Effects of a standardized school-based behavior management program on externalizing behavior, self-esteem and peer relations in 7-9 year old children

Elisabeth van de Leur
Hedda Nyman
Skolbaserade interventioner kan vara effektiva för att främja välmående och förhindra beteendeproblem hos barn. Denna studie undersöker effekter av en skolbaserad beteendeintervention kallad Komet, avseende utagerande beteende, självvärdering och kamratrelationer i en grupp bestående av 2834 lågstadiebarn. Lärarskattning av utagerande beteende, självskattning av självvärdering, samt positiva och negativa kamratnomineringar användes i en klusteranalys som genererade fyra profiler. Profilerna döptes till ”utagerande-avvisade”, ”populära”, ”osäkra” och ”lågmälda-lugna”. Effekter av Komet på de fyra profilerna undersöktes vid tre mättillfällen (före, efter och uppföljning) med MANOVA. Resultaten indikerar att Komet bidrar till att höja självvärdering hos ”osäkra” barn. En trend i resultaten, dock icke-signifikant, antyder att Komet kan förebygga utagerande beteende hos ”utagerande-avvisade” barn samt bidra till positiv utveckling för ”lågmälda-lugna” barn. De identifierade profilerna stödjer till viss del tidigare forskning om kamratstatus.
# INTRODUCTION
Externalizing behavior ........................................................................... 2
Self-esteem .......................................................................................... 3
Peer relations ....................................................................................... 4
Relationships between externalizing behavior, self-esteem and peer-relations .... 7
School-based interventions ....................................................................... 8
Communication Method (Comet) .............................................................. 9
Current study ........................................................................................ 11
Research questions ............................................................................... 12
METHOD ............................................................................................... 12
Participants .......................................................................................... 12
Procedure ............................................................................................. 13
Comet ..................................................................................................... 13
The comparison group ........................................................................... 14
Measures ............................................................................................... 14
  Brief Teacher Rating of Externalizing behavior (BREB) ......................... 14
  I Think I Am (ITIA) ........................................................................... 15
  Social Preference and Social Impact. .................................................... 16
Data analysis .......................................................................................... 16
Cluster analysis ..................................................................................... 16
Mixed ANOVA for repeated measures. .................................................... 17
Attrition analysis .................................................................................. 18
Ethical considerations ........................................................................... 19
RESULTS ............................................................................................... 19
Profiles .................................................................................................. 19
Effects across profiles and assessment points ........................................... 20
  Effects of time and intervention in Profile 1 (“Externalizing-Rejected”) .... 24
  Effects of time and intervention in Profile 2 (“Popular”) ....................... 25
  Effects of time and intervention in Profile 3 (“Insecure”) ...................... 26
  Effects of time and intervention in Profile 4 (“Quiet-Cool”) ................. 27
DISCUSSION .......................................................................................... 28
Profiles .................................................................................................. 28
Effects of Comet and time across profiles ............................................... 31
Limitations of the study ....................................................................... 33
Conclusions .......................................................................................... 38
Recommendations for future research ................................................... 38
REFERENCES ....................................................................................... 39
EFFECTS OF A STANDARDIZED SCHOOL-BASED BEHAVIOR MANAGEMENT PROGRAM ON EXTERNALIZING BEHAVIOR, SELF-ESTEEM AND PEER RELATIONS IN 7–9 YEAR OLD CHILDREN

Elisabeth van de Leur and Hedda Nyman

School-based interventions can be effective in promoting well-being and preventing maladjustment in children. This study investigates the effects of an indicated school-based behavior management program called Comet with regard to externalizing behavior, self-esteem and peer relations in a sample of 2,834 primary school children. Teacher ratings of externalizing behavior, self-reports of self-esteem, and positive and negative peer nominations were used in a cluster analysis, which categorized children into four profiles: “Externalizing-Rejected”, “Popular”, “Insecure”, and “Quiet-Cool”. MANOVA was performed to investigate the effects of Comet across profiles and time: pre; post and follow-up. Results indicate that Comet effectively increases self-esteem in “Insecure” children. Although non-significant, a trend in results suggest that Comet may effectively prevent externalizing behavior in “Externalizing-Rejected” children, and may foster positive development of children with the “Quiet-Cool” profile. The profiles identified in the cluster analysis confirm, to some extent, previous research regarding peer status groups.

School is the place where many children spend most of their time on a day-to-day basis. Apart from being an institution for education and socialization, schools offer a unique forum for reaching large numbers of children and can be used to implement interventions aimed at improving children’s well-being. Schools could also be potential avenues to counterbalance outside factors that may put children at risk for maladjustment in terms such as socioeconomic status, parental ability, or children’s intellectual and emotional capacity.

During recent years, a wide range of school-based interventions have been developed. Considering the amount of available interventions, there is a call for evaluating their efficacy, ensuring interventions fulfill their purpose. Evaluation also supports the maxim of using limited resources in the best way possible (Forster, 2010). One way of achieving intervention efficacy on a larger scale is to implement indicated intervention programs for targeted students at classroom level. This is opposed to, for example, pulling targeted children out of the classroom for intervention delivery. One of the many advantages of this inclusive approach to intervention is that the implementation is potentially useful for their peers as well as the targeted children. With this format intervention programs benefit as many children as possible.

1 We would like to express our gratitude to Lotta Reuterskiöld and Martin Forster for their invaluable help in mentoring this master thesis.
One target area of recently developed school-based interventions is reduction of externalizing behavior (Solomon, Klein, Hintze, Cressey, & Peller, 2012). Researchers have long been investigating the linkages between externalizing behavior, self-esteem and peer relations to find potential causalities and correlates to developmental trajectories during childhood and adolescence (e.g., Chen, 2012; Coie, Dodge, & Coppotelli, 1982; Donnellan, Trzesniewski, Robins, Moffitt & Caspi, 2005; Mathias, Beibl, & Dilalla, 2011; Newcomb, Bukowski, & Pattee, 1993). Available knowledge points toward several associations and interactions between these constructs. Therefore, investigating these associations further by an intervention targeting externalizing behavior is highly motivated. The next section provides descriptions of externalizing behavior, self-esteem and peer relations according to definitions in the literature, followed by a review of associations between the constructs.

Externalizing behavior
The term externalizing behavior is commonly used as an umbrella term to describe several behavioral problem areas, such as aggression, delinquency and antisocial behaviors (Donnellan et al., 2005; Mathias, Beibl, & Dilalla, 2011). Other authors have broadened the term to include conduct problems, hyperactivity, lack of self-control and lack of empathy (Ybrandt & Armelius, 2010). Yet, others have chosen a more blunt definition of the term to simply describe externalizing behavior as ways of “dumping” ones problems onto others (Baumeister, Campbell, Krueger, & Vohs, 2003). Karlberg (2011) speaks of externalizing behavior in terms of behaviors that breach the norms of a given context. Given this definition of externalizing behavior, defining which behaviors are to be considered externalizing is a question of culture. In the case of school children, this would mean that behavior is evaluated on the basis of norms, rules and expectations within the school environment. If children are expected to sit still quietly in their chairs while teachers lecture, walking around the classroom and speaking without permission, in some contexts, could be considered “acting out”. However, in the context of a pre-school or a primary school classroom with less focus on structured teaching, this same type of behavior might not be seen as a problem. Standardization or classification of what is to be considered externalizing behavior is therefore difficult using Karlberg’s definition.

Externalizing behavior can also be viewed from a perspective of gender-related norms. In a recent meta-analysis, Chaplin and Aldao (2013) showed that boys were more prone to engage in externalizing behavior than girls. Regardless of gender, however, all children are expected to engage in externalizing behavior from time to time (Karlberg, 2011). Defying norms and testing limits is a natural part of childhood. Reasonably, interventions to reduce externalizing behavior should only be recommended when a child engages in such behavior repeatedly and in a manner that hinders the child’s learning activities and social interaction at school. The detection of problematic incidences of externalizing behavior within the context of structural demands and individual achievement at school (as well as in other contexts) gradually increases each year children attend school (ibid.). Externalizing behavior is a defined part of neurodevelopmental and psychiatric disorders such as Attention Deficit Hyperactivity Disorder (ADHD), Conduct Disorder (CD) and Opposition Defiant Disorder (ODD) (Diagnostic and Statistical Manual of Mental Disorders, DSM-IV; APA, 2002).
However, externalizing behavior does not automatically imply that a child has a specific diagnosis.

According to behavioristic learning theory, externalizing behavior is considered the product of interaction between a child and its context (Karlberg, 2011). In other words, problematic behavior is not expressed within a vacuum. Efforts to reduce externalizing behavior in children should therefore take into consideration that contextual contingencies are both potential sources as well as solutions to problematic behavior, rather than attributing all responsibility to the child in question – or to a diagnosis, for that matter.

Self-esteem
In a large review of research on associations between self-esteem and different aspects of performance, Baumeister, Campbell, Krueger and Vohs (2003) define self-esteem as “the value people place on themselves” and as “the evaluative component of self-knowledge” (ibid., p. 2). They go on to describe self-esteem as a perception rather than a reality-based construct, emphasizing that both high and low self-esteem can reflect evaluations of varied accuracy. Accordingly, a person with extremely high self-esteem could just as well be truly successful and competent, as narcissistic. Likewise, low self-esteem can be based on correct information of inadequacy and failure, or represent negatively biased perceptions of one’s qualities and characteristics (ibid.). The ability to evaluate oneself requires the cognitive ability to reflect over one’s general self-worth; to make comparisons of these reflections within a social context, and verbal skills to put these reflections into words (Mathias, Beibl, & Dilalla, 2011). Development of these abilities usually commences as children enter primary school. In pre-school, many children display inflated self-esteem and a general absence of negative self-evaluations. This is considered as manifestations of undeveloped self-evaluation skills. The self-esteem of pre-school children also tends to be rather unstable. After sufficient cognitive abilities for self-evaluation have been fully developed, however, self-esteem is a relatively stable trait (ibid.).

Self-esteem poses somewhat of a methodological problem when investigated by researchers. Because self-esteem is, by definition, a subjective psychological construct, the only established way of measuring it is by self-report (Baumeister, Campbell, Krueger, & Vohs, 2003). Greenwald and Farnham (2000) pose that since people are able to process social information about others both in a controlled, declarative mode, as well as and in an automatic, unreflected mode, there is reason to believe that socially evaluative information about oneself, i.e., self-esteem, exists in an explicit as well as implicit form. Since most studies regarding self-esteem measure the construct by self-reports, a possible implicit self-esteem is not taken into account.

Another problem with the self-report procedure for measuring self-esteem is that it complicates distinction between what researchers believe to be different types of self-esteem. Schneider and Turkat (1975), for example, argue for a division between genuine high and defensive high self-esteem. High ratings on self-esteem scales can, per se, be an attempt to gain approval and therefore be considered a defensive type of self-esteem. Schneider’s and Turkat’s method for distinguishing between displays of genuine versus
defensive self-esteem is correlating ratings of the latter with ratings on scales of social desirability, categorizing individuals scoring high on both scales as demonstrating a defensive rather than genuine self-esteem. A person with a high but defensive self-esteem actively guards against failure and repudiates it when it occurs, while a person with genuine high self-esteem does not consider failure to be threatening in the same way (ibid.).

**Peer relations**

Peer relations play an important role in children’s social and emotional development (Newcomb, Bukowski, & Pattee, 1993), as well as in cognitive development and learning (Greenman, Scheider, & Tomada, 2009). When forming peer-relations, children often choose playmates that are like themselves (Newcomb, Bukowski, & Pattee, 1993). However, what is more relevant to the establishment of peer relations is the extent to which children experience their interactions with peers as rewarding (ibid.). In order to engage in rewarding peer interaction children must possess social skills such as empathy and perspective taking. In order to achieve acceptance in the peer group, children must also be able to recognize and adjust to normative expectations of behavior (Gifford-Smith & Brownell, 2003).

Associations between peer-rejection and risk for later maladjustment are well established (Gifford-Smith & Brownell, 2003). Accordingly, peer relations is a well-researched subject. Early investigations led to the formulation of a theory placing children into one of five peer-status categories: *popular, rejected, neglected, controversial* and *average* (Coie, Dodge, & Coppotelli, 1982). These findings are based on a two-dimensional sociometric model, using *social preference* on one hand and *social impact* on the other (Newcomb, Bukowski, & Pattee, 1993). Social preference and social impact are measured through peer nominations. Coie, Dodge and Coppotelli (1982) describe the procedure collecting peer nominations accordingly: children are asked to name three peers whom they would most like to play with, and three peers whom they would least like to play with. Peer nominations are then standardized within classrooms (ibid.). Social preference is calculated by subtracting the number of negative peer nominations from the number of positive peer nominations, and social impact is estimated by adding negative and positive peer nominations together. Thus, social preference refers to whether a child is considered popular or unpopular by its peers, while social impact is a compiled measurement of the extent children are recognized by their peers at all (Newcomb, Bukowski, & Pattee, 1993).
Newcomb, Bukowski and Pattee (1993) conducted a meta-analysis on sociometric peer status as measured among 5-12 year old children. Based on their findings they formulated a set of descriptions of peer status. *Average* peer status refers to children with average scores across both social preference and social impact measures. *Popular* peer status describes a combination of high scores on social preference and average scores on social impact and is associated with high levels of sociability and cognitive abilities, but with low levels of aggression and social withdrawal. The popularity of these children has been attributed to possessing such traits as being helpful, considerate, following rules, and maintaining high academic performance. Children who score average on social preference and high on social impact are called *controversial*. The name controversial refers to children being both highly sociable and highly aggressive, assumedly attracting anger as well as amusement from peers (ibid.).

Rejected peers scoring average on social impact and low on social preference, are characterized as more aggressive, more withdrawn, and less sociable than other children (Newcomb, Bukowski, & Pattee, 1993). Rejected children often lack the social and cognitive skills needed to moderate their aggressive behavior – skills that may differentiate rejected children from controversial children. Withdrawal, demonstrated by rejected children, is expressed through elevated rates of depression and anxiety. These emotional and behavioral states are likely to obstruct a child’s chances of developing mutually rewarding relationships with other children (ibid.). Recent research has acknowledged the heterogeneity of children categorized as rejected, and highlights the importance of distinguishing between different sub-groups of rejected children, such
as *aggressive-rejected* versus *nonaggressive-rejected* children (Bierman, Smoot, & Aumiller, 1993).

*Neglected* peer status describes low scores of social impact and average scores of social preference. *Neglected* children demonstrate slightly fewer positive social interactions than popular peers and somewhat lower levels of aggression than all other children. According to Newcomb, Bukowski and Pattee (1993), however, *neglected* children do not appear to be at risk for social maladjustment to the same extent as *rejected* children. The authors claim that the critical predictive factor for maladjustment is the combination of aggression and withdrawal with social deficits, as specifically exhibited by *rejected* children. In contrast, Salmivalli and Isaacs (2005) claim that all kinds of disengaged social behavior put children at risk for social maladjustment and adverse treatment by peers. As in the case of *rejected* children, more recent research underlines the relevance of taking into account the heterogeneity of *neglected* children when investigating the risks associated with belonging to this particular peer-status group (Harrist, Zaia, Bates, Dodge, & Pettit, 1997).

Dodge et al. (2003) oppose the discourse in which children’s peer relations are described in terms such as “rejected” or “neglected” children. According to the authors, this type of wording implies that peer relations are in some way determined on the base of children’s individual psychological disposition, rather than being effects of peer group dynamics. Dodge et al. (ibid.) argue that peer rejection should instead be viewed as a negative life event, with the potential of creating long-term stress and severe inhibition of social development for children subject of peer-rejection.

A drawback of the research on relationships between peer-rejection and maladjustment is that most studies on this subject are based on correlational studies, hindering conclusions regarding causal links (Newcomb, Bukowski, & Pattee, 1993). These studies are often longitudinal, which is beneficial for the validity of results, although it is not enough to overcome the shortcomings of correlational data. However, correlations suggesting that peer-rejection is a contributing factor to developmental trajectories of social maladjustment and psychological problems have been found (Dodge et al., 2003). Whether peer-rejection functions as a trigger for negative development or merely exacerbates existing problems is still debatable. One hypothesis suggests a reciprocal relationship between a child’s experiences of peer interaction and its social, affective and emotional dispositions. These dispositions are an aggregate of genetic disposition and experiences of social interaction in other contexts. This would imply that lower levels of aggression and withdrawal at the onset of peer-rejection could serve as protection against stress caused by peer-rejection as well as having well-functioning social relations outside of the peer group (Newcomb, Bukowski, & Pattee, 1993). According to Coie et al. (1990; cited by Newcomb, Bukowski & Pattee, 1993), peer-rejection causes children to experience loneliness and anxiety as well as being the subject of deviant social reputation, all of which are stressful experiences. In addition, peer-rejection in itself hinders a child’s efforts to gain social support from both peers and adults. In the long run, fewer opportunities to engage in social interaction make it hard for children to develop social skills and coping strategies (ibid.).
**Relationships between externalizing behavior, self-esteem and peer-relations**

Because externalizing behavior includes verbal and physical aggression towards others, it is not surprising this behavior is related to problems with peer relations. Externalizing behavior is believed to complicate social interaction with peers and thereby endangering long-term social development (Karlberg, 2011). In studies by Coie, Dodge and Coppotelli (1982), and Newcomb, Bukowski and Pattee (1993), aggressive behavior towards others is directly associated with peer rejection. Dodge et al. (2003) investigated relations between peer-rejection and aggressive behavior using longitudinal data of children in pre-school, 1st, 2nd, and 3rd grade. The authors hypothesized that peer-rejection would exacerbate aggressive behavior through impeding development of social processing skills. This hypothesis held to be true. Analyses confirmed that peer-rejection was linked to increasing aggression over time; higher numbers of years of experiencing peer-rejection was associated with an exponential increase in aggression. The study further confirmed that the correlation between peer-rejection and aggression was moderated by social processing skills.

According to Gifford-Smith and Brownell (2003), however, the association between peer-rejection and externalizing behavior is complicated by the fact that externalizing behavior is perceived differently depending on age and norms in a given peer group. An assumption for the link between externalizing behavior and peer-rejection is that the manifestation of externalizing behavior in question is perceived as non-normative in the peer group. In general, the relationship between externalizing behavior and peer-rejection is weaker among younger children, indicating that externalizing behavior is more normative and accepted among peers at this developmental stage. This can, for example, be considered in the light of younger children not yet having developed more sophisticated ways of managing conflicts and communicating with peers (ibid.).

According to some researchers, peer relations are also related to self-esteem. Salmivalli and Isaacs (2005) found positive correlations between peer relations and experiences of self-esteem, which was further confirmed by Thomaes, Reijntjes, Orobio de Castro, Bushman, Poorthius and Telch (2010) who conducted a laboratory experiment with 8-12 year old children who received negative/positive feedback from perceived popular/neutral peers. Results showed that peer approval and disapproval had immediate effects on children’s self-esteem. Thomaes et al. (ibid.) underpins these findings relating cognitive and social development during childhood to the acquired ability to appreciate the approval of others and evaluate one’s global self-worth – abilities that are generally acquired around the primary school years (ibid.). On the contrary, no significant correlation between self-esteem and peer status was found in a study including 542 9th grade students (Bishop & Inderbitzen, 1995). However, having at least one reciprocal friendship relation was found to have a positive effect on self-esteem, as compared with having no friends at all (ibid.). The proposed correlation between peer relations and self-esteem has been questioned further by Baumeister, Campbell, Krueger and Vohs (2003), who found no support for this notion in their review of self-esteem and its effects on different aspects of performance and interpersonal relationships.

As for relationships between self-esteem and externalizing behavior, Donnellan et al. (2005) found a negative correlation between self-esteem on the one hand and aggression...
and anti-social behavior on the other hand, in older children and adolescents (as measured by self-reports, teacher and parent ratings). The correlation was small to moderate. The results of the study indicate low self-esteem in 11-year old children predict higher levels of aggressive and antisocial behavior two years later. However, a causal connection could not be fully established due to limitations in the study’s design. While valid in relation to older children and adolescents, there is some evidence that the relationship between externalizing behavior and self-esteem is somewhat different for younger children. Mathias, Beibl and Dilalla (2011) investigated this relationship within a group of pre-school children and found that children who overestimated their own cognitive and social abilities received higher teacher ratings of externalizing behavior.

As mentioned earlier, research on self-esteem has been questioned because of the limitations associated with self-report measures. Moreover, research examining relationships to other constructs and outcomes has been questioned because of the lack of data supporting a causal link between self-esteem and increased performance or other positive outcomes for the individual (Baumeister, Campbell, Krueger & Vohs, 2003). Studies have established a correlational link between self-esteem and other objective measures such as school performances (e.g., Davies & Brember, 1999) and behavior as rated by others (i.e., aggression; Donnellan et al., 2005; Mathias, Beibl, & Dilalla, 2011). However, it is unclear whether self-esteem is the cause of the outcomes or if it is a product of them, or if these relationships are mediated by some other variable. Assuming that self-esteem is an effect of success rather than a cause of it, Forster (2013) argues that interventions aiming at boosting self-esteem would be pointless. However, even though self-esteem cannot readily be claimed to be the cause of positive outcomes for the individual, it is an interesting subject of investigation because of the established correlations with other constructs and outcomes.

In sum, externalizing behavior, self-esteem and peer-relations seem to be interrelated in some ways. Considering the known associations between these constructs and developmental trajectories in childhood, investigating potential effects on self-esteem and peer-relations of school-based intervention aimed at reducing externalizing behavior in children is highly motivated.

**School-based interventions**

The intervention in focus for this study is a class-level behavior management program called Comet. Comet was developed in collaboration between the Social Services Office in Stockholm and the Department of Psychology at Uppsala University, intended for children exhibiting externalizing behavior and experiencing peer-rejection (Karlberg, 2011). It was developed with the School-Wide Positive Behavior Support (SWPBS) approach in mind because of the growing evidence of its success. In recent years, SWPBS has emerged as a promising brew for the effective prevention of problematic behavior among children at school (Solomon et al., 2012). SWPBS combines the principles of applied behavioral analysis with a focus on teachers’ role in shaping behavior. This approach promotes the reinforcement of desirable behaviors in children. SWPBS also represents a systems approach where existing structures and cultures at schools provide the framework for intervention implementation. SWPBS-intervention implementation is guided by a three-tiered model rooted in the Response to
Intervention-approach (RTI) which involves all of the children in a particular preventative level; providing specialized prevention in smaller groups for children who exhibit mild to moderate behavior problems and treating children with severe behavior problems at an individualized level (ibid.).

A recent meta-analysis showed that SWPBS is effective in changing student behavior (Solomon et al., 2012). Observed effect sizes were in the low to medium range ($R^2 = .44$ to .68), according to the authors’ benchmarks (low = .35; medium = .65; high = .90). In general, behavior oriented school-based interventions seem to be efficient and highly appreciated among school personnel, due to its focus on structure, rules, and reinforcement of desirable behavior (Karlberg, 2011). Interventions based on behavior management consider how the environment can be adapted to fit the needs of children who experience trouble at school. This way, stigma caused by attribution of behavior problems to the targeted child is avoided.

Wilson and Lipsey (2007) conducted a meta-analysis of school-based interventions for aggressive and disruptive behavior and found that behavior focused interventions were superior to all other intervention modalities included in the study (these included cognitively oriented interventions such as anger management or problem-solving training, social skills training, and individual therapy). Wilson and Lipsey (2007) categorized school-based interventions accordingly: Universal programs (interventions delivered in classroom settings without any specific target students), indicated programs (interventions provided to targeted students, either inside or outside the classroom), special schools (interventions delivered to children in an education setting adapted to students with special needs), and multimodal programs (involving multiple intervention approaches). The analysis revealed that indicated programs were more effective in reducing aggressive/disruptive behavior than the other intervention types, by a mean effect size of $d = 0.29$. According to Wilson and Lipsey (ibid.), an effect size of this proportion represents a decrease in problem behavior of practical significance in the school environment.

Wilson and Lipsey (2007) also found that indicated programs were most effective when delivered in a classroom setting, as opposed to taking targeted children out of the ordinary education context. According to Karlberg (2011), both targeted children and their peers benefit from classroom delivered interventions. First of all, class-level interventions have the potential of reaching out to children that have not yet started to display problematic behavior, but may in the future. Secondly, it has been proven that the effects of intervention on targeted children are enhanced when the whole class is included in the program (ibid.). Yet another advantage of an inclusive intervention is stigmatization of the targeted children is minimized.

**Communication Method (Comet).**

Comet, which is the intervention of focus in this study, is an indicated intervention delivered at the classroom level. The long-term goals of Comet are to prevent maladjustment associated with externalizing behavior during childhood, such as delinquency, substance abuse, mental illness and antisocial behavior (Forster, Karlberg, Hammarberg, & Hellman, 2009). The basic assumption underlying Comet is that the
most effective way of modifying problematic student behavior is changing the behavior of teachers (ibid.). Comet uses behavior modification based on behavioristic learning theory of how behavior can be shaped, changed and maintained through contingency alteration (Wadström, 2013). More specifically, behavior can be modified by changing the antecedents or consequences of the behavior (Ramnerö & Törneke, 2006). The basic underlying assumption of behavior modification is that individuals are motivated to engage in behaviors associated with consequences that are considered appetitive such as praise and attention from a teacher or parent, or one’s favorite food. By introducing a desirable stimulus in connection with a desirable behavior, the probability of that behavior occurring in the future increases. Likewise, the frequency of an undesirable behavior decreases when paired with undesirable consequences, e.g., revoked allowance. Changing consequences that maintain a child’s externalizing behavior means staging a new learning process that motivates the child to engage in new and adaptive forms of behavior (Karlberg, 2011; Wadström, 2013). Behavior modification has proven to be an effective intervention method for reducing externalizing behavior (Karlberg, 2011).

In Comet, teachers are encouraged to change their behavior in order to foster constructive interaction in the classroom (Forster, Karlberg, Hammarberg, & Hellman, 2009). As opposed to using punishment and reprimand when responding to externalizing behavior, teachers are encouraged to reinforce adaptive behavior and the development of social skills in students. Beside reduction of externalizing behavior, this is also intended to increase students’ academic motivation and to create a good work environment for both students and teachers (ibid.). In Comet, teachers attend eight group sessions led by certified Comet instructors that provide guidance and consultation throughout the program. The first part of the program is devoted to behavior analysis of the targeted student, followed by planning and implementation of individual and class-level interventions. The final part is devoted to evaluation and maintenance of results. Information gathering of target behaviors and intervention testing take place continuously between sessions, all of which is subject of evaluation and development in cooperation with the certified instructor during the sessions.

Comet has proven to be effective in reducing externalizing behavior in 100 targeted children in a large randomized controlled study (RCT-study) by Forster, Sundell, Morris, Karlberg, and Melin (2012). The study compared Comet to an active control condition represented by a non-structured psychosocial program called Project Charlie, which is a universal program (Karlberg, 2011) created in the 1970’s as a drug prevention program (European Monitoring Centre for Drugs and Drug Addiction, 2012) with the aim of increasing children’s emotional and social skills (Karlberg, 2011). The pronounced goals of Project Charlie are to improve self-esteem and conflict resolution skills, and to reduce bullying (ibid.). The program is based on a manual with descriptions of 74 lesson plans on emotional and social learning (Forster, 2010). Unlike Comet, Project Charlie is not based on behavioristic learning theory or the assumption that teachers’ behavior is the key to changing children’s behavior. Instead, the goals of Project Charlie are sought through traditional teaching and learning exercises on a number of specific topics, such as shaping a positive self-esteem, expressing emotion, improving communication skills and empathic ability, accepting differences between
people, practicing conflict management and cooperation, and resisting peer-pressure (Project Charlie AB).

Project Charlie does not address externalizing behavior in targeted students, nor does it encourage teachers to use behavior management strategies (Forster et al., 2012). The authors did not expect Project Charlie to be effective in reducing externalizing behavior or in reducing teacher reprimand. Instead, the purpose of choosing Project Charlie as comparison group to Comet was to provide the teachers in the comparison group with credible attention, as opposed to using a wait-list or no-treatment control group (ibid.).

The targeted children in the study by Forster et al. (2012) were 7-9 years old and attended first, second or third grade in regular education classrooms. Observer, teacher and student ratings were collected at pre-test, post-test (6 months after pre-test) and follow-up (8 months after post-test). Results at post-test showed that Comet was more efficient than Project Charlie in reducing externalizing behavior in targeted students. At follow-up, significant intervention effects favoring Comet was recorded for teacher-rated externalizing behavior, observer-rated externalizing behavior, hyperactivity, positive peer nominations, and teacher reprimand. Effect sizes were in the medium and large range. Teacher reprimand had a significant moderating effect on externalizing student behavior. Thus, reduction of teacher reprimand enhanced the effects attributed to the intervention. The effects of Comet on externalizing behavior increased by post-test, which the authors discuss in relation to most students having stayed with their teachers during the entire research period. It is possible that the Swedish school system, where students stay with the same teacher for several years, provides a particularly beneficial environment for implementing interventions such as Comet (ibid.).

The existing research done on the efficacy of Comet is positive and provides a base for future investigation. These promising results, together with the growing popularity of the program in Sweden, motivate further evaluation. The data recorded in conjunction to the study included information of both targeted students and their peers (N = 2,834), primarily allowing for normative comparison, but also allowing for further investigation of the data set, which has served as basis for the writing of this master thesis.

**Current study**

The purpose of this study was to explore the potential effects of Comet beyond those already established concerning externalizing behavior in targeted children, in comparison to an active control condition (Project Charlie). Because Comet is an indicated program administered at class level, it may create benefit beyond the targeted students to also include their peers. Because externalizing behavior is related to self-esteem and peer relations, such measures can be combined to create additional insight to class-level effects of Comet. Exploration of class-level effects can be further elaborated by distinguishing different profiles of children, in order to investigate whether potential class-level effects of Comet vary between profiles. By using data of all 2,834 children included in the RCT-study by Forster et al. (2012), profiles based on measures of externalizing behavior, self-esteem and peer relations can be mapped at pre-test, allowing for investigation of effects of Comet at post-test and follow-up with regard to different profiles of children. In addition, observation of general developmental
Trajectories in different profiles can contribute further to the understanding of relationships between externalizing behavior, self-esteem and peer relations in primary school children.

**Research questions**
1. What profiles will emerge from the child data at pre-test, and how do the profiles develop over time?
2. Does the effect of Comet on externalizing behavior, self-esteem and peer relations vary across the profiles?
3. Are the results maintained over time?

**METHOD**

Data material used for this master thesis was collected in conjunction to the RCT-study by Forster et al., published in the Journal of Emotional and Behavioral Disorders in 2012 (issue 20, p. 169-183). Forster (2010) and Karlberg (2011) also used the study as part of their doctoral theses. The data material was retrieved from and by permission of Martin Forster.

**Participants**

The study by Forster et al. (2012), onwards referred to as “the RCT-study”, included 100 target children exhibiting externalizing behavior. In the current study, data from the target children and their classmates was used, adding up to a total number of 2,834 participants. The children were 7-9 years old and attended 1st, 2nd, and 3rd grade in regular education classrooms (ibid.). They were distributed in 88 classrooms at 38 schools in Stockholm, and were a representative sample for the population of Stockholm in terms of socioeconomic factors such as parent income, marital status and employment rate (Karlberg, 2011). 1,598 children were randomized to Comet and 1,236 to Project Charlie. A Chi-Square Test showed no significant differences between the two groups concerning distribution of boys and girls. However, distribution across grades varied significantly between the groups. More 2nd grade children than expected were assigned to Comet, whereas more 1st and 3rd graders than expected were assigned to Project Charlie ($\chi^2(2) = 139.73, p < .001$).

522 children (18.4 % of the sample) were not included in the current study due to missing data at pre-test on one or several outcome measures chosen for analysis, thereby leaving 2,312 children for inclusion in the data analysis. A Chi-Square Test revealed no significant differences between included and excluded students concerning sex ($\chi^2(1) = 2.8, p = .09$) or intervention condition ($\chi^2(1) = 1.7, p = .19$). Variations regarding grade were however significant, with slightly more children in second grade than expected represented in the attrition ($\chi^2(2) = 7.39, p = .03$; see Table 1).
Table 1 Distributions of included/excluded students.

<table>
<thead>
<tr>
<th>N = 2,834</th>
<th>Sex</th>
<th>Grade</th>
<th>Intervention</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Girls</td>
<td>Boys</td>
<td>1st</td>
</tr>
<tr>
<td>Included (n = 2312)</td>
<td>1,109</td>
<td>1,194</td>
<td>796</td>
</tr>
<tr>
<td>Excluded (n = 522)</td>
<td>227</td>
<td>288</td>
<td>156</td>
</tr>
</tbody>
</table>

Procedure
Selection of targeted students for the RCT-study was performed by announcing the study at schools in the area of Stockholm, whereby 189 teachers indicated interest in participating (Forster et al., 2012). Thereafter, targeted students were selected in collaboration between teachers and researchers, after rating externalizing behavior of potential target students using a 5-point rating-scale (1 = not at all, 5 = very much). One hundred 7-8 year old children who were all rated at least one standard deviation above average on externalizing behavior were selected for inclusion in the RCT-study. Due to the greater availability of certified Comet-instructors than Project Charlie-instructors at the time, 60 children were randomized to Comet and 40 to the comparison group, Project Charlie. Observer, teacher and student ratings were collected at pre-test, post-test (6 months after pre-test) and follow-up (8 months after post-test). In addition to data of targeted students, a number of measures were also collected from peers, which altogether formed the data set of 2,834 children that was used in this study. Measures collected from both targeted students and peers were ratings of self-esteem, positive and negative peer nominations, teacher ratings of externalizing behavior and performance in mathematics and word comprehension. The latter two measures were not included in the current study.

Comet
The teachers of target students who were assigned to Comet attended nine three-hour sessions in groups of four to five teachers, over a period of five months (Forster et al., 2012). The group sessions were led by certified Comet-instructors that provided guidance and consultation throughout the program. During the program, teachers were instructed to use the principles of behavior management in performing a Functional Behavioral Assessment (FBA) in order to identify target behaviors and formulate goals for changing these (ibid.). The first six sessions were devoted to creating a case formulation and an intervention plan regarding the targeted student, whereas the last three sessions concerned maintenance and class-wide behavior management strategies (ibid.). The teachers administered all individual and class-wide interventions in the classroom by themselves, which has shown to be more effective than having psychologists, peers, or researchers administrating interventions in classrooms (Wilson, Lipsey, & Derzon, 2003). Examples of class-wide behavior management strategies included formulation of class rules; use of token economy; dividing the class into two teams and letting them compete over which team displayed the largest number of prosocial behaviors in the classroom; conflict management workshops where the teacher let the children practice conflict management strategies with each other; and Class-Wide Peer Tutoring, which aims at increasing learning among students by letting them actively give instructions and feedback in pairs (Karlberg, 2011).
Because the program was tailored to the specific problem behaviors exhibited by the targeted student, interventions delivered at class-level varied somewhat (Karlberg, 2011). All teachers performed the FBA and reinforced target behaviors in targeted students accordingly. According to teacher reports, approximately two thirds of the teachers implemented formulation of class rules and about three quarters chose to implement token economy. Less than half of the teachers implemented the Class-Wide Peer Tutoring-intervention. After the last session, teachers were instructed to choose for themselves whether they wanted to continue to implement interventions after post-test and until follow-up, or not (ibid.). No control of intervention adherence was carried out during the research period.

The comparison group
Teachers of the students assigned to Project Charlie received two full-day workshops in groups of 20 (Forster, 2010). The workshops were administered right after pre-test measurements, and consisted of a review of the 74 lesson plans described in the program. In order to make comparisons between Comet and Project Charlie more plausible, the teachers in the comparison group also received a three-hour session in groups of 4-5 for discussing applications of the program with regard to the targeted students. According to how the program is generally implemented, teachers received no further consultation or instruction after having finished the workshops at the beginning of the study. Teachers were free to choose any elements of the program for implementation in their class, but were asked to implement a total of 15-20 lesson plans during the research period. Since no further instructions were given, teachers were free to continue or to stop implementing lesson plans during the period between post-test and follow-up. According to teacher reports, the mean number of implemented lesson plans was 13 at post-test and 19 at follow-up (Karlberg, 2011). No control of intervention adherence was carried out during the research period.

Measures
For the purpose of measuring externalizing behavior, self-esteem and peer relations, four measures from the original RCT-study were used. The measures are described in more detail in the following section.

Brief Teacher Rating of Externalizing behavior (BREB).
A brief teacher rating was used as a measure of externalizing behavior (BREB). BREB consists of three items, rated from 1 (almost never) to 5 (almost always). The first item, “aggressive”, relates to how often a child starts fights, says mean things to other children, or pushes or hits children. The second item, “disruptive”, relates to how often a child disrupts play, if the child has trouble sharing and only wants to play on his or her own terms. The third item, “peer-aggression”, relates to being intrusive, sabotaging ongoing play, and being disliked by peers. Accordingly, this measure corresponds well with the general definition of externalizing behavior as presented in previous research (e.g., Donnellan et al., 2005; Mathias, Beibl & Dilalla, 2011; Ybrandt & Armelius, 2010; “aggression”, “anti-social behavior”, “conduct problems”, “lack of self-control”). Being a teacher rating scale, BREB can be assumed to reflect teachers’ apprehension of students in relation to norms. Therefore BREB can be considered to correspond to
Karlberg’s definition of externalizing behavior as behaviors that breach the norms of a given context (2011). Scores on the three items are averaged to create a total score for each child. The scale has demonstrated adequate test-retest reliability in previous studies \((r = .72; \text{Forster et al., 2012})\), as well as correlation to peer ratings \((r = .41; \text{ibid.})\). The internal consistency of the scale (Cronbach’s alpha) in our sample was \(\alpha = .82; \alpha = .80\) and \(\alpha = .80\) at pre-test, post-test and follow-up, respectively. An Independent Samples T-test showed no significant variations in BReB-ratings between children assigned to Comet and Project Charlie at pre-test \((t[2310] = .82, p = .41)\).

*I Think I Am (ITIA).*

Self-esteem was measured by a self-rating scale called *I Think I Am* (ITIA). ITIA is a measure of how children evaluate themselves concerning physical characteristics, psychological characteristics, and interpersonal relationships. Thus, as measure of self-esteem, ITIA can be considered consistent with the definition of self-esteem as presented by Baumeister, Campbell, Krueger and Vohs (2003; “the value people place on themselves”; “the evaluative component of self-knowledge”). The scale can be divided into five subscales: physical appearance, talents and skills, psychological well-being (in terms of psychological stability, anxiety and aggressiveness), relationships to parents and family, and relationships to teachers and peers (Ouvinen-Birgenstam, 1985). For the purpose of enabling comparison to previous research, the total scale was used as measure in this study. The scale is available in two versions, a shorter version for 7-9 year olds, which was used in the present study, and a longer version for 10-16 year olds. The shorter version consists of 32 items rated “yes” or “no”. Half of the items are positively phrased (e.g., “I have a nice face”), and half are negatively phrased (e.g., “My parents are disappointed in me”).

The scale has demonstrated adequate internal consistency by split-half method \((r = .82\) and .85 for boys and girls, respectively). The concept validity of the scale has been evaluated in four studies (Ouvinen-Birgenstam, 1985): First, scores on ITIA of children with high rates of non-attendance at school were compared to scores of children with less non-attendance. The hypothesis of a negative relation between non-attendance at school and psychological well-being was confirmed. Second, scores on ITIA were compared to observations made by psychologists, with the assumption of children with low self-esteem being more reserved than children with high self-esteem, which was also confirmed. Third, children that experience peer exclusion were assumed to demonstrate poorer psychological well-being. This was confirmed by a negative correlation between peer-exclusion and scores on ITIA. Last, scores on ITIA were compared to another self-report measure of self-esteem, demonstrating positive correlations. Altogether, ITIA is considered to be a valid measure of self-evaluation (ibid.). Cronbach’s alpha was calculated for correlations in the current data set between the five subscales, demonstrating adequate internal consistencies of \(\alpha = .83; \alpha = .76\) and \(\alpha = .77\) at pre-test, post-test and follow-up, respectively. An Independent Samples T-test showed that children assigned to Comet demonstrated a significantly lower overall mean on the total scale at pre-test \((M = -.03, SD = 1.01)\) than children assigned to Project Charlie \((M = .07, SD = .97), (t[2310] = 2.36, p = .02)\).
Social Preference and Social Impact.
Scores of social preference and social impact were used as measures of peer-relations. Scores were calculated according to how Coie, Dodge & Coppotelli (1982) originally described the procedure of measuring peer relations: social preference was calculated by subtracting the number of negative peer nominations from the number of positive peer nominations, and social impact was estimated by adding negative and positive peer nominations together. Peer nominations have demonstrated good reliability and validity, as well as correlations to observed behavior as well as behavior problems later in life (Forster et al., 2012). Peer nominations were collected by research assistants in accordance to descriptions by Coie, Dodge and Coppotelli (1982). First, the research assistant handed out lists with the names of all children in a given class. Second, the research assistant read each name out loud, as to ensure that all children recognized each name on the list. Last, children were asked to circle the names of three peers whom they would most like to play with, and the names of three peers whom they would least like to play with. Afterwards, peer nominations were standardized within classrooms (Forster et al., 2012). An Independent Samples T-test revealed no significant differences concerning social preference or social impact scores between children assigned to Comet and Project Charlie at pre-test ($t[2310] = .08$, $p = .94$; $t[2310] = -.14$, $p = .89$).

Data analysis
The data analysis was performed in two steps: first, by cluster analysis to answer the question as to what profiles will emerge from the child data at pre-test. Thereafter Mixed analysis of variance (MANOVA) for repeated measures was performed with regard to the entire sample as well as for each individual profile, to answer the questions as to whether the effect of intervention on externalizing behavior, self-esteem and peer relations varied across the profiles. MANOVA was also performed in order to investigate the maintenance of intervention effects over time as well as to observe general developmental trajectories in the profiles. All analyses were performed using the SPSS 21.0 software.

Cluster analysis.
Cluster analysis is a multivariate statistical method for identifying underlying structures in a data set, by grouping respondents (or objects) into profiles on the basis of similarities over a given set of variables (Verma, 2012). As described by Pell and Hargreaves (2011), this method is relevant when looking at relationships between people rather than variables. In cluster analysis, people are sorted into groups with people whom they are more similar to, than to the extent which they are dissimilar to people in other groups (ibid.). Individuals within profiles will thereby be homogenous in regard to one or several aspects, and heterogeneous to respondents in other profiles. Cluster analysis is a useful tool for conducting exploratory research and generating hypotheses, especially when handling large data sets. Thus, cluster analysis was considered a suitable method of analysis for identifying profiles in our data.

The SPSS 21.0 software offers the possibility to perform hierarchical, K-means, or TwoStep cluster analysis (Verma, 2012). According to Verma (ibid.), hierarchical cluster analysis is most suitable for smaller data sets, and K-means cluster analysis is
useful for conducting confirmatory analysis when there is solid theory for making
assumptions of what number of relevant profiles a data set should aggregate. The
purpose of TwoStep cluster analysis is to identify natural cluster formations in large
data sets. TwoStep cluster analysis can be performed either with or without
predefinition of number of profiles, and is the only method available in SPSS that
allows combinations of continuous and ordinal data in the analysis (ibid.). Because the
purpose of this study demanded exploratory analysis of a very large dataset with
continuous (ITIA; Social Preference and Social Impact) as well as ordinal (BREB)
variables, TwoStep cluster analysis was chosen as the preferred method.

In order to achieve equal effect of each variable on a profile solution, data must be
standardized (ibid.). In the case of externalizing behavior and self-esteem, this was
achieved by transforming all data into z-scores. As for scores of social preference and
social impact, data was already standardized within classrooms. The log-likelihood
criterion was chosen as distance measure for the cluster analysis. The distance measure
is the condition under which profiles are distinguished, and log-likelihood is the only
choice offered in SPSS 21.0 that allows for analysis when data is both continuous and
ordinal (Norusis, 2011), as was the case with our analysis. Assumptions for performing
an optimal TwoStep cluster analysis are the following: variables should be independent
so as to not demonstrate strong bivariate correlations, data should be normally
distributed, and categorical data should have a multinominal distribution (Verma, 2012).
Bivariate correlations between the dependent variables were plotted, revealing moderate
correlations between social preference and externalizing behavior ($r = .40; p < .001$) and
between externalizing behavior and social impact ($r = .24; p < .001$). Histograms
showed that scores of social impact and social preference were normally distributed,
whereas ratings of BREB were positively skewed and ratings of ITIA were negatively
skewed. Thus, our data did not meet all of the criteria assumed for cluster analysis.
However, being primarily a method of exploratory analysis and a mean of testing
hypothesis and calculating significance levels secondarily, clustering data that do not
fully meet the assumptions is considered perfectly acceptable (Norusis, 2011).

**Mixed ANOVA for repeated measures.**

After the children had been clustered into profiles, intervention effects and general
developmental trajectories were analyzed through Mixed ANOVA (MANOVA) for
repeated measures. The MANOVA was designed accordingly: intervention * outcome
measures * assessment point (2 * 4 * 3). Missing values at post-test and follow-up was
handled by intent-to-treat method, that is, last observation was carried forward. Intent-
to-treat method is a conservative way to handle missing data and is often used in clinical
research. Data for each variable was plotted into histograms by profile (as determined in
the previous cluster analysis), assessment point and intervention type, indicating
moderately negatively skewed data of ITIA and moderately positively skewed data for
BREB. Transformations of data (by $\log_{10}$ and square root) did not improve skewness on
these measures; hence, untransformed data was used. Moreover, Mauchly’s Test of
Sphericity generated significant results for externalizing behavior, self-esteem, social
preference and social impact ($\chi^2 (2) = 127.7; 171.4; 87.1; 60.5, p < .001$), thereby
violating the assumption of homogenous variances. This is especially problematic in
small samples. As recommended in such cases a more conservative criterion, Huyhn-
Feldt, was used to interpret observed values and level of significance. Huyhn-Feldt corrects the degrees of freedom to take the violation of the assumption of homogenous variances into account (Borg & Westerlund, 2012). Post hoc-tests by Least Significant Difference were performed to investigate main and interaction effects in more detail.

Follow-up analyses with the purpose of investigating recorded effects in more detail were performed by MANOVA for repeated measures for each individual profile. The Huyhn-Feldt criterion was used for interpretation since the assumption of sphericity was violated. Post Hoc-tests by Least Significant Difference were performed. Bonferroni-adjustment was used to reduce the risk of Type I error, which increases when multiple analyses are performed on a single dataset. Because the design was 5 x MANOVA, the alpha-level was set at $p = .01 \, (0.05/5)$.

Cohen’s $d$ ($d$) was used as effect size measure to compare Comet and Project Charlie in the case of significant interaction effects of time and intervention within profiles. The guidelines for interpreting Cohen’s $d$ are: 0.2 (small effect size); 0.5 (medium effect size), and 0.8 (large effect size; Cohen, 1988). Eta-squared ($\eta^2$) was used as effects size measure to describe effects other than those regarding intervention, e.g., main effects of time between or within profiles. The value of eta-square represents the portion of variance in a dependent variable accounted for by the independent variable/variables in a sample (Borg & Westerlund, 2012). To interpret eta-squared values, guidelines proposed by Cohen (1988) were used: $\eta^2 = .01$ (1% accounted variance, small effect size); $\eta^2 = .06$ (6% accounted variance, medium effect size); $\eta^2 = .14$ (14% accounted variance, large effect size).

**Attrition analysis**

Missing data from the 522 children excluded from the cluster analysis at pre-test, as well as missing data at post-test and follow-up of children included in the analysis, was considered at large to be attrition due to expected difficulties with coordinating retrieval of teacher and student ratings in this large project. A Missing Values Analysis (MVA) was performed to investigate patterns of missing data of the children included in the MANOVA. The overall attrition was 12% at post-test and 25% at follow-up. Cases with missing data generally displayed attrition at more than one variable – particularly at follow-up, where the predominant pattern was attrition at all four variables. These cases were often represented at group level in the data set, suggesting that attrition belonged to entire classes. A possible explanation for this is that some teachers quit their job during the research period, which meant that the intervention would have been discontinued and that research assistants thereby might have chosen to cancel further data gathering from these classes. For 87 children, all data from both post-test and follow-up was missing. A likely explanation to this is that some children moved to a different school during the research period.

Another notable pattern of attrition were that in some cases, only one of the following outcome measures was missing: externalizing behavior at post-test, self-esteem at post-test, externalizing behavior at follow-up, or self-esteem at follow-up. Unlike peer nominations, measures of externalizing behavior and self-esteem were retrieved by teachers and not by research assistants. Therefore, this might have been a reflection of
factors such as time pressure or other work-related issues having affected whether teachers completed their ratings or not. Regarding cases of attrition that were considered to be random – that is, attrition that did not coincide with the general patterns – it is plausible that some students may not have been in class at one assessment point when student ratings were obtained, which would however not prevent the teacher from completing teacher ratings of the students in question.

Ethical considerations
At the time when this data was collected (1999–2000), any general demand of formal ethical approval of research studies had not yet been implemented in Sweden. Nevertheless, appropriate measures were taken by the authors of the RCT-study to assure the welfare and anonymity of the participants (Forster, personal communication, 9-23-2013). The parents were informed about the study and were required to give their written consent in order for the children to be included in the study. All data had already been anonymized when received from Martin Forster.

Concerns regarding collection of peer nominations have been raised in previous studies, since children are asked to explicitly name the children they like the best and the least in the group (Coie, Dodge, & Coppotelli, 1982; Newcomb, Bukowski, & Pattee, 1993). However, collection of negative and positive peer nominations provides information that is both vital and impossible to gain if you only ask for positive peer nominations (e.g., for distinguishing rejected from neglected children). The authors of the RCT-study considered possible disadvantages with the procedure of gathering peer nominations, but concluded that they did not outweigh the necessity of the information. Nevertheless, they opted for an acceptable solution by letting research assistants collect peer nominations instead of the children’s teachers.

RESULTS

Profiles
The TwoStep cluster analysis generated a solution of four profiles based on the outcome measures (Table 2). The profiles were of relative even sizes and the outcome measures were weighted equally in the analysis, thereby producing an acceptable profile solution. The profiles are described below.

Profile 1, “Externalizing-Rejected”: Children within the first profile received the highest ratings on both externalizing behavior and social impact. Ratings of externalizing behavior were markedly higher than in the other profiles; over one standard deviation above average. Social impact scores were only slightly higher than those of the profile with the second highest scores, and still within one standard deviation above average. Social preference scores were the lowest of all profiles, over one standard deviation below average. Self-esteem ratings were in the average range.

Profile 2, “Popular”: Children within this profile were characterized as being most popular among peers, with the highest social preference scores of all profiles. These scores were however within one standard deviation above average. Scores of social impact and self-esteem were the second highest among the profiles, although still within
the average range. Children in this profile exhibited levels of externalizing behavior slightly below the overall mean.

Profile 3, “Insecure”: These children were characterized by low scores on self-esteem, which exceeded one standard deviation below the overall mean and all other profiles. They were rated average on externalizing behavior and social preference, while slightly below average on social impact.

Profile 4, “Quiet-Cool”: Children in the fourth profile stood out by having the least social impact and exhibiting the lowest levels of externalizing behavior, although both measures were within one standard deviation below average. Self-esteem ratings were the highest of all profiles, but only slightly higher than those of Profile 2. Social preference scores were average.

Table 2 Mean scores of outcome measures by profiles (based on data from pre-test).

<table>
<thead>
<tr>
<th>Profile</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>1-4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
<td>18.1 %</td>
<td>27.3 %</td>
<td>26.7 %</td>
<td>27.9 %</td>
<td>M (SD)</td>
</tr>
<tr>
<td></td>
<td>(n=419)</td>
<td>(n=631)</td>
<td>(n=617)</td>
<td>(n=645)</td>
<td>(N = 2312)</td>
</tr>
<tr>
<td>Externalizing behavior</td>
<td>1.40</td>
<td>-0.33</td>
<td>-0.06</td>
<td>-0.55</td>
<td>0 (1)</td>
</tr>
<tr>
<td>Self-esteem</td>
<td>-0.02</td>
<td>0.53</td>
<td>-1.22</td>
<td>0.71</td>
<td>0 (1)</td>
</tr>
<tr>
<td>Social Impact</td>
<td>0.88</td>
<td>0.76</td>
<td>-0.37</td>
<td>-0.90</td>
<td>0 (1.15)</td>
</tr>
<tr>
<td>Social Preference</td>
<td>-2.00</td>
<td>1.43</td>
<td>0.18</td>
<td>-0.12</td>
<td>0 (1.58)</td>
</tr>
</tbody>
</table>

Chi-Square Tests revealed no significant differences between the profiles with regard to intervention assignment, however, differences concerning sex and grade were significant ($\chi^2 (3) = 85.86, p < .001; \chi^2 (6) = 160.85, p < .001$). Boys were overrepresented in Profile 1 (“Externalizing”), whereas girls were overrepresented in Profile 4 (“Quiet/Cool”). As for grade, a dominant part of all 3rd graders were clustered into profile 3 (“Insecure”), a profile in which 1st graders were underrepresented. Instead, a larger portion of 1st graders than expected was found in Profile 2 (“Popular”).

Effects across profiles and assessment points

Results from the MANOVA for repeated measures including all profiles are initially described, followed by reports of analyses of each individual profile. Recorded effects in the MANOVA including all profiles are displayed in Table 3 and Figure 2 on page 22. Means and standard deviations of outcome measures from each MANOVA of each individual profile are presented in tables 4, 5, 6, and 7 on pages 22-23.

Multivariate tests of within-subjects effects showed a significant interaction effect of assessment point and profile ($F [24, 16052.18] = 88.09, p < .001$), which confirmed that the profiles developed independently over time. The effect size for this interaction was large. Univariate tests showed that the interaction effect of assessment point and profile was significant for each outcome measure. Furthermore, multivariate tests of within-subjects effects also revealed a small but significant interaction effect of assessment point, profile and intervention ($F [24, 16052.18] = 2.36, p < .001$), indicating that effects
of Comet in the different profiles were distinguished from effects of the comparison intervention.

Tests of between-subjects effects yielded significant main effects of profile on all outcome measures, which again confirmed that the cluster analysis had produced independent profiles. The profiles accounted for 42% of the variance in externalizing behavior ($F[3, 2302] = 546.91, p < .001$); 35% of the variance in self-esteem ($F[3, 2302] = 420.31, p < .001$); 49.8% of the variance in social preference ($F[3, 2302] = 574.00, p < .001$); and 30% of the variance concerning social impact ($F[3, 2302] = 336.78, p < .001$). Post-Hoc tests showed that the main effect of profile was valid for differences between all profiles on every outcome measure.

![Graphs showing estimated marginal means of profiles across assessment points, plotted by outcome measure.](image)

**Figure 2** Estimated marginal means of profiles across assessment points, plotted by outcome measure.
Table 3 MANOVA for repeated measures including all profiles. Interactions effects of Time and Profile (T x P); Time, Profile and Intervention (T x P x I); and Profile and Intervention (P x I) on outcome measures (pre-test to follow-up).

<table>
<thead>
<tr>
<th></th>
<th>N = 2,312</th>
<th>T x P</th>
<th>T x P x I</th>
<th>P x I</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F(df)</td>
<td>F(df)</td>
<td>F(3)</td>
<td></td>
</tr>
<tr>
<td>Ext. behavior</td>
<td>32.09(5.70)**</td>
<td>3.41(5.70)*</td>
<td>2.911'</td>
<td></td>
</tr>
<tr>
<td>Self-esteem</td>
<td>226.72(5.50)**</td>
<td>2.28(5.50)'</td>
<td>3.59*</td>
<td></td>
</tr>
<tr>
<td>Social Preference</td>
<td>35.84(5.81)**</td>
<td>.62(5.81)</td>
<td>.41</td>
<td></td>
</tr>
<tr>
<td>Social Impact</td>
<td>76.02(5.87)**</td>
<td>3.35(5.87)*</td>
<td>1.75</td>
<td></td>
</tr>
</tbody>
</table>

**Within-subjects effects; ** Between-subjects effects.
* Significant at the .01-alpha level; ** p < .001.
1 Trend (p < .05).

Table 4 Means (and Standard Deviations) of Outcome Measures for Profile 1 at Pre-test, Post-test and Follow-up.

<table>
<thead>
<tr>
<th></th>
<th>n = 419</th>
<th>Externalizing behavior</th>
<th>Self-esteem</th>
<th>Social Preference</th>
<th>Social Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Pre-test</td>
<td>Post-test</td>
<td>Follow-up</td>
<td></td>
</tr>
<tr>
<td>Comet</td>
<td></td>
<td>3.11 (1.00)</td>
<td>2.75 (1.07)</td>
<td>2.60 (1.12)</td>
<td></td>
</tr>
<tr>
<td>(n = 238)</td>
<td></td>
<td>16.79 (9.67)</td>
<td>19.00 (9.45)</td>
<td>19.90 (9.84)</td>
<td>-2.00 (1.43)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-2.00 (1.43)</td>
<td>-1.65 (1.79)</td>
<td>-1.54 (1.76)</td>
<td>.84 (1.12)</td>
</tr>
<tr>
<td>Charlie</td>
<td></td>
<td>2.89 (1.10)</td>
<td>2.69 (1.11)</td>
<td>2.56 (1.12)</td>
<td></td>
</tr>
<tr>
<td>(n = 181)</td>
<td></td>
<td>17.74 (8.82)</td>
<td>19.90 (8.35)</td>
<td>20.78 (8.86)</td>
<td>-1.62 (1.55)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-2.00 (1.36)</td>
<td>-1.72 (1.53)</td>
<td>-1.62 (1.55)</td>
<td>.93 (1.04)</td>
</tr>
</tbody>
</table>

Table 5 Means (and Standard Deviations) of Outcome Measures for Profile 2 at Pre-test, Post-test and Follow-up.

<table>
<thead>
<tr>
<th></th>
<th>n = 631</th>
<th>Externalizing behavior</th>
<th>Self-esteem</th>
<th>Social Preference</th>
<th>Social Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Pre-test</td>
<td>Post-test</td>
<td>Follow-up</td>
<td></td>
</tr>
<tr>
<td>Comet</td>
<td></td>
<td>1.42 (.59)</td>
<td>1.42 (.62)</td>
<td>1.40 (.58)</td>
<td></td>
</tr>
<tr>
<td>(n = 357)</td>
<td></td>
<td>23.20 (6.14)</td>
<td>24.27 (6.77)</td>
<td>24.64 (6.63)</td>
<td>1.41 (1.02)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.41 (1.02)</td>
<td>1.01 (1.30)</td>
<td>.88 (1.32)</td>
<td>.74 (.71)</td>
</tr>
<tr>
<td>Charlie</td>
<td></td>
<td>1.41 (.57)</td>
<td>1.41 (.65)</td>
<td>1.37 (.61)</td>
<td></td>
</tr>
<tr>
<td>(n = 274)</td>
<td></td>
<td>22.88 (5.78)</td>
<td>23.01 (6.97)</td>
<td>23.83 (6.79)</td>
<td>1.46 (1.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.46 (1.01)</td>
<td>.90 (1.24)</td>
<td>.87 (1.32)</td>
<td>.77 (.73)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.37 (.61)</td>
<td>23.83 (6.79)</td>
<td></td>
<td>.87 (1.32)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.26 (.96)</td>
</tr>
</tbody>
</table>
### Table 6 Means (and Standard Deviations) of Outcome Measures for Profile 3 at Pre-test, Post-test and Follow-up.

<table>
<thead>
<tr>
<th></th>
<th>Comet (n = 373)</th>
<th></th>
<th>Charlie (n = 244)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre-test</td>
<td>Post-test</td>
<td>Follow-up</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n = 617</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Externalizing behavior</td>
<td>Self-esteem</td>
<td>Social Preference</td>
</tr>
<tr>
<td>Comet (n = 373)</td>
<td>1.63 (.68)</td>
<td>4.31 (6.19)</td>
<td>.21 (1.10)</td>
</tr>
<tr>
<td>Comet (n = 373)</td>
<td>1.51 (.66)</td>
<td>17.26 (10.67)</td>
<td>.20 (1.35)</td>
</tr>
<tr>
<td>Comet (n = 373)</td>
<td>1.54 (.70)</td>
<td>19.80 (9.78)</td>
<td>.13 (1.43)</td>
</tr>
<tr>
<td>Charlie (n = 244)</td>
<td>1.72 (.77)</td>
<td>4.61 (6.96)</td>
<td>.14 (1.21)</td>
</tr>
<tr>
<td>Charlie (n = 244)</td>
<td>1.68 (.80)</td>
<td>15.92 (11.85)</td>
<td>.14 (1.41)</td>
</tr>
<tr>
<td>Charlie (n = 244)</td>
<td>1.61 (.81)</td>
<td>17.25 (11.60)</td>
<td>.06 (1.44)</td>
</tr>
</tbody>
</table>

### Table 7 Means (and Standard Deviations) of Outcome Measures for Profile 4 at Pre-test, Post-test and Follow-up.

<table>
<thead>
<tr>
<th></th>
<th>Comet (n = 349)</th>
<th></th>
<th>Charlie (n = 296)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre-test</td>
<td>Post-test</td>
<td>Follow-up</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n = 645</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Externalizing behavior</td>
<td>Self-esteem</td>
<td>Social Preference</td>
</tr>
<tr>
<td>Comet (n = 349)</td>
<td>1.21 (.34)</td>
<td>24.69 (4.75)</td>
<td>-.14 (.85)</td>
</tr>
<tr>
<td>Comet (n = 349)</td>
<td>1.29 (.51)</td>
<td>24.23 (6.22)</td>
<td>-.02 (1.15)</td>
</tr>
<tr>
<td>Comet (n = 349)</td>
<td>1.30 (.53)</td>
<td>24.27 (6.87)</td>
<td>.07 (1.21)</td>
</tr>
<tr>
<td>Charlie (n = 296)</td>
<td>1.22 (.38)</td>
<td>25.13 (4.35)</td>
<td>-.10 (.83)</td>
</tr>
<tr>
<td>Charlie (n = 296)</td>
<td>1.23 (.46)</td>
<td>24.77 (6.01)</td>
<td>.11 (1.33)</td>
</tr>
<tr>
<td>Charlie (n = 296)</td>
<td>1.22 (.49)</td>
<td>24.97 (6.91)</td>
<td>.10 (1.37)</td>
</tr>
</tbody>
</table>
Effects of time and intervention in Profile 1 ("Externalizing-Rejected").

A significant interaction effect of assessment point and intervention was recorded for social impact (F [1.89, 787.96] = 5.00, \( p = .01 \)), with a larger decrease attributed to Project Charlie than Comet. The effect size at follow-up was in the small range (\( d = .29 \)). A trend was also found for the interaction effect of assessment point and intervention on externalizing behavior, represented by a slightly larger decrease in children assigned to Comet (F [1.89, 787.43] = 3.24, \( p = .04 \)). The effect size at follow-up was small (\( d = .18 \)). Analysis of within-subjects effects showed that assessment point had a significant main effect on all outcome measures (see Figure 3). An overall decrease in externalizing behavior was demonstrated (F [1.89, 787.43] = 55.49, \( p < .001 \)), which was shown by Post-Hoc test to be significant at both post-test and follow-up. There was a significant overall increase in ratings of self-esteem (F [1.78, 743.88] = 27.39, \( p = .001 \)). Post-Hoc tests showed that the difference was significant through all assessment points. Social preference increased over time (F [1.92, 802.45] = 19.55, \( p < .001 \)), but only significantly between pre-test and post-test, whereas social impact decreased significantly through all three assessment points (F [1.89, 787.96] = 26.04, \( p < .001 \)). Effect sizes of the differences over time were in the small to medium range.

![Figure 3](image.png)

**Figure 3** Estimated marginal means of Comet and Project Charlie in Profile 1 (\( n = 419 \)), plotted by outcome measure.
Effects of time and intervention in Profile 2 ("Popular"). No significant main or interaction effects regarding Comet were recorded in Profile 2. In this profile, self-esteem, social preference and social impact changed significantly over time. Effects recorded for Profile 2 are displayed in Figure 4. Self-esteem increased (F [1.93, 1214.97] = 10.56, \( p < .001 \)), while social preference and social impact decreased (F [1.91, 1202.14] = 73.82, \( p < .001 \); F [1.97, 1238.45] = 99.54, \( p < .001 \)). The increase in self-esteem scores was significant across all assessment points, but the effect size fell short of the benchmark for ‘small’. Differences regarding social preference and social impact were significant between pre-test and post-test, but not between post-test and follow-up. Effect sizes were in the medium to large range.

**Figure 4** Estimated marginal means of Comet and Project Charlie in Profile 2 (n = 631), plotted by outcome measure.
Effects of time and intervention in Profile 3 (“Insecure”)

The analysis showed a significant interaction effect of assessment point and intervention on self-esteem (F [1.70, 1041.08] = 4.92, p = .01), illustrated by a greater increase accounted for by Comet than Project Charlie. The effect size was in the small to medium range (d = .43). Significant main effects of assessment point on externalizing behavior, self-esteem and social impact were recorded. Effects recorded for Profile 3 are displayed in Figure 5. Ratings of externalizing behavior decreased (F [1.86, 1144.76] = 9.70, p < .001) – significantly between pre-test and post-test but not between post-test and follow-up. However, the effect size fell short of the benchmark for ‘small’. Self-esteem scores increased significantly between each assessment point (F [1.70, 1041.08] = 564.90, p < .001), by a large effect size. Social impact increased as well (F [1.96, 1202.52] = 12.35, p < .001), but not significantly between post-test and follow-up. Also, the effect size was less than small.

Figure 5 Estimated marginal means of Comet and Project Charlie in Profile 3 (n = 617), plotted by outcome measure.
Effects of time and intervention in Profile 4 (“Quiet-Cool”)

Assessment point and intervention accounted for a trending interaction effect on externalizing behavior, with a larger increase in children assigned to Comet (F [1.91, 1228.71] = 3.93, p = .02). The effect size was small (d = .26). Significant main effects of assessment point on externalizing behavior, social preference and social impact were found in this group. Effects recorded for Profile 4 are displayed in Figure 6. An overall increase was observed in externalizing behavior (F [1.91, 1228.71] = 5.06, p < .001). Post-Hoc tests showed that this increase was significant at post-test, but not between post-test and follow-up. However, the effect size was below ‘small’. Significant overall increases were also recorded for social preference (F [1.93, 1236.33] = 11.64, p = .001) and social impact (F [1.99, 1275.03] = 108.31, p < .001), but the differences between post-test and follow-up were not significant. The effect size was large in the case of social impact, but very small in the case of social preference.

Figure 6 Estimated marginal means of Comet and Project Charlie in Profile 4 (n = 645), plotted by outcome measure.
DISCUSSION

The purpose of the current study was to explore the potential class-level effects of Comet in comparison to an active control condition (Project Charlie) on externalizing behavior, self-esteem and peer relations in a large sample of primary school children. Exploration of class-level effects was further elaborated by distinguishing different profiles of children, in order to investigate whether potential class-level effects of Comet varied between profiles. In addition, general developmental trajectories in different profiles were observed in order to contribute further to the understanding of relationships between externalizing behavior, self-esteem and peer relations in primary school children.

This section will start off with a discussion of the profiles obtained through the exploratory cluster analysis, followed by a discussion of the effects recorded in the analyses of variance. Next, limitations due to the choice of method and outcome measures will be presented and in closing, concluding remarks and recommendations for future research.

Profiles

The cluster analysis revealed that children in this large sample could be categorized into four different profiles. The profiles were named “Externalizing-Rejected”, “Popular”, “Insecure” and “Quiet-Cool”. When compared to categories of peer status as described by Coie, Dodge and Copotelli (1982) – that is, popular, rejected, neglected, controversial and average peer status – some noteworthy similarities are observed. Children in Profile 1 (“Externalizing-Rejected”) share the characteristics of children labeled as rejected by Coie, Dodge and Copotelli (1982), with regard to low social preference and high levels of externalizing behavior (Newcomb, Bukowski & Pattee, 1993), although “Externalizing-Rejected” children’s levels of social impact were higher than what would be considered as typical for rejected children. Children in Profile 2 (“Popular”) resemble the popular peer status group, with high scores on social preference and relatively low levels of externalizing behavior. However, based on evidence of positive correlations between peer-relations and self-esteem (Salmivalli & Isaacs, 2005; Thomaes et al., 2010), children in Profile 2 would have been expected to report higher levels of self-esteem. A possible reason for this discrepancy could be the fact that ITIA-scores were overall negatively skewed, which indicate that scores of self-esteem in Profile 2 (approximately 0.5 SD above average) in fact quite well would resemble what is considered as high self-esteem in previous studies. According to Forster (2013), reports of self-esteem have been observed to become increasingly negatively skewed during recent decades, that is, more and more children report high levels of self-esteem. The scores of self-esteem in Profile 2 could therefore be considered to reflect high self-esteem.

The fact that children in Profile 1 and 2 received almost equally high scores on social impact can be viewed as an indication of externalizing behavior and pro-social behavior being equally determining of the visibility of children in this sample. In comparison to the rejected peer status group that score average on social impact according to Coie, Dodge and Copotelli (1982), children in this sample that demonstrate externalizing
behavior seem to receive more attention from peers. However, the image of disruptive children receiving a lot of attention from peers does not at all seem far-fetched. Moreover, social impact scores of both Profile 1 and 2 were still within one standard deviation above average, making direct comparisons to previous research somewhat precarious.

Children in Profile 3 (“Insecure”) resemble that which Coie, Dodge and Coppotelli (1982) refer to as the average peer status group, demonstrating average social impact and social preference. “Insecure” children were rated average on externalizing behavior and reported the lowest ratings of self-esteem in the entire sample, exceeding one standard deviation below average. The demonstrated relation between average peer status and low self-esteem is somewhat unexpected, based on evidence of low self-esteem being associated with peer-rejection (Salmivalli & Isaacs, 2005; Thomaes et al., 2010), which implies that children in Profile 1 (“Externalizing-Rejected”) would have been more likely to report such low self-esteem. Instead, self-esteem in Profile 1 was in the average range. It is possible that the low self-esteem displayed by children in Profile 3 relates to some psychological or social occurrence other than those captured by the outcome measures in this study. Another possible interpretation is that peer status is, in fact, not that important in relation to self-esteem, as suggested by Baumeister, Campbell, Krueger and Vohs (2003) and by Bishop and Inderbitzen (1995). Moreover, measures in the present study do not take into account the presence or absence of at least one reciprocal relationship, which was found to be the most critical predictor of self-esteem in the study by Bishop and Inderbitzen (ibid.).

The perceived contradiction of children in Profile 1 being subject of peer-rejection while reporting average self-esteem could be partly clarified by taking into account that 1st graders were overrepresented in this group. In children of this age, cognitive skills required for self-evaluation are still developing, and as noted earlier, pre-school children exhibiting externalizing behavior tend to demonstrate inflated self-esteem (Mathias, Beibl, & Dilalla, 2011; Thomaes et al., 2010). Reports of average self-esteem among “Externalizing-Rejected” children in this sample might thereby be interpreted as an artefact of the many 1st graders represented in the profile; 1st graders that are not yet fully able to assess and evaluate the meaning of being disliked by peers. Likewise, the observation of children in Profile 3 (“Insecure”) scoring low on self-esteem could perhaps be attributed partly to the fact that 3rd grade children were overrepresented in the group, since evidence suggests that children’s self-evaluations become more realistic as cognitive skills develop over the years (ibid.). However, because 3rd graders only represented 100 out of 617 children in the “Insecure” profile, and 1st graders only represented 263 out of 631 children in the “Popular” profile, these interpretations are presented by a cautionary note awaiting further research.

Children in Profile 4 (“Quiet-Cool”) share the same characteristics as those reported for the neglected peer status group by Coie, Dodge and Coppotelli (1982), with low scores on social impact and average scores on social preference. They also demonstrated the lowest ratings of externalizing behavior. Somewhat unexpectedly, these children reported higher levels of self-esteem than children in any other profile. This might be interpreted as a confirmation of the notion expressed by Newcomb, Bukowski and Pattee, of the critical differences between rejected and neglected children (1993).
Rejected children are actively deselected by their peers because of their aggressive behavior and social skills deficits, which has been seen to impede self-esteem and social development. Neglected children, on the other hand, do not seem to suffer from being positioned outside of the peer group, but instead seem rather content about their situation (ibid.). Recent research highlights the heterogeneity of children within peer-status groups, e.g., distinguishing between aggressive-rejected and nonaggressive-rejected children (Bierman, Smoot, & Aumiller, 1993), and between neglected children with and without at least one reciprocal friendship (Harrist et al., 1997). Such a distinction could not be made on basis of scores of social preference and social impact. It is possible that use of other measurement instruments that take the aspect of heterogeneity into account would have produced different results.

Considering the exploratory setting for the cluster analysis and the fact that measures of externalizing behavior and self-esteem were included in the analysis besides measures of social preference and social impact, it is interesting to see that the aggregated profiles to such a great extent matched peer status groups described by Coie, Dodge and Coppotelli (1982) and by Newcomb, Bukowski and Pattee (1993). Nevertheless, these authors performed peer status group classification by using cut-off limits concerning social preference and social impact of one standard deviation below or above average, while mean scores used for distinguishing profiles in the present study were mostly within the average range, making direct comparisons to previous research somewhat precarious.

Boys were overrepresented in the “Externalizing-Rejected” group, whereas girls were overrepresented in the “Quiet-Cool” profile. The predominance of boys among “Externalizing-Rejected” children could perhaps be seen as an expression of gender differences in externalizing behavior, with boys being more prone to engage in this kind of behavior (Chaplin & Aldao, 2013). Chaplin and Aldao (2013) investigated gender differences in emotion expression in children and found that boys used externalizing behavior as a means of emotional expression more often than girls. This pattern was found to emerge in children during pre-school, which suggests externalizing behavior may not only be a manifestation of innate biological differences, but also the result of adaptation to social expectations. Higher arousal levels, lower inhibitory control and inferior language abilities are considered biologically-based conditions attributed to boys from birth, which contribute to male difficulties with regulating and expressing emotion (ibid.). By the same notion, more developed female inhibitory control and language skills help girls manage emotions constructively from an early age. Social norms and expectations are thought to exacerbate these gender differences as children reach pre- and primary school when the development of cognitive schemas that guide gender-appropriate behavior take place. Chaplin and Aldao (2013) support this idea with evidence of the expectancy that boys express more negative emotions and utilize externalizing behavior to a greater extent than girls, who in turn, are expected to express more positive emotion and internalized negative emotion in terms of guilt and anxiety.

Although beyond the scope of this study, the question of gender and gender bias in relation to externalizing behavior, peer status and self-esteem, is highly relevant. A possible presence of gender bias in Swedish primary schools is problematic because it carries a risk of narrowing behavior repertoires of children at school, possibly impeding
socio-emotional development during childhood. Chaplin and Aldao (ibid.) argue the importance of raising awareness on this matter, since previous research suggests that limiting the range of emotions expressed or encouraging expression of certain emotions to the exclusion of others increases the risk of social maladjustment and psychopathology later in life. In the case of children in the “Externalizing” group, characterized not only by externalizing behavior but of low social preference scores as well, opportunities of engagement in social interaction are likely to be few. As stated by Newcomb, Bukowski and Pattee (1993), the combination of externalizing behavior and peer-rejection tend to obstruct chances of developing positive social interaction with peers. Assuming that gender bias contributes to the reinforcement of externalizing behavior in boys within this group, a change of attitudes might help them to engage less in behaviors that continuously serve alienation from peers, while at the same time developing more pro-social behaviors. Along the same line, reducing eventual gender bias might as well lead to girls developing more assertive behaviors.

**Effects of Comet and time across profiles**

Results from the analyses of variance show that variations in outcome measures differed significantly between profiles, while intervention had little detectable impact on differences within each profile. Generally, the overall pattern of development across profiles was movement and stabilization toward overall mean scores.

With respect to previous research on relations between externalizing behavior, self-esteem and peer relations, the most predictable development found in this sample is the one displayed by children in Profile 1. In this profile, significant overall decreases in externalizing behavior and social impact coincided with an increase in self-esteem and social preference. Although research on self-esteem has demonstrated contradictory results, some of the research (e.g., Salmivalli & Isaacs, 2005; Thomaes et al., 2010) indicate that parallel increases of self-esteem and social preference are to be expected, which is the case for the “Externalizing-Rejected” children. Intervention and time generated interaction effects in two cases: Comet accounted for a smaller decrease in social impact than Project Charlie, and for a trend of being more effective in reducing externalizing behavior. In these cases, however, effect sizes were small. Thus, differences in Profile 1 could be interpreted as manifestations of a general pattern of progression in children displaying these profile-specific characteristics at the ages of 7-9. On the other hand, recorded intervention effects could also be considered meaningful, given that small effect sizes should be interpreted somewhat differently in the case of preventive interventions than with indicated interventions, because the level of target behaviors are closer to average to begin with. Simply put – if an intervention is to have an effect, the level of the target behavior or psychological construct in question must originate from a notable level (Karlberg, 2011).

Hence, the result in Profile 1 might indicate that the effect of Comet on externalizing behavior would have been significant if these children had demonstrated higher ratings at pre-test. In fact, this is indeed the case when looking at the results from the RCT-study by Forster et al. (2012), which evinced the efficacy of Comet compared to Project Charlie with regard to target students. With regard to the sample in the current study, were target students were bundled together with large numbers of non-target students,
effects are naturally smaller. That being said, reduction of externalizing behavior at a universal level of an effect size at $d = .18$ is likely to be both noticeable and relevant in the school setting. This can be compared to results presented in the meta-analysis of school-based interventions by Wilson and Lipsey (2007), where universal level effect sizes of around $d = .20$ are considered to be of practical significance.

Accordingly, the results in Profile 1 indicate that Comet could indeed be effective in reducing externalizing behavior at a non-indicated level. Usually, externalizing behavior tends to emerge in conjunction to the transition from pre-school to primary school, as increasing demands become hard to meet for some children (Karlberg, 2011). Profile 1 demonstrated an opposite pattern over time, which could be interpreted as an expression of successful adaptation to the primary school form. In this case, there is further a possibility of Comet improving this adaptation process. Nevertheless, these inferences should be considered with caution, given that the effect is non-significant ($p = .04$) by the use of Bonferroni-correction. Furthermore, in line with previous research on the connection between externalizing behavior and peer relations (Coie, Dodge, & Coppotelli, 1982; Karlberg, 2011; Newcomb, Bukowski, & Pattee, 1993), children in Profile 1 seemed to benefit from lower levels of externalizing behavior, in terms of improved relations to peers, as expressed by somewhat higher ratings of social preference.

In Profile 2, no significant effects concerning intervention were recorded. These “Popular” children displayed a trajectory of stable levels of externalizing behavior, a small increase in scores on self-esteem and a medium to large decrease in levels of social preference and social impact. Although becoming less popular among peers than before, this does not seem to affect the self-esteem of these children. According to Thomaes et al. (2010), children by this age should have acquired the cognitive ability to make realistic self-evaluations. Therefore, at the time they reach follow-up, the “Popular” children should reasonably be better at evaluating the fact that they are less liked by peers. On the contrary, other research evidence suggests that there is in fact no relation between peer status and self-esteem (Baumeister, Kruger, & Vohs, 2003; Bishop & Inderbitzen, 1995). Yet others highlight the possibility of high self-esteem reports reflecting defensive self-esteem rather than genuine self-esteem (Schneider & Turkat, 1975).

Forster (2013) point out that while high ratings of self-esteem have increased during recent decades, rates of psychological well-being have developed in the opposite direction during the same time period. This has led to some researchers drawing the conclusion that optimism in self-esteem reports is an expression of self-enhancement rather than genuine self-evaluation (ibid.). The inconsistent pattern of self-esteem in relation to peer status in Profile 2 might therefore be interpreted as an expression of biased rather than genuine self-esteem. Another interpretation is that the increase in self-esteem scores is related to some other factor not measured in the current study, such as social support outside the school setting, which can serve as a protective factor for stress caused by peer-rejection (Newcomb, Bukowski, & Pattee, 1993). One might also point out that these children’s scores on peer relations, externalizing behavior and self-esteem are still within the normal range, hence, it is arguable that variations are in fact too small to generate any conclusions.
The most prominent variation in Profile 3 ("Insecure") was attributed to increases in self-esteem over time, while changes in the other variables were negligible. Results show that children assigned to Comet displayed a larger increase in self-esteem than children in Project Charlie, and the effect size was in the small to medium range. Again, taking into account that this is at universal level, this effect size can be considered to represent an effect of practical significance. The increase in self-esteem scores was substantial – from exceeding one standard deviation below average at pre-test, to within a half standard deviation below the mean at follow-up. It is interesting that Comet represented a larger increase in self-esteem scores, since increasing self-esteem is an explicit goal of Project Charlie but not of Comet. Perhaps, certain changes of teacher behavior encouraged in Comet could contribute to positive self-evaluation in students with low self-esteem, such as teachers using less reprimand and instead reinforcing the development of social skills in students. However, this is merely a speculation and further research is needed to establish the course through which Comet can support the development of self-esteem in "Insecure" children.

The development of self-esteem for children in Profile 3 could also be viewed in light of evidence presented by Mathias, Beibl and Dilalla (2011), suggesting that self-esteem stabilizes as children enter primary school. Thus, a possible hypothesis would be that cognitive development in these children fostered acquisition of a more realistic self-esteem, which was in fact higher than the self-esteem demonstrated at pre-test. On the other hand, research shows that self-esteem in pre-school children tends to be positively biased rather than negatively biased (ibid.). It is possible that the development of self-esteem displayed by children in Profile 3 relates to some psychological or social occurrence beyond the scope of outcome measures in this study, which would also be likely to influence the positive effect of Comet.

At pre-test, children in Profile 4 ("Quiet-Cool") demonstrated the lowest level of externalizing behavior and social impact in the sample, combined with the highest level of self-esteem. At follow-up, externalizing behavior, social preference and social impact all increased, while self-esteem stayed stable. There was a trend of intervention effect of Comet being related to a larger increase in externalizing behavior than Project Charlie ($p = .02$). The effect size of $d = 0.26$ should be considered of practical significance in this context. In relation to the overall increase in social impact, added levels of externalizing behavior could perhaps be viewed in relation to increased visibility in the peer group. Given the low levels of externalizing behavior in Profile 4 at pre-test, increased visibility is considered as a positive development. The boost in social impact is also manifested in slightly increased social preference scores, which implies that the increase of externalizing behavior might have served these children well, in the sense of commanding attention and space for themselves. The results imply that Comet could potentially foster this process.

Limitations of the study
In this study, effects of Comet were compared to Project Charlie – an active control group in terms of teachers in this group receiving credible attention, as opposed to having a no-treatment control group (Forster et al., 2012). Project Charlie is indeed an “active” intervention, although differing from Comet in terms of purpose and content.
While Project Charlie was originally chosen as control group because it was not expected to be effective in reducing externalizing behavior (ibid.), the conditions under which Project Charlie functions as control group in this study might differ, taking into account the addition of self-esteem and peer relations as outcome measures. Previous evidence only supports effects of Comet with regard to reducing externalizing behavior. Project Charlie, on the other hand, explicitly aims at boosting self-esteem and improving peer-relations. Under these terms, Project Charlie could in fact have been expected to be more effective than Comet in some aspects. An advantage of comparing Comet to Project Charlie is that the observation of Comet being more effective in boosting self-esteem for children with the “Quiet-Cool” profile can be considered even more reliable, than if Comet had been compared to an intervention not aimed at boosting self-esteem. Along the same line, a potential disadvantage of using Project Charlie as comparison group is increased risk of not detecting or underestimating effects of Comet with regard to self-esteem and peer relations, especially when considering the exploratory setting for the current study.

The attrition at pre-test, which affected the cluster analysis, as well as attrition of approximately 12 % at post-test and 25 % at follow-up, should be taken into account as a limitation of the study. Although attrition was mainly considered to be random, there is indeed an increased risk of Type II-error associated with attrition. Apart from random attrition, certain observations of systematic attrition were made. The first observation is concerned with the overrepresentation of 2nd grade children in the attrition at pre-test. Although this might be explained in part by the fact that 2nd grade children were overrepresented in the study to begin with, it cannot be excluded that this had some effect on the cluster analysis. The second observation is the recurrence of missing data at only one of the following outcome measures: externalizing behavior at post-test, self-esteem at post-test, externalizing behavior at follow-up, or self-esteem at follow-up. Since measures of externalizing behavior and self-esteem were collected by teachers, this might have been a reflection of teachers experiencing time pressure or other work-related issues, thereby affecting the response rate. The work situation for teachers in Sweden is generally considered to be strained due to, for example, decreased teacher-student ratio and increasing demands of written documentation. It is possible that the teachers responsible for the missing data experienced an even morepressuring work situation than teachers completing all evaluations, possibly biasing the results in some way. Hence, future research might benefit from including measures other than teacher reports, alternatively relying on research assistants to assist the collection of teacher ratings.

With regard to the risk of Type II-error, something can also be said concerning Bonferroni-adjustment, which was applied in this study. Bonferroni-adjustment is commonly used in psychological research to reduce the inflated risk of Type I-error associated with performing multiple analyses on a single dataset (Eichstaedt, Kovatch & Aaron Maroof, 2013). However, there is some controversy as to whether the conservative alpha-level produced by Bonferroni-adjustment makes it appropriate for use in some cases, e.g., when conducting exploratory analyses. This is because a significant reduction of the risk for Type I-error simultaneously lowers power and increases the risk for Type II-error (ibid). Thus, a possible limitation of this study is the choice of a conservative criterion regarding alpha-level, which may have led to false
conclusions of some effects being non-significant. After all, being able to detect effects of an indicated intervention at a universal and preventive level is likely to be relatively difficult to begin with.

Since the efficiency of Comet is mainly attributed to changes in teacher behavior (Forster et al., 2012), which assumingly affects all students in a class, the question as to whether the program can benefit students at group level seem reasonable. Without application of Bonferroni-adjustment, as argued by Eichstaedt, Kovatch and Aaron Maroof (2013), the effects of Comet on reducing externalizing behavior in Profile 1 and on increasing externalizing behavior in Profile 4 would indeed have been considered significant. This would have made it possible to demonstrate group-level effects of Comet with more certainty. Regardless of whether the adjustment should have been applied or not, results produced within the exploratory setting for this study should be considered as ground work for future research.

Moreover, teachers assigned to Comet were free to choose class-level interventions for implementation at their own liking. Approximately two thirds of the teachers implemented formulation of class rules and about three quarters chose to implement token economy. Less than half of the teachers implemented the Class-Wide Peer Tutoring-intervention. The fact that the program was not implemented consistently throughout the entire sample is a limitation of the current study. Also, intervention adherence was not controlled for. Thus, future research on group-level effects of Comet could benefit from a stricter implementation of the program, which would allow for investigation of how different elements of the program contribute to its effects.

Furthermore, two limitations can be pointed out with regard to the data analysis. First, choosing different combinations of variables in the cluster analysis would have been likely to generate alternative cluster solutions, all of which would probably be interesting to investigate further in some aspect. For example, inclusion of measures of performance in mathematics and word comprehension, which were available in the data set, could have generated alternative profiles. Including other variables in the cluster analysis could also have altered the somewhat problematic finding of bivariate correlations between externalizing behavior and social preference, and between externalizing behavior and social impact. Although these conditions can be considered as reasonable or even expected, given the evidence of relatedness between externalizing behavior and peer relations as presented by Coie, Bukowski and Pattee (1993), using non-related variables would probably have generated a more robust cluster analysis.

Second, data failed to meet the demand of sphericity for the analyses of variance, which had to be corrected by using a more conservative criterion for interpretation of observed values and significance level. Restriction of range of chosen variables might have contributed to this issue. In all measures except for ITIA, raw scores ranged only within one to five. In addition, data of externalizing behavior and self-esteem were moderately skewed, which means that a majority of data points accumulated within an even smaller range. The issue of sphericity could have been avoided by choosing a different method for the follow-up analyses, e.g., paired samples t-test. Performing t-tests instead of MANOVAs would however command choosing a more conservative alpha-level using Bonferroni-adjustment, since the number of t-test needed to cover all follow-up analyses
would exceed the number of performed MANOVAs. Moderator analysis would have been another possible choice of data-analysis, which would have allowed for investigation of all effects of interest in one analysis at the same time, using profile as a moderating factor instead of independent variable. However, MANOVA was chosen as the preferred method of analysis because it meant performing fewer analyses than by choosing t-test, and because it was considered more reasonable than Moderator analysis, given the authors’ prior knowledge and experience of statistical analyses.

Furthermore, there is some controversy as of how to best interpret effect sizes. Guidelines such as those presented by Cohen (1988) are general and not always suitable for application. According to Peng, Chen, Chiang & Chiang (2013), effect sizes should be interpreted by considering the research design, characteristics of independent and dependent variables, as well as sample characteristics in the study at hand. Because observed variations were generated on the base of data aggregated by an explorative analysis, previous research could not be used as reference for interpreting effect sizes. For lack of other plausible references, Cohen’s guidelines for small, medium and large effect sizes were considered to be acceptable for use in this study, although interpretation was also made by the notion of differences between expected effect sizes regarding effects at indicated and universal level (Wilson & Lipsey, 2007). In the case of effect sizes reported in eta-square, which is known to be a somewhat upwardly biased measure of effect size with regard to the population in question (Pierce, Block, & Aguinis, 2004), these should be interpreted with this reservation in mind.

The observation of children assigned to Comet demonstrating a significantly lower overall mean on ITIA than children in Project Charlie at pre-test could not be traced in ratings of self-esteem over time. Although significant, the difference measured only 0.10 in z-score at pre-test. However, the lack of equivalence may have hampered detection of potential differences in scores of self-esteem between the two groups over time.

On account of observed resemblance of the four profiles to peer status groups as described by Coie, Dodge and Coppotelli (1982), a reservation must be made with regard to limited variances of social preference and social impact scores. As noted earlier, the original peer status classification was performed on basis of cut-off limits of one standard deviation below or above average, whereas scores used for distinguishing profiles in the present study were mostly within the average range. Thereby, comparison of profiles to previous research and interpretations of developmental trajectories should be made with caution.

Limitations can also be discussed with regard to outcome measures. Although demonstrating adequate reliability, the measure of externalizing behavior, BREB, is a short teacher-report of only three items. First of all, restriction of range in the scale may have contributed to the problem of sphericity in the analyses of variance. Second of all, blind observer ratings would have been preferable to teacher ratings, in order to reduce the risk of bias. Moreover, recent research distinguishes between different types of externalizing behavior that might have different implications for peer relations (Gifford-Smith & Brownell, 2003). This is also related to how different types of externalizing behavior are perceived across developmental stages in childhood. Three types of
externalizing behaviors – or aggression, which is the term used by the authors – are defined: instrumental aggression (aggression to obtain a personal gain of some kind), reactive aggression (aggression in response to provocation), and bullying (unprovoked aggression in relation to a specific person). Among younger children, instrumental aggression is the only type related to peer-rejection, whereas all types of aggression are related to peer-rejection among older children (ibid.). Gifford-Smith and Brownell (ibid.) further highlight the importance of studying externalizing behavior in relation to what kind of aggression is considered normative in a given peer group. Evidence suggests that only non-normative types of aggression are related to peer-rejection, which means that different types of externalizing behavior can cause different responses depending on peer group. Furthermore, research on externalizing behavior has often focused on boys, which has caused a knowledge gap concerning externalizing behavior exhibited by girls and its relatedness to peer relations. By using BREB as outcome measure, the different considerations with regard to externalizing behavior as described by Gifford-Smith and Brownell (ibid.) has not been successfully captured, thus indicating a limitations of the current study.

As for the measure of self-esteem, ITIA, negatively skewed reports not only posed a problem of restricted range, but also raise questions of the scale at a more general level. As mentioned before, the increased number of observations of negatively skewed self-esteem reports during recent years highlights the uncertainty regarding reliability and validity of self-esteem as a construct (Forster, 2013). Self-esteem measures have also been questioned with regard to differences between global and domain-specific self-esteem (Baumeister, Kruger, & Vohs, 2003). The use of global self-esteem measures is considered to be somewhat problematic, since global reports often fail to predict specific behaviors or outcomes. In this study, being able to compare the results to previous research was the main reason for choosing measures of global self-esteem instead of domain-specific self-esteem. The question as to exactly what a measure such as ITIA actually reflects is nevertheless highly relevant.

Using peer-status as a measure of peer-relations can also be considered a limitation of the present study. In recent research, the importance of other dimensions of peer relations beside peer-status is discussed (Gifford-Smith & Brownell, 2003). Two aspects of peer relations that have been in the interest of researchers because of their associations with psychological well-being and social maladjustment in children are social network centrality and number of mutual friendships (ibid.). In a study by Gest, Graham-Bermann, and Hartup (2001), peer-status was found to correlate only moderately with social network centrality and number of mutual friendships. Moreover, 39 % of rejected children in the study had at least one mutual friendship, while 31 % of popular children did not. Having at least one reciprocal friendship has been found to buffer for ill-effects of being rejected at group-level (Bishop & Inderbitzen, 1995; Hodges, Boivin, Vitaro, & Bukowski, 1999). Thus, using peer-status as the sole measure of peer relations is a limitation of the current study, inasmuch as other relevant aspects of peer relations are not taken into account.
**Conclusions**

The study shows that Comet might benefit children beyond expected effects with regard to problematic levels of externalizing behavior in targeted students. For children with low self-esteem, the results show that Comet has a positive effect on the development of a higher, and possibly more realistic, self-esteem. Although not reaching the required level of significance, Comet showed some effect on reducing externalizing behavior in Profile 1 (“Externalizing-Rejected”), which could be viewed as an indication of Comet possibly being relevant for the purpose of preventing problematic behavior at universal level in primary schools. Comet also seems to be effective in addressing externalizing behavior from the other end of the spectrum – that is, by having positive effect on children exhibiting very low levels of self-esteem to command more attention and space for themselves in a healthy way. However, this effect was not significant.

Moreover, the four profiles generated by the cluster analysis is an interesting finding. The profiles resemble peer status groups as described in previous research, thus lending some support to such a categorization. Analyses showed that developmental trajectories were different across profiles, with a common pattern of movement and stabilization toward the overall mean scores. Some observations of relations between self-esteem, externalizing behavior and peer relations diverged from knowledge established through previous research, which hampered interpretations of how these constructs might be related to each other in developmental trajectories and observed intervention effects. For example, the parallel increase in self-esteem scores and decrease in social preference in Profile 2, and the lack of relation between self-esteem scores and peer status in Profile 1, could be considered unexpected taking into account some of the research in the area. These diversions were partly considered in the light of ongoing controversy regarding the validity and reliability of self-esteem as a construct, as well as limitations due to chosen measures of externalizing behavior and peer relations.

A major strength of the study was that the sample was large and that data had been collected through randomized controlled design, which benefits the validity and generalizability of the results. Indicated group-level effects of Comet, as well as the identified profiles, can therefore be considered representative in relation to the population of primary school children in Sweden.

**Recommendations for future research**

The inconsistencies of the findings in relation to previous research raise questions about the validity and relevance of self-esteem as a psychological construct. Perhaps a more fruitful approach than the one in this study would be investigating connections to externalizing behavior and peer status by looking at different domains of self-esteem. On the other hand, considering the difficulties and the incertitude associated with measuring a non-observable psychological construct such as self-esteem, it may be even more constructive to focus on other variables that are measurable in ways that offer greater reliability and validity. For example, behavior observations could be used to operationalize different aspects of self-esteem, instead of self-ratings. Future research should also acknowledge the need for considering various aspects of externalizing behavior and peer relations when choosing appropriate outcome measures.
Moreover, further knowledge of the effects of Comet, as well as of developmental trajectories in primary school children in general, could potentially be found by including additional outcome measures besides the ones used in this study. For example, investigating potential group-level effects of Comet on academic performance would be a relevant course of action. A plausible hypothesis behind such an investigation could be that by being effective in reducing externalizing behavior in target children, Comet might contribute to a better work environment in classrooms (Karlberg, 2011), which could possibly foster enhanced academic performance.

REFERENCES


