maladjustment. *Child Development*, *79*(5), 1185–1229.

Carrera, M. V., DePalma, R., & Lameiras, M. (2011). Toward a More Comprehensive Understanding of Bullying in School Settings. *Educational Psychology Review*, *23*(4), 479–499. doi:10.1007/s10648-011-9171-x

Cascardi, M., Brown, C., Iannarone, M., & Cardona, N. (2014). The Problem With Overly Broad Definitions of Bullying: Implications for the Schoolhouse, the Statehouse, and the Ivory Tower. *Journal of School Violence*, *13*(3), 253–276. doi:10.1080/15388220.2013.846861

CBC News. (2014). Human Rights Tribunal of Alberta sides with bullied teacher. Retrieved September 05, 2014, from http://www.cbc.ca/news/canada/edmonton/human-rights-tribunal-of-alberta-sides-with-bullied-teacher-1.2730483

Chan, J. H. F. (2009). Where Is the Imbalance? *Journal of School Violence*, *8*(2), 177–190. doi:10.1080/15388220802074199

Chang, L., Schwartz, D., Dodge, K. A., & McBride-Chang, C. (2003). Harsh parenting in relation to child emotion regulation and aggression. *Journal of Family Psychology*, *17*(4), 598.

Cillessen, A. H. N., & Rose, A. J. (2005). Understanding popularity in the peer system. *Current Directions in Psychological Science*, *14*(2), 102.

Coie, J. D., Dodge, K. A., & Kupersmidt, J. B. (1990). Peer group behavior and social status. In S. R. Asher & J. D. Coie (Eds.), (pp. 17–59). Cambridge: Cambridge University Press. Retrieved from http://books.google.ca/books?id=OC84AAAAIAAJ

Coie, J. D., Terry, R., Lenox, K., Lochman, J., & Hyman, C. (1995). Childhood peer rejection and aggression as predictors of stable patterns of adolescent disorder. *Development and Psychopathology*, *7*, 697–714.

Cole, D. A. (1990). Relation of social and academic competence to depressive symptoms in childhood. *Journal of Abnormal Psychology*, *99*(4), 422–9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2266218

Cole, J. C., Cornell, D. G., & Sheras, P. (2006). Identification of School Bullies by Survey Methods. *Professional School Counseling*, *9*(4), 305–313.

Cook, C. R., Williams, K. R., Guerra, N. G., & Kim, T. E. (2010). Variability in the prevalence of bullying and victimization: a cross-national and methodological analysis. In S. R. Jimerson, S. M. Swearer, & D. L. Espelage (Eds.), *Handbook of bullying in schools: An international perspective*. (pp. 347–362). New York: Routledge.

Cook, C. R., Williams, K. R., Guerra, N. G., Kim, T. E., & Sadek, S. (2010). Predictors of Bullying and Victimization in Childhood and Adolescence: A Meta-analytic Investigation. *School Psychology Quarterly*, *25*(2), 65–83. doi:10.1037/a0020149

Copeland, W. E., Wolke, D., Angold, A., & Costello, E. J. (2013). Adult psychiatric outcomes of bullying and being bullied by peers in childhood and adolescence. *JAMA Psychiatry (Chicago, Ill.)*, *70*(4), 419–26. doi:10.1001/jamapsychiatry.2013.504

Cornell, D. G., & Bandyopadhyay, S. (2010). The Assessment of Bullying. In S. R. Jimerson, S. M. Swearer, & D. L. Espelage (Eds.), *Handbook of bullying in schools: An international perspective*. (pp. 265–276). New York: Routledge/Taylor & Francis Group.

Cornell, D. G., Sheras, P., & Cole, J. (2006). Assessment of bullying. Mahwah, NJ: Erlbaum.

Côté, S. M., Vaillancourt, T., Barker, E. D., Nagin, D. S., & Tremblay, R. E. (2007). The Joint Development of Physical and Indirect Aggression: Predictors of Continuity and Change During Childhood. *Development and Psychopathology*, *19*(01), 37–55. doi:10.1017/S0954579407070034

Coyne, S. M., Nelson, D. A., & Underwood, M. K. (2011). Aggression in Children. In P. K. Smith & C. H. Hart (Eds.), (2nd ed., pp. 491–509). Chichester West Sussex ;;Malden MA: Wiley-Blackwell.

Craig, W. M., Harel-Fisch, Y., Fogel-Grinvald, H., Dostaler, S., Hetland, J., Simons-Morton, B., … Pickett, W. (2009). A cross-national profile of bullying and victimization among adolescents in 40 countries. *International Journal of Public Health*, *54*(S2), 216–224. doi:10.1007/s00038-009-5413-9

Crick, N. R. (1995). Relational aggression: The role of intent attributions, feelings of distress, and provocation type. *Development and Psychopathology*, *7*(02), 313–322.

Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social information-processing mechanisms in children's social adjustment. *Psychological Bulletin*. *Vol 115(1)*, *115*(1), 74–101. Retrieved from http://csaweb106v.csa.com/ids70/view_record.php?id=1&recnum=13&log=from_toc&SID=sgbgv7cubsraohj7b6902tu4h1&mark_id=cache:0,0,21

Crick, N. R., & Dodge, K. A. (1996). Social Information-Processing Mechanisms in Reactive and Proactive Aggression. *Child Development*, *67*(3), 993–1002. doi:10.1111/1467-8624.ep9704150179

Crick, N. R., Murray-Close, D., Marks, P. E. L., & Mohajeri-Nelson, N. (2009). Aggression and Peer Relationships in School-Age Children: Relational and Physical Aggression in Group and Dyadic Contexts. In K. H. Rubin, W. M. Bukowski, & B. P. Laursen (Eds.), (pp. 287–302). New York: Guilford Press.

Cross, J., & Newman-Gonchar, R. (2004). Data quality in student risk behavior survery and administrator training. *Journal of School Violence*, *3*, 89–109.

Cuadrado-Gordillo, I. (2012). Repetition, Power Imbalance, and Intentionality: Do These Criteria Conform to Teenagers' Perception of Bullying? A Role-Based Analysis. *Journal of Interpersonal Violence*, *27*(10), 1889–1910. doi:10.1177/0886260511431436

Dawes, M., & Xie, H. (2014). The role of popularity goal in early adolescents' behaviors and popularity status. *Developmental Psychology*, *50*(2), 489–97. doi:10.1037/a0032999

De Bruyn, E. H., Cillessen, A. H. N., & Wissink, I. B. (2009). Associations of Peer Acceptance and Perceived Popularity With Bullying and Victimization in Early Adolescence. *The Journal of Early Adolescence*, *30*(4), 543–566. doi:10.1177/0272431609340517

Dijkstra, J. K., Lindenberg, S., & Veenstra, R. (2008). Beyond the Class Norm: Bullying Behavior of Popular Adolescents and its Relation to Peer Acceptance and Rejection. *Journal of Abnormal Child Psychology*, *36*(8), 1289–1299. doi:10.1007/s10802-008-9251-7

Dishion, T. J., Andrews, D. W., & Crosby, L. (1995). Antisocial Boys and Their Friends in Early Adolescence: Relationship Characteristics, Quality, and Interactional Process. *Child Development*, *66*(1), 139–151. doi:10.1111/j.1467-8624.1995.tb00861.x

Dodge, K. A. (1991). The structure and function of reactive an proactive aggression. In D. J. Pepler & K. H. Rubin (Eds.), . Routledge. Retrieved from http://books.google.ca/books?id=4HQPVyg6qMEC

Dodge, K. A., Coie, J. D., & Lynam, D. R. (2006). Aggression and antisocial behavior in youth. *Handbook of Child Psychology*, *3*, 719–778. Retrieved from http://books.google.com/books?id=sg4Qr7qZrXYC

Dollard, J., Doob, L. W., Miller, N. E., Mower, O. H., & Sears, R. R. (1989). *Frustration and aggression* (p. 232). New Haven: Yale University Press. Retrieved from http://books.google.ca/books?id=EIpqAAAAMAAJ

Egan, S. K., & Perry, D. G. (1998). Does Low Self-Regard Invite Victimization? *Developmental Psychology*, *34*(2), 299–309. Retrieved from http://resolver.scholarsportal.info/resolve/00121649/v34i0002/299_dlsiv.xml

Espelage, D. L., Holt, M. K., & Henkel, R. R. (2003). Examination of Peer–Group Contextual Effects on Aggression During Early Adolescence. *Child Development*, *74*(1), 205–220.

Espelage, D. L., & Swearer, S. M. (2003). Research on School Bullying and Victimization: What Have We Learned and Where Do We Go from Here? *School Psychology Review*, *32*(3), 365–383.

Estell, D. B., Cairns, R. B., Farmer, T. W., & Cairns, B. D. (2002). Aggression in Inner-City Early Elementary Classrooms: Individual and Peer-Group Configurations. *Merrill-Palmer Quarterly*, *48*(1), 52–76. Retrieved from http://resolver.scholarsportal.info/resolve/15350266/v48i0001/52_aiieeciapc.xml

Estell, D. B., Farmer, T. W., & Cairns, B. D. (2007). Bullies and victims in rural African American youth: Behavioral characteristics and social network placement. *Aggressive Behavior*, *33*(2), 145–159.

Farmer, T. W., Estell, D. B., Bishop, J. L., O'Neal, K. K., & Cairns, B. D. (2003). Rejected Bullies or Popular Leaders? The Social Relations of Aggressive Subtypes of Rural African American Early Adolescents. *Developmental Psychology*, *39*(6), 992.

Farmer, T. W., Xie, H., & Cairns, B. D. (2007). Social Synchrony, peer networks, and aggression in school. In P. Hawley, T. D. Little, & P. C. Rodkin (Eds.), (pp. 209–233). Mahwah N.J.: L. Erlbaum Associates.

Farrington, D. P. (1991). Childhood aggression and adult violence: Early precursors and life outcomes. In D. J. Pepler & K. H. Rubin (Eds.), (pp. 5–29). Hillsdale NJ: Erlbaum.

Felix, E. D., Sharkey, J. D., & Green, J. G. (2011). Getting precise and pragmatic about the assessment of bullying: The development of the California Bullying Victimization Scale. *Aggressive …*. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/ab.20389/full

Finkelhor, D., Turner, H. a, & Hamby, S. (2012). Let's prevent peer victimization, not just bullying. *Child Abuse & Neglect*, *36*(4), 271–4. doi:10.1016/j.chiabu.2011.12.001

Frisén, A., Hasselblad, T., & Holmqvist, K. (2012). What actually makes bullying stop? Reports from former victims. *Journal of Adolescence*, *35*(4), 981–90. doi:10.1016/j.adolescence.2012.02.001

Frisén, A., Holmqvist, K., & Oscarsson, D. (2008). 13-year-olds' perception of bullying: definitions, reasons for victimisation and experience of adults' response. *Educational Studies*, *34*(2), 105–117. doi:10.1080/03055690701811149

Furlong, M. J., Sharkey, J. D., Felix, E. D., Tanigawa, D., & Green, J. G. (2010). Bullying Assessment: A Call for Increased Percision of Self-Reporting Procedures. In S. R. Jimerson, S. M. Swearer, & D. L. Espelage (Eds.), *Handbook of bullying in schools: An international perspective*. (1st ed., pp. 329–346). New York: Routledge.

Garandeau, C. F., Lee, I. a., & Salmivalli, C. (2014). Differential effects of the KiVa anti-bullying program on popular and unpopular bullies. *Journal of Applied Developmental Psychology*, *35*(1), 44–50. doi:10.1016/j.appdev.2013.10.004

Gilbert, P. (1992). *Depression: the evolution of powerlessness* (p. 580). Psychology Press. Retrieved from http://books.google.ca/books?id=toRX7Zyo_pMC

Gilbert, P. (2000). The relationship of shame, social anxiety and depression: the role of the evaluation of social rank. *Clinical Psychology & Psychotherapy*, *7*(3), 174–189. doi:10.1002/1099-0879(200007)7:3<174::AID-CPP236>3.0.CO;2-U

Gilbert, P., & Allan, S. (1994). Assertiveness, submissive behaviour and social comparison. *British Journal of Clinical Psychology*, *33*, 295–306.

Gilbert, P., Broomhead, C., Irons, C., McEwan, K., Bellew, R., Mills, A., … Knibb, R. (2007). Development of a striving to avoid inferiority scale. *British Journal of Social Psychology*, *46*(3), 633–648. doi:10.1348/014466606X157789

Gladden, R. M., Vivolo-Kantor, A., Hamburder, M., & Lumpkin, C. (2014). *Bullying Surveillance Among Youths: Uniform Definitions for Public Health and Recommended Data Elements, Version 1.0*. Atlanta, GA: National Center for Injusry Prevention and Control, Centers for Disease Control and Prevention and U.S. Department of Education. Retrieved from https://calio.dspacedirect.org/handle/11212/1303

Glew, G. M., Fan, M.-Y., Katon, W., Rivara, F. P., & Kernic, M. A. (2005). Bullying, psychosocial adjustment, and academic performance in elementary school. *Archives of Pediatrics & Adolescent Medicine*, *159*(11), 1026–1031. doi:10.1001/archpedi.159.11.1026

Goodenow, C. (1993). The psychological sense of school membership among adolescents: Scale Development and Educational Correlates. *Psychology in the Schools*, *30*(1), 79–90.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: a research note. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *38*(5), 581–586. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9255702

Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child and Adolescent Psychiatry*, *40*(11), 1337–45. doi:10.1097/00004583-200111000-00015

Goodwin, C. J. (2003). *Research in Psychology: Methods and Design* (3rd Editio.). John Wiley and Sons.

Graham, S., & Juvonen, J. (1998). Salf-Blame and peer victimization in middle school: An attributional analysis. *Developmental Psychology*, *34*, 587–599.

Green, J. G., Felix, E. D., Sharkey, J. D., Furlong, M. J., & Kras, J. E. (2013). Identifying bully victims: definitional versus behavioral approaches. *Psychological Assessment*, *25*(2), 651–7. doi:10.1037/a0031248

Greif, J. L., & Furlong, M. J. (2006). The Assessment of School Bullying. *Journal of School Violence*, *5*(3), 33–50. doi:10.1300/J202v05n03_04

Hagborg, W. J. (1994). An Exploration of School Membership Among Middle- and High-School Students. *Journal of Psychoeducational Assessment*, *12*(4), 312–323. Retrieved from http://jpa.sagepub.com/content/12/4/312.short

Hagborg, W. J. (1998). An investigation of a brief measure of school membership. *Adolescence*, *33*(130), 461–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9706331

Hamburder, M., Basile, K. C., & Vivolo, A. M. (2011). *Measuring bullying victimization, perpetration and bystander experiences: A compendium of assessment tools*. Atlanta, GA.

Hanish, L. D., Bradshaw, C. P., Espelage, D. L., Rodkin, P. C., Swearer, S. M., & Horne, A. (2013). Looking Toward the Future of Bullying Research: Recommendations for Research and Funding Priorities. *Journal of School Violence*, *12*(3), 283–295. doi:10.1080/15388220.2013.788449

Hart, E. J., & Ostrov, J. M. (2013). Functions of aggressive behavior and future functional impairment. *Early Childhood Research Quarterly*, *28*(4), 683–691. doi:10.1016/j.ecresq.2013.05.005

Hawker, D. S. J., & Boulton, M. J. (2000). Twenty Years' Research on Peer Victimization and Psychosocial Maladjustment: A Meta-analytic Review of Cross-sectional Studies. *Journal of Child Psychology and Psychiatry*, *41*(4), 441–455. Retrieved from http://resolver.scholarsportal.info/resolve/00219630/v41i0004/441_tyropvamrocs.xml

Hawkins, D. L., Pepler, D. J., & Craig, W. M. (2001). Peer Interventions in Playground Bullying. *Social Development*, *28*, 450–470.

Hawley, P. (1999). The ontogenesis of social dominance: A strategy-based evolutionary perspective. *Developmental Review*, *19*, 97–132.

Hawley, P. (2003). Prosocial and Coercive Configurations of Resource Control in Early Adolescence: A Case for the Well-Adapted Machiavellian. *Merrill-Palmer Quarterly*, *49*(3), 279–309. doi:10.1353/mpq.2003.0013

Hawley, P., Little, T. D., & Rodkin, P. C. (Eds.). (2007). *Aggression and adaptation : the bright side to bad behavior*. Mahwah N.J.: L. Erlbaum Associates.

Hellström, L., Beckman, L., & Hagquist, C. (2013). Self-Reported Peer Victimization: Concordance and Discordance Between Measures of Bullying and Peer Aggression Among Swedish Adolescents. *Journal of School Violence*, *12*(4), 395–413. doi:10.1080/15388220.2013.825626

Henry, D. (2006). Associations between peer nominations, teacher ratings, self-reports, and observations of malicious and disruptive behavior. *Assessment*, *13*, 241–252.

Hodges, E. V. E., Boivin, M., Vitaro, F., & Bukowski, W. M. (1999). The Power of Friendship: Protection Against an Escalating Cycle of Peer Victimization. *Developmental Psychology*, *35*(1), 94–101. Retrieved from http://resolver.scholarsportal.info/resolve/00121649/v35i0001/94_tpofpaaecopv.xml

Hodges, E. V. E., & Perry, D. G. (1999). Personal and interpersonal antecedents and consequences of victimization by peers. *Journal of Personality and Social Psychology*, *76*(4), 677–685. doi:10.1037/0022-3514.76.4.677

Hoff, K. E., Reese-Weber, M., Joel Schneider, W., & Stagg, J. W. (2009). The association between high status positions and aggressive behavior in early adolescence. *Journal of School Psychology*, *47*(6), 395–426.

Houghton, S. J., Nathan, E., & Taylor, M. (2012). To Bully or Not to Bully, That Is Not the Question: Western Australian Early Adolescents' in Search of a Reputation. *Journal of Adolescent Research*, *27*(4), 498–522. doi:10.1177/0743558411432638

Huesmann, L. R. (1988). An information processing model for the development of aggression. *Aggressive Behavior*, *14*(May 1987), 13–24. Retrieved from http://deepblue.lib.umich.edu/handle/2027.42/83387

Huesmann, 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. doi:10.1037//0022-3514.72.2.408

Hunter, S. C., Boyle, J. M. ., & Warden, D. (2007). Perceptions and correlates of peer-victimization and bullying. *British Journal of Educational Psychology*, *77*(4), 797–810.

Irons, C., & Gilbert, P. (2005). Evolved mechanisms in adolescent anxiety and depression symptoms: the role of the attachment and social rank systems. *Journal of Adolescence*, *28*(3), 325–341. Retrieved from http://resolver.scholarsportal.info/resolve/01401971/v28i0003/325_emiaaataasrs.xml

Jimerson, S. R., Swearer, S. M., & Espelage, D. L. (Eds.). (2010a). *Handbook of bullying in schools: An international perspective*. *Handbook of bullying in schools: An international perspective*. New York: Routledge/Taylor & Francis Group. Retrieved from http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc6&NEWS=N&AN=2010-06797-000

Jimerson, S. R., Swearer, S. M., & Espelage, D. L. (2010b). International Scholarship Advances Science and Practice Addressing Bullying in Schools. In S. R. Jimerson, S. M. Swearer, & D. L. Espelage (Eds.), *Handbook of bullying in schools: An international perspective*. (pp. 1–6). New York: Routledge.

Juvonen, J., & Graham, S. (2001). Self-views versus peer perceptions of victim status among early adolescents. In J. Juvonen & S. Graham (Eds.), (1st ed., pp. 105–124). The Guilford Press.

Juvonen, J., & Graham, S. (2014). Bullying in schools: the power of bullies and the plight of victims. *Annual Review of Psychology*, *65*, 159–85. doi:10.1146/annurev-psych-010213-115030

Juvonen, J., Nishina, A., & Graham, S. (2000). Peer harassment, psychological adjustment, and school functioning in early adolescence. *Journal of Educational Psychology*. *Vol 92(2)*, *92*(2), 349–359. Retrieved from http://csaweb110v.csa.com/ids70/view_record.php?id=2&recnum=19&log=from_toc&SID=himo7f6q1rbfg2uahqifahf0n5&mark_id=cache:1,0,34

Kert, A. S., Codding, R. S., Tryon, G. S., & Shiyko, M. (2010). Impact of the word "bully" on the reported rate of bullying behavior. *Psychology in the Schools*, *47*, 193–204.

Kessler, R. C. (1997). The effects of stressful life events on depression. *Annual Review of Psychology*, *48*(1), 191–214.

Kikas, E., Peets, K., Tropp, K., & Hinn, M. (2009). Associations Between Verbal Reasoning , Normative Beliefs About Aggression , and Different Forms of Aggression. *Journal of Research on Adolescence*, *19*(1), 137–149. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1532-7795.2009.00586.x/full

Kochenderfer, B. J., & Ladd, G. W. (1996). Peer Victimization: Cause or Consequence of School Maladjustment? *Child Development*, *67*(4), 1305–1317. doi:10.1111/1467-8624.ep9704040996

Kochenderfer-Ladd, B., & Wardrop, J. L. (2001). Chronicity and Instability of Children's Peer Victimization Experiences as Predictors of Loneliness and Social Satisfaction Trajectories. *Child Development*, *72*(1), 134–151. doi:10.1111/1467-8624.00270

Kraha, A., Turner, H. a, Nimon, K., Zientek, L. R., & Henson, R. K. (2012). Tools to support interpreting multiple regression in the face of multicollinearity. *Frontiers in Psychology*, *3*(March), 44. doi:10.3389/fpsyg.2012.00044

Ladd, G. W. (2005). *Children's Peer Relations and Social Competence: A Century of Progress*. Yale University Press. Retrieved from http://www.amazon.com/dp/0300106432

Ladd, G. W., & Kochenderfer-Ladd, B. (2002). Identifying victims of peer aggression from early to middle childhood: Analysis of cross-informant data for concordance, estimation of relational adjustment, prevalence of victimization, and characteristics of identified victims. *Psychological Assessment*, *14*(1), 74.

Ladd, G. W., & Troop-Gordon, W. (2003). The role of chronic peer difficulties in the development of children's psychological adjustment problems. *Child Development*, *74*(5), 1344–1367.

LaFontana, K. M., & Cillessen, A. H. N. (2002). Children's perceptions of popular and unpopular peers: A multimethod assessment. *Developmental Psychology*, *38*(5), 635.

Land, D. (2003). Teasing apart secondary students' conceptualizations of peer teasing, bullying and sexual harassment. *School Psychology International*, *24*(2), 147–165. doi:10.1177/0143034303024002002

Laurent, J., Catanzaro, S. J., Joiner, T. E., Rudolph, K. D., Potter, K. I., Lambert, S., … Gathright, T. (1999). A measure of Positive and Negative Affect for Children: Scale Development and Preliminary Validation. *Psychological Assessment*, *11*(3), 326–338.

Lease, A. M., Musgrove, K. T., & Axelrod, J. L. (2002). Dimensions of social status in preadolescent peer groups: Likability, perceived popularity, and social dominance. *Social Development*, *11*(4), 508–533.

Li, Y., & Wright, M. F. (2014). Adolescents' social status goals: relationships to social status insecurity, aggression, and prosocial behavior. *Journal of Youth and Adolescence*, *43*(1), 146–60. doi:10.1007/s10964-013-9939-z

Messick, S. (1995). Validity of Psychological Assessment: Validation of Inferences From Persons' Responses and Performances as Scientific Inquiry Into Score Meaning. *American Psychologist*, *50*(9), 741–749. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:No+Title#0

Miles, S. B., & Stipek, D. (2006). Contemporaneous and Longitudinal Associations Between Social Behavior and Literacy Achievement in a Sample of Low-Income Elementary School Children. *Child Development*, *77*(1), 103–117. Retrieved from http://resolver.scholarsportal.info/resolve/00093920/v77i0001/103_calabssolesc.xml

Moffitt, T. E. (1993). Adolescence-limited and life-course-persistent antisocial behavior: A developmental taxonomy. *Psychological Review*. *Vol 100(4)*, *100*(4), 674–701. Retrieved from http://csaweb108v.csa.com/ids70/view_record.php?id=1&recnum=15&log=from_toc&SID=tjdt8m1rebgtcj5318qvvsptt6&mark_id=cache:0,0,28

Moffitt, T. E. (2005). The New Look of Behavioral Genetics in Developmental Psychopathology: Gene-Environment Interplay in Antisocial Behaviors. *Psychological Bulletin*, *131*, 533–554. doi:10.1037/0033-2909.131.4.533

Monks, C. P., & Smith, P. K. (2006). Definitions of bullying: Age differences in understanding of the term, and the role of experience. *British Journal of Developmental Psychology*, *24*(4), 801–821.

Morrow, M. T., Hubbard, J. A., McAuliffe, M. D., Rubin, R. M., & Dearing, K. (2006). Childhood aggression, depressive symptoms, and peer rejection: The mediational model revisited. *International Journal of Behavioral Development*, *30*(3), 240–248. doi:10.1177/0165025406066757

Mynard, H., & Jospeh, S. (2000). Development of the multidimensional peer-victimization scale. *Agressive Behavior*, *26*, 169–178.

Nagin, D. S., & Tremblay, R. E. (1999). Trajectories of Boys' Physical Aggression, Opposition, and Hyperactivity on the Path to Physically Violent and Nonviolent Juvenile Delinquency. *Child Development*, *70*(5), 1181–1196. doi:10.1111/1467-8624.00086

Nansel, T. R., Overpeck, M., Pilla, R. S., Ruan, J., Simons-Morton, B., & Scheidt, P. (2001). Bullying behaviors among US Youth: Prevalence and association with psychosocial adjustment. *Journal of the American Medical Association*, *285*(16), 2094–2100.

Nathans, L. L., Oswald, F. L., & Nimon, K. (2012). Interpreting Multiple Linear Regression : A Guidebook of Variable Importance. *Practical Assessment, Research & Evaluation*, *17*(9), 1–19.

Newcomb, A. F., & Bukowski, W. M. (1983). Social impact and social preference as determinants of children's peer group status. *Developmental Psychology*, *19*(6), 856.

Newcomb, A. F., Bukowski, W. M., & Pattee, L. (1993). Children's peer relations: A meta-analytic review of popular, rejected, neglected, controversial, and average sociometric status. *Psychological Bulletin*. *Vol 113(1)*, *113*(1), 99–128. Retrieved from http://csaweb113v.csa.com/ids70/view_record.php?id=1&recnum=11&log=from_toc&SID=ucuqcp1rngreo9tanc5q83evr7&mark_id=cache:0,0,23

Nishina, A., & Juvonen, J. (2005). Daily reports of witnessing and experiencing peer harassment in middle school. *Child Development*, *76*(2), 435–450.

Norušis, M. (2012). *IBM SPSS Statistics 19 Statistical Procedures Companion*. Upper Saddle River, NJ: Prentice Hall. Retrieved from http://www.chapters.indigo.ca/books/IBM-SPSS-Statistics-Statistical-Procedures-Marija-Norusis-Inc-Spss/9780321748423-item.html

Olthof, T., Goossens, F. A., Vermande, M. M., Aleva, E. A., & van der Meulen, M. (2011). Bullying as strategic behavior: relations with desired and acquired dominance in the peer group. *Journal of School Psychology*, *49*(3), 339–359. doi:10.1016/j.jsp.2011.03.003

Olweus, D. (1996). *The Revised Olwues Bully/Victim Questionnaire*. Bergen, Norway: Mimeo, Reserach Centre for Health Promotion (HEMIL), UIniversity of Bergen.

Olweus, D. (2010). Understanding and researching bullying: Some critical issues. In S. R. Jimerson, S. M. Swearer, & D. L. Espelage (Eds.), *Handbook of bullying in schools: An international perspective*. (1st ed., pp. 9–34). New York: Routledge.

Olweus, D. (2013). School bullying: development and some important challenges. *Annual Review of Clinical Psychology*, *9*, 751–80. doi:10.1146/annurev-clinpsy-050212-185516

Parker, J. G., & Asher, S. R. (1987). Peer relations and later personal adjustment: Are low-accepted children at risk? *Psychological Bulletin*. *Vol 102(3)*, *102*(3), 357–389. Retrieved from http://csaweb116v.csa.com/ids70/view_record.php?id=4&recnum=12&log=from_toc&SID=mp9kv70kddqn4qrge443ulq3o7&mark_id=cache:3,0,24

Peeters, M., Cillessen, A. H. N., & Scholte, R. H. J. (2009). Clueless or Powerful? Identifying Subtypes of Bullies in Adolescence. *Journal of Youth and Adolescence*, *39*(9), 1041–1052. doi:10.1007/s10964-009-9478-9

Pellegrini, A. D., & Long, J. D. (2002). A longitudinal study of bullying, dominance, and victimization during the transition from primary school through secondary school. *British Journal of Developmental Psychology*, *20*(2), 259–280.

Pepler, D. J., Craig, W. M., Connolly, J. A., Yuile, A., & Mcmaster, L. (2005). A Developmental Perspective on Bullying, *384*(April), 376–384. doi:10.1002/ab

Pepler, D. J., Craig, W. M., & Roberts, W. R. (1998). Observations of aggressive and nonaggressive children on the school playground. *Merrill-Palmer Quarterly*, *1*(44), 55–76.

Perry, D. G., Hodges, E. V. E., & Egan, S. K. (2001). Determinants of chronic victimization by peers: A review and a new model of family influence. In J. Juvonen & S. Graham (Eds.), (1st ed., pp. 73–104). The Guilford Press.

PREVNet. (2014). Bullying: Definitions. Retrieved September 07, 2014, from http://www.prevnet.ca/research/fact-sheets/bullying-definitions

Puissant, S. P., Gauthier, J.-M., & Van Oirbeek, R. (2011). The contribution of social rank and attachment theory to depression in a non clinical sample of adolescents. *The Spanish Journal of Psychology*, *14*(2), 832–42. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/22059328

Putallaz, M., Grimes, C. L., Foster, K. J., Kupersmidt, J. B., Coie, J. D., & Dearing, K. (2007). Overt and relational aggression and victimization: Multiple perspectives within the school setting. *Journal of School Psychology*, *45*(5), 523–547. doi:10.1016/j.jsp.2007.05.003

Raine, A. (2008). From Genes to Brain to Antisocial Behavior. *Current Directions in Psychological Science*, *17*(5), 323 –328. doi:10.1111/j.1467-8721.2008.00599.x

Reijntjes, A., Vermande, M., Goossens, F. A., Olthof, T., van de Schoot, R., Aleva, L., & van der Meulen, M. (2013). Developmental trajectories of bullying and social dominance in youth ☙. *Child Abuse & Neglect*, *37*(4), 224–34. doi:10.1016/j.chiabu.2012.12.004

Reijntjes, A., Vermande, M., Olthof, T., Goossens, F. A., van de Schoot, R., Aleva, L., & van der Meulen, M. (2013). Costs and benefits of bullying in the context of the peer group: A three wave longitudinal analysis. *Journal of Abnormal …*, (41), 1217–1229. doi:10.1007/s10802-013-9759-3

Rigby, K. (2000). Effects of peer victimization in schools and perceived social support on adolescent well-being. *Journal of Adolescence*, *23*(1), 57–68. Retrieved from http://resolver.scholarsportal.info/resolve/01401971/v23i0001/57_eopvispssoaw.xml

Rigby, K. (2003). Consequences of Bullying in Schools. *Canadian Journal of Psychiatry*, *48*(9), 583.

Rigby, K. (2012). Bullying in Schools: Addressing Desires, Not Only Behaviours. *Educational Psychology Review*, *24*(2), 339–348. doi:10.1007/s10648-012-9196-9

Ringrose, J., & Renold, E. (2010). Normative cruelties and gender deviants: the performative effects of bully discourses for girls and boys in school. *British Educational Research Journal*, *36*(4), 573–596. doi:10.1080/01411920903018117

Rodkin, P. C., & Berger, C. (2008). Who bullies whom? Social status asymmetries by victim gender. *International Journal of Behavioral Development*, *32*(6), 473–485. doi:10.1177/0165025408093667

Rodkin, P. C., Farmer, T. W., Pearl, R., & Acker, R. V. (2006). They're cool: Social status and peer group supports for aggressive boys and girls. *Social Development*, *15*(2), 175–204.

Rodkin, P. C., Farmer, T. W., Pearl, R., & Van Acker, R. (2000). Heterogeneity of popular boys: Antisocial and prosocial configurations. *Developmental Psychology*, *36*(1), 14.

Rodkin, P. C., & Wilson, T. (2007). Aggression and adaptation: Psychological record, educational promise. In P. Hawley, T. D. Little, & P. C. Rodkin (Eds.), (pp. 235–267). Mahwah N.J.: L. Erlbaum Associates.

Rosch, E., & Mervis, C. G. (1975). Family Resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *75*, 573–605.

Rose, A. J., & Swenson, L. P. (2009). Do perceived popular adolescents who aggress against others experience emotional adjustment problems themselves?. *Developmental Psychology*, *45*(3), 868.

Rose, A. J., Swenson, L. P., & Waller, E. M. (2004). Overt and relational aggression and perceived popularity: developmental differences in concurrent and prospective relations. *Developmental Psychology*, *40*(3), 378.

Rosen, P. ., Milich, R., & Harris, M. (2007). Victims of their own cognitions: Implicit social cognitions, emotional distress, and peer victimization. *Journal of Applied Developmental Psychology*, *28*, 211–226.

Rowell, T. E. (1974). The concept of social dominance. *Behavioral Biology*, *11*(2), 131–154. doi:10.1016/S0091-6773(74)90289-2

Rubin, K. H., Bukowski, W. M., & Laursen, B. P. (Eds.). (2009). *Handbook of peer interactions, relationships, and groups* (p. 673). Guilford Press. Retrieved from http://books.google.com/books?id=ZxqG6O28r1QC

Rule, B. G., Nesdale, A. R., & McAra, M. J. (1974). Children's reactions to information about the intentions underlying an aggressive act. *Child Development*, 794–798.

Sacco, D. T., Silbaugh, K., Corredor, F., Casey, J., & Doherty, D. (2012). An Overview of State Anti-Bullying Legislation and Other Related Laws. *The Kinder and Braver World Project Research Series*.

Salmivalli, C. (2005). Consequences of school bullying and violence. Retrieved from http://www.oecd.org/dataoecd/28/6/33866604.pdf

Salmivalli, C. (2010). Bullying and the peer group: A review. *Aggression and Violent Behavior*, *15*(2), 112–120.

Salmivalli, C., & Peets, K. (2009). Bullies, victims and bully-Victim Relationships in middle childhood and early adolescence. In K. H. Rubin, W. M. Bukowski, & B. P. Laursen (Eds.), (pp. 322–340). Guilford Press.

Salmivalli, C., Peets, K., & Hodges, E. V. E. (2011). Bullying. In P. K. Smith & C. H. Hart (Eds.), (2nd ed., pp. 510–528). Chichester West Sussex ;;Malden MA: Wiley-Blackwell.

Salmon, G., James, A., & Smith, D. M. (1998). Bullying in schools: self reported anxiety, depression, and self esteem in secondary school children. *BMj*, *317*(7163), 924.

Schäfer, M., Werner, N. E., & Crick, N. R. (2002). A comparison of two approaches to the study of negative peer treatment: General victimization and bully/victim problems among German schoolchildren. *British Journal of Developmental Psychology*, *20*(2), 281–306.

Scholte, R. H. J., Burk, W. J., & Overbeek, G. (2013). Divergence in self- and peer-reported victimization and its association to concurrent and prospective adjustment. *Journal of Youth and Adolescence*, *42*(12), 1789–800. doi:10.1007/s10964-012-9896-y

Schwartz, D. (2000). Subtypes of victims and aggressors in children's peer groups. *Journal of Abnormal Child Psychology*, *28*(2), 181–192.

Schwartz, D., Dodge, K. A., Coie, J. D., Hubbard, J. A., Cillessen, A. H. N., Lemerise, E. A., & Bateman, H. (1998). Social-cognitive and behavioral correlates of aggression and victimization in boys' play groups. *Journal of Abnormal Child Psychology*, *26*(6), 431–440.

Sercombe, H., & Donnelly, B. (2013). Bullying and agency: definition, intervention and ethics. *Journal of Youth Studies*, *16*(4), 491–502. doi:10.1080/13676261.2012.725834

Sijtsema, J. J., Veenstra, R., Lindenberg, S., & Salmivalli, C. (2009). Empirical test of bullies' status goals: Assessing direct goals, aggression, and prestige. *Aggressive Behavior*, *35*(1), 57–67.

Smith, P. K., & Brain, P. (2000). Bullying in Schools: Lessons From Two Decades of Research. *Aggressive Behavior*, *26*, 1–9.

Smith, P. K., Cowie, H., Olafsson, R. F., Liefooghe, A. P. D., Almeida, A., Araki, H., … Wenxin, Z. (2002). Definitions of bullying: a comparison of terms used, and age and gender differences, in a fourteen-country international comparison. *Child Development*, *73*(4), 1119–33. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12146737

Smith, P. K., del Barrio, C., & Tokunaga, R. S. (2013). Definitions of Bullying and Cyberbullying. In S. Bauman, D. Cross, & J. Walker (Eds.), *Principles of Cyberbullying Research: Definitions, Measures, and Methodology* (pp. 26–40). New York: Routledge.

Smith, P. K., & Monks, C. P. (2008). Concepts of bullying: developmental and cultural aspects. *International Journal of Adolescent Medicine and Health*, *20*(2), 101–12. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/18714549

Solberg, M. E., & Olweus, D. (2003). Prevalence estimation of school bullying with the Olweus Bully/Victim Questionnaire. *Aggressive Behavior*, *29*(3), 239–268. doi:10.1002/ab.10047

Stassen Berger, K. (2007). Update on bullying at school: Science forgotten? *Developmental Review*, *27*(1), 90–126. doi:doi: 10.1016/j.dr.2006.08.002

Statistics Canada. (2006). Visible minority groups, percentage distribution, for Canada, provinces and territories - 20% sample data.

Stuart-Cassel, V., Bell, A., & Springer, J. F. (2011). Analysis of State Bullying Laws and Policies. *Office of Planning, Evaluation and Policy Development, US Department of Education*.

Sutton, J., Smith, P. K., & Swettenham, J. (1999). Bullying and "Theory of Mind": A Critique of the "Social Skills Deficit" View of Anti-Social Behaviour. *Social Development*, *8*(1), 117–127.

Swearer, S. M., Espelage, D. L., Vaillancourt, T., & Hymel, S. (2010). What Can Be Done About School Bullying?: Linking Research to Educational Practice. *Educational Researcher*, *39*(1), 38–47. doi:10.3102/0013189X09357622

Swearer, S. M., Siebecker, A. B., Johnsen-Frerichs, L. A., & Wang, C. (2010). Assessment of Bullying/Victimization: The Problem of Comparability Across Studies adn Across Methodologies. In S. R. Jimerson, S. M. Swearer, & D. L. Espelage (Eds.), *Handbook of bullying in schools: An international perspective*. (pp. 305–328).

Swearer, S. M., Turner, R. K., Givens, J. E., & Pollack, W. S. (2008). "You're so gay!": Do different forms of bullying matter for adolescent males? *School Psychology Review*, *37*, 160–173.

Tabachnick, B., & Fidell, L. (2007). *Using multivariate statistics* (5th ed.). Boston: Allyn & Bacon. Retrieved from http://www.amazon.ca/dp/0205459382

Terry, R. (2000). Recent Advances in Measurement Theory and the Use of Sociometric Techniques. *New Directions for Child & Adolescent Development*, *2000*(88), 27–53.

Thomas, H. J., Connor, J. P., & Scott, J. G. (2014). Integrating Traditional Bullying and Cyberbullying: Challenges of Definition and Measurement in Adolescents – a Review. *Educational Psychology Review*. doi:10.1007/s10648-014-9261-7

Thomas, K. K., & Bowker, J. C. (2013). An Investigation of Desired Friendships During Early Adolescence. *The Journal of Early Adolescence*, *33*(6), 867–890. doi:10.1177/0272431612469725

Thornberg, R. (2013). Distressed Bullies, Social Positioning and Odd Victims: Young People's Explanations of Bullying. *Children & Society*, n/a–n/a. doi:10.1111/chso.12015

Toblin, R. L., Schwartz, D., Hopmeyer Gorman, A., & Abou-ezzeddine, T. (2005). Social-cognitive and behavioral attributes of aggressive victims of bullying. *Journal of Applied Developmental Psychology*, *26*(3), 329–346.

Tomada, G., & Schneider, B. H. (1997). Relational aggression, gender, and peer acceptance: Invariance across culture, stability over time, and concordance among informants. *Developmental Psychology*. *Vol 33(4)*, *33*(4), 601–609. Retrieved from http://csaweb110v.csa.com/ids70/view_record.php?id=4&recnum=9&log=from_toc&SID=bsq1lhhu3ccvfk0lbq53ugu3o4&mark_id=cache:0,0,26

Vaillancourt, T., & Hymel, S. (2006). Aggression and social status: the moderating roles of sex and peer-valued characteristics. *Aggressive Behavior*, *32*(4), 396–408. doi:10.1002/ab.20138

Vaillancourt, T., Hymel, S., & McDougall, P. (2003). Bullying Is Power: Implications for School-Based Intervention Strategies. *Journal of Applied School Psychology*, *19*(2), 157–176. doi:10.1300/J008v19n02_10

Vaillancourt, T., McDougall, P., Hymel, S., Krygsman, A., Miller, J., Stiver, K., & Davis, C. (2008). Bullying: Are researchers and children/youth talking about the same thing? *International Journal of Behavioral Development*, *32*(6), 486–495. doi:10.1177/0165025408095553

Vaughn, B. E., & Santos, A. J. (2007). An evolutionary/ecological account of aggressive behavior and trait aggression in human children and adolescents. In P. Hawley, T. D. Little, & P. C. Rodkin (Eds.), (pp. 31–63). Mahwah N.J.: L. Erlbaum Associates.

Veenstra, R., Lindenberg, S., De Winter, A. F., Zijlstra, B. J. H., Verhulst, F. C., & Ormel, J. (2007). The Dyadic Nature of Bullying and Victimization: Testing a Dual-Perspective Theory. *Child Development*, *78*(6), 1843–1854. doi:10.1111/j.1467-8624.2007.01102.x

Veldman, D. J., & Sheffield, J. R. (2010). Educational and Psychological Measurement. doi:10.1177/001316447903900114

Vernberg, E. M. (1990). Psychological adjustment and experiences with peers during early adolescence: Reciprocal, incidental, or unidirectional relationships? *Journal of Abnormal Child Psychology*, *18*, 187–198. doi:10.1007/BF00910730

Vitaro, F., Brendgen, M., & Barker, E. D. (2006). Subtypes of aggressive behaviors: A developmental perspective. *International Journal of Behavioral Development*, *30*(1), 12–19. doi:10.1177/0165025406059968

Volk, A. a., Camilleri, J. a, Dane, A. V, & Marini, Z. a. (2012). Is adolescent bullying an evolutionary adaptation? *Aggressive Behavior*, *38*(3), 222–38. doi:10.1002/ab.21418

Walcott, C. M., Upton, A., Bolen, L. M., & Brown, M. B. (2008). Associations between peer-perceived status and aggression in young adolescents. *Psychology in the Schools*, *45*(6), 550–561.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and Validation of Brief Measures of Positive and Negative Affect: The PANAS Scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070.

Werner, N. E., & Nixon, C. L. (2005). Normative Beliefs and Relational Aggression: An Investigation of the Cognitive Bases of Adolescent Aggressive Behavior. *Journal of Youth and Adolescence*, *34*(3), 229–243. doi:10.1007/s10964-005-4306-3

Ybarra, M. L., Espelage, D. L., & Mitchell, K. J. (2014). Differentiating Youth Who Are Bullied From Other Victims of Peer-Aggression: The Importance of Differential Power and Repetition. *The Journal of Adolescent Health : Official Publication of the Society for Adolescent Medicine*, 1–8. doi:10.1016/j.jadohealth.2014.02.009

You, S., Furlong, M. J., Felix, E. D., Sharkey, J. D., Tanigawa, D., & Green, J. G. (2008). Relations among school connectedness, hope, life satisfaction, and bully victimization. *Psychology in the Schoolsi*, *45*(5), 446–460. doi:10.1002/pits

Appendix A: Ethical Approval Notice

File Number: 04-10-05 **Date (mm/dd/yyyy):** 10/18/2010

Université d'Ottawa University of Ottawa
Bureau d'éthique et d'intégrité de la recherche Office of Research Ethics and Integrity

Ethics Approval Notice

Social Science and Humanities REB

Principal Investigator / Supervisor / Co-investigator(s) / Student(s)

First Name	Last Name	Affiliation	Role
Darcy	Santor	Social Sciences / School of Psychology	Supervisor
Margaret	McGugan	Social Sciences / School of Psychology	Student Researcher

File Number: 04-10-05

Type of Project: PhD Thesis

Title: You and Your School

Approval Date (mm/dd/yyyy)	Expiry Date (mm/dd/yyyy)	Approval Type
10/18/2010	10/17/2011	Ia

(Ia: Approval, Ib: Approval for initial stage only)

Special Conditions / Comments:
N/A

550, rue Cumberland 550 Cumberland Street
Ottawa (Ontario) K1N 6N5 Canada Ottawa, Ontario K1N 6N5 Canada
(613) 562-5841 • Téléc./Fax (613) 562-5338
http://www.research.uottawa.ca/ethics/index.html
http://www.recherche.uottawa.ca/deontologie/index.html

Appendix B: Classroom Script

Hello everyone. I am going to start by making sure that we have the right people in the room. If your name is not on this list, please raise your hand. (Read list)

Alright, it seems like we have the right people in the room. Lets get started. If you have questions about how to do the survey as we are going, please raise your hand.

Lets start by reading the first page of the survey. It says (read first page, pause, and then collect coding page).

The first page of questions, marked with a number 2 at the top, asks you some questions about yourself. Please fill these in now. (1 minute).

Now turn to the next page, marked 3 at the top. Here we ask "Are there some kids in your school who hang around together a lot? Who are they?" After the question there are some instructions. It asks you to think of different groups that hand around together and then identify them using the list below. Lets do this together. In one of the columns below, find your own name. Put a checkmark in the box next to it, and then think about who you hang around with a lot. Put checkmarks next to the names of those people. My lists have made up names on them, but say my name was Lisa L, I would but a check mark next to it here. I hang around a lot with Jake J and Nadia N, so I would also put checkmarks next to their names. There is not right number of names to check off. Check as many or as few names as you need to show who you hang around with. After you have finished identifying the group that you hang around with, go to a different column and check off the names of other people who hang around together a lot. Each group should be in a different column. If you look on the next couple of pages you will see there are more columns if you need them. Keep going until you have listed all the groups that you can think of, and then wait for my next instruction.

(pause)

OK, please turn to page 6. If you're not done with the groups you can come back to it later. Page 6 looks like this (show). Here we want to know "**Who are the most and least popular people in your class?** Think about who the most popular people in your class are. Once you have thought of who they are, check the box next to their names in the list on the left below. You can name as many people as you like. Once you have done this repeat the process except this time think of the people who are the least popular. Check the box next to their name in the list on the right.

The question on page 7 is similar to the last one, except this time we want to know **Who seems to have a lot of power over others?** and **Who does not seem to have a lot of power over others?** Check the box next to the names of powerful student in the list on the left and not powerful students on the list on the right.

Ok, now turn to the page 8 in your booklet. Let's read this definition. (read)

There was a lot of information contained in that definition, so let me summarise for you. Bullying is intentionally hurtful, it happens repeatedly, and it is hard to escape because the person doing the bullying is more powerful than the person they bully.

Now turn to page 8. On this page we want to know "Are there some people in your class who bully other people? Who are they and who do they bully?" Thinking about the definition that you just read, I want you to think of a person who bullies. You're going to check their name in the left column like this. Then you're going to think about who they bully, and check the names of the people who bully in the other column (demo). If you can't think of anyone on this list who is a bully, check this box at the bottom next to "no one on this list is a bully".

This page should have only one bully, so only one person's name should be check in the "who is a bully column". They may bully more than one person so you may have a number of names checked in the "whom do they bully column" like I did in the example, but it's also possible there may be only one name checked. After you do this, you can use the following pages to identify other bullies. There is space to identify 9 bullies. You don't need to use every page, just enough to identify all the people on the list who you think bully other people. Work on answering this question now.

(Pause)

Alright, when you are done that, please turn to page 18. You'll see that this page has more columns. This time we want to know **who in your class hits, kicks, shoves, or pushes others or takes others things by force**. Like when we asked about bullies, you'll think of one person who does this at a time and mark their name in the left hand column. Then you'll think of the people that they do this to and mark their names in the next column. Then we want you to answer 3 questions for every person that this is done to. The questions are (read). Just like last time there should only be one names checked in the first column, but you can check as many or as few names as you want in the second column.

In my example the person I thought of who hits, kicks shoves or pushes others or takes their thing by force is the first person on the list, Amanada A. She does these things to Grant G, so I checked his names. I think Amanda means to hurt grant so I check yes in the box under that question. Amanda also seems more powerful than Grant, so I checked yes

there too, but I can only remember this type of thing happening once, so I say no to the question “does this type of thing happen repeatedly”. Amanda also hit Nadia N once, so I check her name, and I think it was purposely hurtful, so I check yes in the first column, but I check no in the next 2 columns because I don’t think Amanda is more powerful, and it only happened once.

If I thought that no one on the list hits, kicks, trips, shoves, or pushes others or takes others things by force, I would check this box down here.

Once you’ve identified one person who does this you can go to the next page and name someone else who hits kicks, shoves, or pushes others and takes others things by force. There are 9 pages, but only use as many as you need.

(Pause)

Turn to page 27. You are going to answer these questions the same way, but this time we are asking you to think of **people in your classroom who tease, insult or yell at other kids and call people mean names**. It can be the same kids you’ve already mentioned, or different kids. Name as many people in your class who you think do this. If there is no one who does these things, check the box at the bottom of the page. Go ahead.

(Pause)

Turn to page 36. Same thing again, one last time. This time we want to know **who gossips, says bad things about others behind their backs, or tries to get others to dislike someone else.** If there is no one who does these things, check the box at the bottom of the page. Go ahead.

(Pause)

Turn to page 45. The rest of the questions ask you about yourself. You are asked to circle different numbers to indicate your answers. For example in the first questionnaire you are asked to indicate how you have been feeling over the past few weeks. So if I had been feeling really interested over the past couple of weeks I would circle 5 to show that I had been really interested. There slightly different instructions for different parts of this last section, so make sure you read any instructions you see. Go ahead and work on your own for the rest of the survey. If you have any questions raise your hand.

Appendix C: Methods of Operationalizing Peer-Report Bullying Nominations

Given the peer-report methodology used in this research, some thought was given to how to best aggregate multiple ratings for a given individual to create categorical scores. In both our definitional and explicit behavioural bullying measures, peers are asked to nominate classmates who display a given behaviour (e.g., aggression, bullying). Given that there will inevitably be some disagreement among reporters about the behaviour of individuals, a decision rule must be adopted that will dictate how many nominations an individual must receive before they are considered to display that characteristic.

There are several methods of aggregating multiple peer reports into a single score (e.g., person A is a bully, yes or no). In Rodkin & Berger's (2008) original work with the *Who Bullies Whom?* measure, participants were classified as bullies or victims of bullying if they received at least two nominations from classmates. In that study, this method successfully differentiated among students who bully, students who are victimized by bullying, and students who do both ("bully/victims") on several social status and social-personality characteristics, and it identified bully/victim dyads that were characterised by power differentials. A two-nomination method of classification has also been supported by Bronson and Cornell (2009) on the basis that it captures involvement with bullying real enough to be known to the classroom. Advantages of using a two-nomination cut-off include the simplicity and face validity of the method and the ability to retain real differences between classrooms. It also captures all bullying involvement, rather than identifying only an extreme group, as is seen in other methods. Disadvantages of the two-nomination method include the fact that it doesn't compensate for differences in the size of the nominating pool between classrooms – two nominations from a nominating pool of 50 represents a lower bar than two nominations from a nominating pool of 20. The two-nomination method is also

relatively insensitive to the severity or unusualness of bullying involvement, which may be seen as a disadvantage in some contexts, but is seen as an advantage in the context of this study (see below).

Another method that could be used to quantify peer nominations of bullying involvement is standardizing scores. This method is similar to the method used in this study for quantifying popularity. The number of nominations that perpetrator/victim dyads receive would be standardized within classrooms and participants whose standardized scores are above a given cut-off would be categorized. Cut-offs used typically range from one-standard-deviation to half of a standard deviation above the mean (Hoff, Reese-Weber, Joel Schneider, & Stagg, 2009; Schwartz, 2000; Vaillancourt et al., 2003). This method is more commonly used than a two-nomination cut-off (Branson & Cornell, 2009). Advantages of this method are that it controls for differences between classrooms in the size of the nominating pool and it allows us to set a relatively strict classification to identify an extreme group. Disadvantages are that it may mask meaningful differences between classrooms in the number of students involved in bullying since only extreme groups of those involved in bullying are identified, so differences in the absolute level of bullying within classrooms will be masked. Another disadvantage is that cut-offs are based on statistical artefacts rather than meaningful differences (Olweus, 2013).

While neither the two-nomination nor the standardization method of operationalizing peer reports of bullying can be identified as the gold-standard method for operationalizing peer reports of bullying, it is believed that the two-nomination cut-off approach is more appropriate for this study. Since the study is concerned with the construct validity of Olweus' definition of bullying, measuring this definition with as much fidelity as possible is a study goal. Olwues' definition is strictly based on the absolute presence of given behavioural

Table C1

Number of Participants with Given Bullying Involvement Status with Different Methods for Determining Definitional Bullying Status

Method	# of individuals with specified bullying status			
	Bully	Victim	Bully/Victim	Uninvolved
2 Nomination	77 (14%)	162 (30%)	66 (12%)	230 (43%)
1 SD above	68 (13%)	73 (14%)	9 (2%)	385 (72%)

characteristics, rather than the relative unusualness of one's behaviour. The two-nomination cut-off captures all behaviours that seem to meet the criteria for bullying, whereas the one-standard-deviation cut-off only captures the bullying of the most involved students in a classroom. Since, however, the two-nomination cut-off is a relatively unusual way of categorizing peer nomination, we present data here on the number of participants identified as involved in bullying using both cut-off methods for the purpose of comparison.

Comparison of the two-nomination and one-standard-deviation cut-offs for creating definitional bullying status groups.

The numbers of participants nominated in each definitional bullying involvement category using both the two-nomination and one-standard-deviation cut-off can be seen in Table C1. A two-way contingency table analysis was conducted to evaluate whether the two-nomination method and the standard deviation method resulted in similar bullying status identifications for participants. As could be expected, since the two methods are based on the same data, they were found to be related, Pearson χ^2 (9, N = 535) = 446.86, p > .001, Cramér's V = .528. Table C2 shows the relationship between the two classification methods. Across the four bullying status groups, 65% of participants were classified the same way with both classification methods. A one-sample chi-square goodness of fit analysis was run to determine whether the proportion of participants identified in each group differed

Table C2

Correspondence between Methods of Calculating Definitional Bullying Status

		1 SD Cut-off			
		Uninvolved	Bully	Victim	Bully/Victim
2 Nomination Cut-off	Uninvolved	**230**	0	0	0
	Bully	27	**50**	0	0
	Victim	104	0	**58**	0
	Bully/Victim	24	18	15	**9**

Note: Numbers in bold represent agreement between the two methods.

significantly across methods. There was a significant difference, χ^2 (3, N = 535) = 533.10, p > .001. Specifically, while the proportion of students in the Bully group did not differ, χ^2 (1, N = 535) = 1.37, p = .243, the Victim, χ^2 (1, N = 535) = 125.65, p > .001 and Bully/Victim, χ^2 (1, N = 535) = 367.18, p > .001 groups were larger when the two-nomination cut-off was used, and the Uninvolved group was correspondingly smaller, χ^2 (1, N = 535) = 222.57, p > .001. The two-nomination method captures more victimization by bullying. As long as the victimization captured is believed to be meaningful, this is consistent with our reasoning for preferring the two-nomination method as we are interested in all bullying involvement, not just extreme examples. This suggests that the two-nomination method captures victimization by bullying that is real enough to be known to the classroom, but not so extreme as to be captured in measures of extreme groups. Therefore, the two-nomination method of identifying labelled bullying involvement seems more appropriate for our study, as it is more a measure of the absolute level of bullying activity and our study is more focused on the form than the severity of the behaviour.

The results also indicate that there is a large variation across classrooms in the amount of bullying behaviour reported. The number of bullies nominated per classroom ranges from two to 14. Using the two-nomination method of classification will allow us to

Table C3

Number of Participants with Given Bullying Involvement Status with Different Methods for Determining Explicit Behavioural Bullying Status

Method	# of individuals with specified behavioural bullying status			
	Bully	Victim	Bully/Victim	Uninvolved
2 Nomination	68 (13%)	92 (17%)	17 (3%)	358 (67%)
1 SD above	72 (14%)	63 (12%)	6 (1%)	394 (74%)

capture these real differences between classrooms that would become obscured if the one-standard-deviation cut-off were used.

Comparison of the two-nomination and one-standard deviation cut-offs for creating explicit behavioural bullying status groups.

The numbers of participants nominated in each behavioural bullying involvement category with each method can be seen in Table C3. It appears that classifications using the two different cut-offs are quite similar. A two-way contingency table analysis was conducted to evaluate whether the two-nomination method and the standard deviation method resulted in similar explicit behavioural bullying status identifications for participants. As could be expected, the two methods were found to be related, Pearson χ^2 (9, N = 535) = 587.22, p > .001, Cramér's V = .605. Approximately 82% of participants were classified in the same group using both methods. Table C4 shows the relationship between the two classification methods. A one-sample chi-square goodness of fit analysis showed that the proportion of

Table C4

Correspondence between Methods of Calculating Explicit Behavioural Bullying Status

		1 SD Cut-off			
		Uninvolved	Bully	Victim	Bully/ Victim
2 Nomination Cut-off	Uninvolved	**331**	15	11	1
	Bully	18	**50**	0	0
	Victim	41	0	**51**	0
	Bully/Victim	4	7	1	**5**

Note: Numbers in bold represent agreement between the two methods.

participants identified in each category using the two-nomination method differed significantly from the proportions found using the one-standard-deviation nomination method, χ^2 (3, N = 535) = 37.03, p > .001. Interestingly, the same pattern of difference was found as in the definitional bullying measure. The proportion of behaviourally definitional bullies did not differ, χ^2 (1, N = 535) = 0.26, p = .61, but more victims, χ^2 (1, N = 535) = 15.13, p > .001, and bully/victims, χ^2 (1, N = 535) = 20.40, p > .001, and correspondingly less uninvolved participants, χ^2 (1, N = 535) = 12.48, p > .001, were identified using the two-nomination method. As with the definitional bullying variables, for explicit behaviourally-based measures of bullying, the two-nomination appears to capture more victimization that exists in the classroom, although the magnitude of the difference between the two methods appears smaller in this case. Because we are interested in the absolute level of bullying behaviour, and to preserve classroom differences, we will use the two-nomination method for creating our categorical explicit behavioural bullying involvement variable.

Comparison of the two-nomination and one-standard-deviation cut-offs for creating aggression status groups.

The number of participants nominated in each aggression involvement category with both the two-nomination cut-off and the one-standard-deviation cut-off can be seen in Table C5. It appears that classifications using the two cut-offs are quite different. A two-way contingency table analysis was conducted to evaluate whether the two-nomination method

Table C5

A Comparison of Methods for Determining Aggression Status

Method	# of individuals with specified aggression status			
	Aggressor	Victim	Aggressor/Victim	Uninvolved
2 Nomination	42 (8%)	187 (35%)	209 (39%)	97 (18%)
1 SD above	64 (12%)	68 (13%)	17 (3%)	386 (72%)

Table C6

Correspondence between Methods of Calculating Aggression Status

		1 SD Cut-off			
		Uninvolved	Aggressor	Victim	Aggressor/ Victim
2	Uninvolved	**97**	0	0	0
Nomination	Aggressor	28	**14**	0	0
Cut-off	Victim	153	0	**34**	0
	Aggressor/Victim	108	50	34	**17**

Note: Numbers in bold represent agreement between the two methods.

and the standard deviation method produce similar results. As could be expected (since they are based on the same data), the two methods of classification were found to be related, Pearson χ^2 (9, N = 535) = 150.84, p > .001, Cramér's V = .31. However, only 30% of participants were classified in the same group using both methods. Table C6 shows the relationship between the two classification methods. A one-sample chi-square goodness of fit analysis showed that the proportion of participants identified in each category using the two-nomination method differed significantly from the proportions found using the one SD nomination method, χ^2 (3, N = 535) = 2600.70, p < .001. The proportions differed for each group (ps between <.001 and .003). The largest difference occurred in the proportion of participants who were found to be uninvolved in aggression. Many more participants were identified as involved in aggression using the two-nomination cut-off than the standard deviation method. This is consistent with the pattern of results seen in our measure of definitional bullying behaviour. The most parsimonious explanation for the differences in findings across methods seen here is that the two-nomination method captures all aggression within a classroom, whereas the one-standard deviation cut-off captures only individuals who display an unusually high level of aggression or victimization compared to their peers. Our results suggest that most individuals display some aggression and/or experience some victimization by aggression. For our study, since we are interested in capturing different

outcomes associated with different types of aggressive behaviour, the two-nomination cut-off is deemed the more appropriate measure, since it allows us to capture the full range of aggressive involvement.

Validity check for two-nomination method.

The two-nomination method is preferred for this study on conceptual grounds; however, concerns may remain about whether bullying and aggressive involvement captured this way are meaningful. To determine whether bullying status groups created using a two-nomination cut-off predicted differences in student outcomes that we would expect to be associated with bullying involvement, I completed a one-way MANOVA analysis. I chose to use a MANOVA analysis despite the fact that the dependent variables were not highly correlated (absolute values of *r* ranged from .65 to .003), as a way of controlling for Type 1 error. The independent variable for this analysis is definitional bullying status created using a two-nomination cut-off. There were four outcome groups: the Bully group, the Victim group, the Bully/Victim group, and the Uninvolved group. The dependent variables (DV) were self-perceived social rank, social acceptance and submissive behaviour, connection to the school, positive and negative affect, beliefs about aggression, and teacher-rated academic performance and behaviour problems. All of the dependent variables are continuous.

Before running the analysis, an evaluation of MANOVA assumptions was completed. Sample sizes for the four bullying status groups differ. There are 77 participants in the Bully group, 162 participants in the Victim group, 66 participants in the Bully/Victim group, and 230 participants in the uninvolved group. Despite this size difference, there are well over the minimum of 20 participants per cell, required to ensure robustness to violation of multivariate normality. Five multivariate outliers were found and removed from the remainder of the analyses. Preliminary assessments of the homogeneity of Variance-

Covariance Matrices assumption are favourable, as the ratio of largest to smallest group variance for each DV were within an acceptable range (Tabachnick & Fidell, 2007, p. 86). Box's test of Equality of Covariance Matrices was significant, however, Box' $M = 317.75$, $F(135, 96577) = 2.23$, $p < .001$, indicating a possible violation of the homogeneity of Variance-Covariance Matrices assumption. This test is known to be "too strict" (IBID), but to be conservative, Pillai's criterion will be used to evaluate multivariate significance instead of Wild's Lambda.

Significant differences were found between the four bullying involvement groups on the dependent measures, Pillai's Trace = .30, $F(27, 1191) = 4.89$, $p < .001$. The effect size was medium, partial $\eta^2 = .10$. Given this overall significant finding showing group differences with a bullying status variable calculated using a two-nomination cut-off, a series of post-hoc tests (one-way AVOVAs) were performed to understand the source of this difference. No correction was made to the alpha level of these tests given the overall significant result. As displayed in Table C7, there were significant group differences for all dependent variables except Negative Affect. Follow-up tests were conducted to evaluate

Table C7

Follow-up One-Way ANOVAs Examining Definitional Bullying Status Group Differences

Variable	*F*	df	*P*	Partial Eta Squared
Social Rank (ASCS)*	16.25	(3, 469)	>.001	.09
Submissive Behaviour (ASBS)	4.63	(3, 467)	.003	.03
Social Acceptance (SAIS)	6.54	(3, 455)	.03	.02
Negative Affect (PANAS – NA)	1.57	(3, 435)	.20	.01
Positive Affect (PANAS – PA)	5.49	(3, 443)	.03	.02
School Membership (PSSM)*	9.13	(3, 470)	.002	.03
Beliefs about Aggression (NBAAS)*	6.58	(3, 462)	>.001	.04
Behaviour Problems*	38.139	(3, 516)	>.001	.18
Academic Performance	4.10	(3, 519)	.007	.02

Note: * indicates that Levene's test of Equality of Error Variance was significant, indicating significant variation in variance scores.

Table C8

Means and Standard Deviations of Participant Functioning by Definitional Bullying Status Group

	Uninvolved	Bully	Victim	Bully/Victim
Social Rank (ASCS)	6.65 (1.35)	7.33 (1.14)	5.94 (1.53)	6.45 (1.64)
Submissive Behaviour (ASBS)	2.55 (0.66)	2.33 (0.61)	2.68 (0.66)	2.67 (0.75)
Social Acceptance (SAIS)	3.89 (0.82)	3.82 (0.79)	3.67 (0.91)	3.58 (0.88)
Positive Affect (PANAS – PA)	3.58 (0.75)	3.54 (0.79)	3.35 (0.75)	3.34 (0.85)
School Membership (PSSM)	3.84 (0.75)	3.68 (0.68)	3.64 (0.84)	3.41 (0.82)
Beliefs about Aggression (NBAAS)	1.37 (0.40)	1.58 (0.45)	1.34 (0.37)	1.54 (0.66)
Behaviour Problems	1.10 (0.19)	1.51 (0.49)	1.18 (0.34)	1.43 (.455)
Academic Performance	4.80 (1.61)	4.13 (1.73)	4.55 (1.66)	4.24 (1.62)

Note: Standard Deviations are shown in brackets.

pairwise differences among the means for the eight variables for which significant differences were indicated. These tests were done using the Tukey test. Group means and standard deviations (in brackets) are presented in the Table C8.

The results of the post-hoc testing are as follows:

- Bullies had significantly higher self-perceived social rank (ASCS scores) than all other groups (*p*s between .003 and <.001). The victim group was also significantly lower than the uninvolved group on self-perceived social rank ($p < .001$).
- Students in the bully group reported less submissive behaviour than students in the Victim ($p = .003$) or Bully/Victim group ($p = .03$).
- No group differences were significant for social acceptance.
- Students in the victim group had lower levels of positive affect than students in the uninvolved group ($p = .04$).
- Students in the Bully/Victim group had less of a psychological sense of school membership (PSSM) than students in the uninvolved group ($p = .002$).
- Students in the Bully and Bully/Victim groups had higher beliefs in the acceptability of aggression than students in the victim group ($p = .001$ and $p = .02$ respectively).

Students in the bully group also had higher belief in the acceptability of aggression than the uninvolved group (p = .007).

- Students in the Bully and Bully/Victim group both had more teacher-reported behaviour problems than students in the Victim and Uninvolved groups (ps < .001).
- Students in the Bully group had lower academic performance than students in the Uninvolved group (p = .013).

Summary and conclusions.

Two methods of aggregating peer-report data into a categorical aggression and bullying statuses are presented. In all cases, the two methods were significantly related. For categorical rating of aggressive involvement, only 30% of participants are classified the same way with both methods. Agreement was much higher for ratings of bullying. For the definitional bullying status method, 65% of participants were categorized the same way using each method, and for the explicit behavioural method, 82% of participants were categorized the same way using each method. There were some differences, however. More aggressive behaviour and bullying involvement, particularly more victimization, was identified using the two-nomination method. This could be expected, as the two-nomination method measures the absolute level of behaviours in the classroom, whereas the standard deviation method identifies only extreme groups. Therefore, since our study is focused on the presence or absence of certain behaviours and characteristics, rather than the severity of these behaviours and characteristics, the two-nomination method is deemed to be the more appropriate method for aggregating peer report data to use going forward. It also allows us to maintain apparently real differences between classrooms in the level of aggression reported. The validity of the two-nomination method was demonstrated in that bullying status groups

formed using that method were shown to be related to our individual outcomes in an expected manner.

Appendix D: Methods of Operationalizing Characteristics of Aggressive Behaviour

In this study, three characteristics of aggressive behaviour are measured. These are the characteristics that define aggression as bullying: repetition, intention to harm, and whether the aggression took place in the context of a power imbalance. Participants' nominations on these variables were used to calculate explicit behavioural bullying scores (i.e., if all three characteristics were present, aggression or victimization met the criteria to be considered bullying involvement) and are also of interest as potential predictors of participant outcomes. The characteristics of aggression were measured using this study's newly developed extension of Rodkin and Berger's (2008) paired nomination procedure which uses items from Bjorkqvist & Osterman's (1998) *Peer Estimated Conflict Behaviour* (PECOBE) Scale to measure aggressive behaviour and victimization and original items to measure whether the aggression is repeated, intentionally harmful, and in the context of a power imbalance.

As with converting peer bullying or aggression nominations to a single bullying involvement classification, a decision rule will have to be adopted about how to convert peer nomination data to individual scores for characteristics of aggressive behaviour. Reporting participants may not, of course, agree 100% about, for example, whether a participant's victimization is repeated. Peer nomination data can be converted to dichotomous, count, or proportional scores. Descriptive statistics for each method will be presented below before they are compared and conclusions will be drawn about the best method to use for the rest of the study.

Dichotomous Scores – The "Once is Enough" Method

To convert peer-nomination data about the characteristics of aggression into dichotomous scores, a method very similar to the two-nomination method used to classify

bullying and aggression will be used. In this case, if two nominators say that the aggression or victimization of an individual has a certain characteristic, it will be considered to have it, even if other reporters say that the characteristic is not present. This is based on the previously discussed reasoning that if two people in a classroom identify a behaviour then it is real within the social world of the classroom. For ease of reference, this method will be labelled the "Once is Enough" method (even though two nominations are actually required) because it reflects the idea that the presence of repeated aggression, rather than the frequency with which the occurrence is reported or the proportion of relationships characterised by repeated aggression, is likely predictive of participant outcomes. The number of individuals nominated to have each characteristic using the "Once is Enough" method can be seen in Table D1.

Count Scores – The "More is Worse" Method

The second strategy used to create a single score based on the information from multiple reporters is the "More Is Worse" method. In this method, the individual's score will be a count of how many times the individual's behaviour is nominated as having a certain

Table D1

Frequencies of Nominations for Given Characteristics of Aggressive Behaviour Using the "Once is Enough" Method

	# of individuals nominated to have a given characteristic and percent of total sample	
	Aggression	Victimization
Any of the Behaviour Present	251 (47%)	396 (74%)
Behaviour is Repeated	149 (28%)	255 (48%)
Behaviour is Intentional	155 (29%)	258 (48%)
Behaviour is in the Context of a Power Imbalance	130 (24%)	208 (39%)
Behaviour Has All Three Characteristics	85 (16%)	109 (20%)

Note: "Any of the Behaviour Present" represents those who have been identified as displaying some aggressive behaviour or victimization by aggression using a two-nomination cut-off as seen in our measurement of general aggressive behaviour.

characteristic. This results in a continuous variable and is very similar to the way that continuous methods of bullying involvement are calculated. For example, if three classmates indicate that an individual is victimized repeatedly, their score on the repetition variable will be three. Descriptive statistics for each characteristic variable calculated using the "More is Worse" method can be seen in Table D2 below.

Proportion Scores – The "Proportion is What Counts" Method

The final strategy we will use to create a single score for aggression characteristics based on the information from multiple reporters is the "Proportion is What Counts" model. We will use the proportion of people who nominated the person as being involved in aggression and who also said that that aggression had a given characteristic as the final characteristic score. For example, if someone is nominated as the victim of aggression by three people, but only one of those nominators says the victimization is characterised by a power imbalance, the individual's score on the power imbalance variable would be 1/3 or 0.33. This method will allow us to determine whether the agreement among nominators

Table D2

Characteristics of Aggressive Behaviour Using the "More is Worse" Method

Aggression						
	Total Sample			Aggressive Participants		
	Mean	*SD*	Range	Mean	*SD*	Range
Repeated	1.25	1.97	0-11	2.49	2.30	0-11
Intentional	1.3	2	0-11	2.57	2.31	0-11
Power Imbalance	1.12	1.86	0-10	2.25	2.20	0-10
Victimization						
	Total Sample			Victimized Participants		
	Mean	*SD*	Range	Mean	*SD*	Range
Repeated	1.92	2.04	0-13	2.5	2.07	0-13
Intentional	1.96	2.16	0-14	2.56	2.21	0-14
Power Imbalance	1.59	1.99	0-11	2.07	2.10	0-11

Note: Aggressive participants are those who have been identified as displaying aggressive behaviour on the PECOBE using a two-nomination cut-off (n = 251) and Victimized Participants are those who have been identified as having been victimized by aggressive behaviour (n = 396).

Table D3

Characteristics of Aggressive Behaviour Using the "Proportion is What Counts" Method

	Aggressive Participants		Victimized Participants	
	Mean	*SD*	Mean	*SD*
Repeated	.58	.60	.60	.34
Intentional	.60	.39	.57	.38
Power Imbalance	.51	.40	.47	.37

Note: Aggressive participants are those who have been identified as displaying some aggressive behaviour using a two-nomination cut-off (n = 251) and Victimized Participants are those who have been identified as having been victimized by aggressive behaviour (n = 396).

(which would be thought to be associated with higher salience and consistency of behaviour) is predictive. Descriptive statistics for each characteristic variable calculated using the "More is Worse" method can be seen in Table D3.

Comparing Methods of Operationalizing Peer-Ratings of Aggression Characteristics

We have three possible methods for operationalizing behavioural ratings: "Once is Enough," "More is Worse," and "Proportion is What Counts." To help us determine whether one might be better than others, we first examine the extent to which these methods are correlated. Results are presented in the two tables below. Table D4 contains correlations on ratings of aggression, while Table D5 contains correlations on rating of victimization.

As we can see, all correlations are significant. We would expect this, given that all three method scores are created from the same data. What is potentially interesting is that the Once is Enough and the More is Morse method scores are strongly correlated (especially considering that the Once is Enough variable is dichotomous), but the correlation between these two methods and the Proportion is What Counts method appear to be somewhat weaker. This means that having at least two people say that you display one of these characteristics is highly correlated with the overall number of people who will report that you have this characteristic, but less strongly correlated with the proportion of people who say you have that characteristic.

Table D4

Correlations Among Methods of Operationalizing Behavioural Ratings of Aggression

Behaviour Characteristic		Once is Enough	More is Worse	Proportion is What Counts
Intention to Harm	Once is Enough	1	.79*	.41*
	More is Worse		1	.45*
	Proportion is What Counts			1
Power Imbalance	Once is Enough	1	.83*	.53*
	More is Worse		1	.56*
	Proportion is What Counts			1
Repetition	Once is Enough	1	.80*	.44*
	More is Worse		1	.47*
	Proportion is what Counts			1
All Three (Meets Criteria for Bullying)	Once is Enough	1	.84*	.50*
	More is Worse		1	.59*
	Proportion is What Counts			1

Note: * Indicates that the correlation is significant at the .01 level (2-tailed)

An examination of the distribution of proportion scores indicated something interesting. Most of these distributions were tri-modal, with histogram peaks at 0 and 1 as well as the centre of the distribution (.5). This phenomenon is illustrated in Table D6. You can see that for all variables, a substantial proportion of individuals have a Proportion is

Table D5

Correlations Among Methods of Operationalizing Behavioural Ratings of Victimization

Behaviour Characteristic		Once is Enough	More is Worse	Proportion is What Counts
Intention to Harm	Once is Enough	1	.74*	.59*
	More is Worse		1	.56*
	Proportion is What Counts			1
Power Imbalance	Once is Enough	1	.76*	.55*
	More is Worse		1	.59*
	Proportion is What Counts			1
Repetition	Once is Enough	1	.74*	.52*
	More is Worse		1	.50*
	Proportion is What Counts			1
All Three (Meets Criteria for Bullying)	Once is Enough	1	.79*	.57*
	More is Worse		1	.67*
	Proportion is What Counts			1

* Indicates that the correlation is significant at the .01 level (2-tailed).

Table D6

Percentage of Participants with "Proportion is what Counts" Scores of 0, 1, or Chance Levels.

	Proportion score of 0	Proportion score of 1	Proportion score between .4 and .6
Aggression			
Intention	20.8%	36.6%	15.2%
Power Imbalance	30.1%	28.3%	15.0%
Repetition	23.3%	32.7%	17.9%
All Three	48.0%	11.6%	15.6%
Victimization			
Intention	20.9%	29.2%	13.8%
Power Imbalance	28.0%	18.6%	20.1%
Repetition	15.5%	27.0%	19.2%
All Three	49.2%	5.7%	14.9%

What Counts score of either 1 or zero, indicating perfect agreement among reporters. Scores in the range of chance (.4-.6) have also been separated out in the table below to illustrate the proportion of individuals for whom there is little agreement.

This pattern of findings could be interpreted to mean that reporters showed a lot of agreement about some people in their class, but that there were other classmates who were more difficult to identify. However, another possibility is that prefect agreement in proportion scores is related to having fewer nominators. If only two people have identified someone as being aggressive or a victim of aggression, then simply by chance the proportion score would either be one or zero 50% of the time. To test whether the overall number of nominators heavily influences the proportion scores, two independent sample *t*-test were conducted. The first test concerned proportion scores for the perpetration of aggression. The groups were created based on proportion scores for aggression with all three characteristics of bullying. One group was composed of those whose proportion scores showed perfect agreement (i.e., 0 or 1) and the other group included all other participants who had been nominated by at least two people as being aggressive. The dependent variable was the

number of participants who nominated the target participant as being aggressive. Participants with perfect proportion scores had an average of 2.32 nominators, with a standard deviation of 1.64. The average number of nominators for those who did not have perfect proportion scores was 5.98, with a standard deviation of 3.15. The *t*-test results (equality of variances not assumed) indicate that this is a significant difference between groups, $t(179.41) = 12.26$, $p < .001$. The effect size of this difference was large, $d = 1.38$. This indicates that the "Proportion is What Counts" score is highly influenced by the number of reporters, meaning that its utility as a measurement of agreement among reporters is likely compromised.

The independent *t*-test was repeated, this time focusing on proportion scores for those nominated to be victims of aggression. The groups were created based on proportion scores for victimization by aggression that had all three characteristics of bullying. One group was composed of those whose proportion scores showed perfect agreement (i.e., 0 or 1) and the other group included all other participants who had been nominated by at least two people as being aggressive. The dependent variable was the number of participants who nominated the target participant as being the victim of aggression. Participants with perfect proportion scores had an average of 2.89 nominators, with a standard deviation of 1.82. The average number of nominators for those who did not have perfect proportion scores was 5.73, with a standard deviation of 2.77. The *t*-test results (equality of variances not assumed) indicate that this is a significant difference between groups, $t(341.06) = 12.67$, $p < .001$. The effect size of this difference was large, $d = 1.19$. Again, this indicates that the "Proportion is What Counts" score is highly influenced by the number of reporters, meaning that its utility as a measurement of agreement among reporters is likely compromised.

Summary and Conclusions

In this study, we measured three characteristics of aggressive behaviour that define aggression as bullying: repetition, intention to harm, and whether the aggression took place in the context of a power imbalance. As with converting peer bullying nominations to a single bullying involvement classification, a decision rule had to be adopted about how to convert peer-nomination data to individual scores. Peer nomination data could be converted to dichotomous, count, or proportional scores. We have termed these the "Once is Enough" method, the "More is Worse" method, and the "Proportion is What Counts" methods, respectively. Descriptive data for each of the three characteristics of bullying calculated using each of these three methods has been presented as well as information about the relationship among the three methods. We found that all methods indicated that there were some people who displayed each of the characteristics measured. The "Once is Enough" and "More is Worse" methods were highly correlated (*r*s for individual characteristics ranged from .79 to .83). This was expected, given that they are based on the same data. Difference between the two types of measurement were likely a result of the restriction of range in the dichotomous "Once is Enough" variable. The "More is Worse" method of operationalizing characteristics retained more information from the original data. Nevertheless, both the "Once is Enough" and the "More is Worse" methods appear to be valid and will be used depending on the type of variable called for in subsequent analyses (e.g., creating explicit behavioural bullying classifications will require a categorical variable, so the "Once is Enough" method will be used in that case; however, the "More is Worse" method will be used when characteristics are used as predictors in regression analyses in order to preserve the maximum amount of data from the original data). The "Proportion is What Counts" method of operationalizing peer reports will not be used in further analyses, as proportion

scores were shown to be strongly related to the number of nominators, limiting their utility as measures of agreement.

Appendix E: Indicators of Variable Importance in Multiple Regression

Table E1

Purpose of Statistical Measures Used to Investigate the Importance of Individual Predictors in Multiple Regressions with Multicollinearity

Statistics	Purpose of Measure
Beta weight	The independent variables' contribution to prediction within a linear regression equation while holding all other independent variables constant.
Structure coefficient	The correlation between the independent variable and the proportion of variance in the effect (or of the expected values of the dependent variable, $\hat{Y}$).
Squared structure coefficient	The proportion of variance in the effect (or $\hat{Y}$) that can be accounted for by the predictor alone, irrespective of collinearity with other predictors.
Correlation coefficient	The magnitude and direction of the bivariate linear relationship between each independent variable and the dependant variable.
Squared correlation coefficient	The proportion of variance in the dependant variable that can be accounted for by the predictor alone, irrespective of collinearity with other predictors.
Commonality Coefficients:	
Unique effects	The amount of variance each independent variable contributes to a regression equation that is not shared with other independent variables (squared semi-partial correlation).
Common effects	A quantification of how much variance is shared between independent variables.
Dominance weights	A quantification of whether one independent variable contributes more variance than other independent variables calculated by comparing the relative contribution of predictors across pairwise variable sets.
Relative importance weights	The variable importance based on a method that addresses multicollinearity by creating variables' uncorrelated "counterparts." Does not aid in investigating suppression effects.

Appendix F: Commonality Analysis and All Possible Subsets Regressions for the Prediction of Participant Popularity with Behaviourally Measured Characteristics of Aggression.

Table F1

Commonality Analysis and All Possible Subsets Regressions for the Prediction of Participant Popularity with Behaviourally Measured Characteristics of Aggression

Predictor Combination	Commonality Analysis		All Possible Subsets Regression R^2
	Commonality Coefficient	% of Variance Explained	
X1 (# of Nominators)	0.000	0.024	0.000
X2 (Aggression)	0.010	6.745	0.023
X3 (Intention)	0.017	11.470	0.006
X4 (Power Imbalance)	0.063	41.971	0.060
X5 (Repetition)	0.021	14.081	0.005
X6 (All Three Characteristics)	0.002	1.201	0.025
X1 X2	0.000	-0.004	0.023
X1 X3	0.000	0.138	0.006
X2 X3	-0.007	-4.565	0.036
X1 X4	0.000	0.208	0.060
X2 X4	0.039	26.188	0.069
X3 X4	0.000	-0.297	0.118
X1 X5	0.000	0.028	0.005
X2 X5	-0.009	-6.143	0.045
X3 X5	0.015	9.861	0.006
X4 X5	0.005	3.357	0.118
X1 X6	0.000	0.000	0.025
X2 X6	0.005	3.337	0.028
X3 X6	0.005	3.606	0.035
X4 X6	0.036	24.421	0.086
X5 X6	0.005	3.218	0.035
X1 X2 X3	0.000	-0.125	0.036
X1 X2 X4	0.000	-0.218	0.069
X1 X3 X4	0.000	0.159	0.119
X2 X3 X4	-0.008	-5.270	0.121
X1 X2 X5	0.000	0.032	0.046
X1 X3 X5	0.000	-0.190	0.006
X2 X3 X5	0.016	10.697	0.047
X1 X4 X5	0.000	-0.255	0.118
Predictor Combination	Commonality Coefficient	% of Variance Explained	R^2

Table F1 continued.

Predictor Combination	Commonality Analysis		All Possible Subsets Regression
	Commonality Coefficient	% of Variance Explained	R^2
X2 X4 X5	-0.014	-9.603	0.124
X3 X4 X5	0.002	1.024	0.132
X1 X2 X6	0.000	0.034	0.028
X1 X3 X6	0.000	0.042	0.035
X2 X3 X6	-0.001	-0.898	0.060
X1 X4 X6	0.001	0.328	0.086
X2 X4 X6	-0.012	-7.987	0.096
X3 X4 X6	-0.020	-13.100	0.127
X1 X5 X6	0.000	0.029	0.035
X2 X5 X6	-0.003	-2.042	0.069
X3 X5 X6	0.015	10.095	0.037
X4 X5 X6	-0.019	-12.730	0.129
X1 X2 X3 X4	0.000	-0.180	0.121
X1 X2 X3 X5	0.000	0.199	0.048
X1 X2 X4 X5	0.000	0.213	0.125
X1 X3 X4 X5	0.000	-0.025	0.132
X2 X3 X4 X5	-0.024	-16.307	0.147
X1 X2 X3 X6	0.000	-0.092	0.060
X1 X2 X4 X6	-0.001	-0.338	0.096
X1 X3 X4 X6	0.000	-0.223	0.127
X2 X3 X4 X6	0.015	9.878	0.128
X1 X2 X5 X6	0.000	0.014	0.070
X1 X3 X5 X6	0.000	-0.068	0.037
X2 X3 X5 X6	-0.002	-1.067	0.086
X1 X4 X5 X6	0.000	-0.189	0.129
X2 X4 X5 X6	0.015	9.865	0.132
X3 X4 X5 X6	-0.021	-13.875	0.139
X1 X2 X3 X4 X5	0.000	0.013	0.147
X1 X2 X3 X4 X6	0.000	0.255	0.128
X1 X2 X3 X5 X6	0.000	0.009	0.086
X1 X2 X4 X5 X6	0.000	0.128	0.132
X1 X3 X4 X5 X6	0.000	0.167	0.139
X2 X3 X4 X5 X6	0.004	2.869	0.149
X1 X2 X3 X4 X5 X6	0.000	-0.115	0.149
Predictor Combination	Commonality Coefficient	% of Variance Explained	R^2