**Teacher-rated aggression and co-occurring problems and behaviors among schoolchildren: A comparison of four population-based European cohorts**

Whipp A[1*], Vuoksimaa E[1], Bolhuis K[2], de Zeeuw EL[3], Korhonen T[1], Mauri M[4], Pulkkinen L[5], Rimfeld K[6], Rose RJ[7], van Beijsterveldt T CEM[3], Bartels M[3,8], Plomin R[6], Tiemeier H[9], Kaprio J[1,10], Boomsma DI[3,8]

**Affiliations**

(1) Institute for Molecular Medicine Finland (FIMM), University of Helsinki; (2) Department of Child and Adolescent Psychiatry, Erasmus Medical Center, Rotterdam, The Netherlands; (3) Department of Biological Psychology, Faculty of Behavioural and Movement Sciences, Vrije Universiteit Amsterdam, The Netherlands; (4) University of Cagliari, Cagliari, Italy; (5) Department of Psychology, University of Jyvaskyla, Jyvaskyla, Finland; (6) Social, Genetic, and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology, and Neuroscience, King's College London, London, England; (7) Department of Psychological & Brain Sciences, Indiana University, Bloomington, Indiana, USA; (8) Amsterdam Public Health Research Institute, The Netherlands; (9) Department of Social and Behavioral Science, Harvard TH Chand School of Public Health, Boston, MA, USA; (10) Department of Public Health, University of Helsinki

***Corresponding Author:**

Alyce M Whipp
P.O. Box 20 (Tukholmankatu 8 B)
00014 University of Helsinki
Helsinki, Finland
+358451075040
alyce.whipp@helsinki.fi

**Acknowledgements**

This work is part of the ACTION consortium which is supported by funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 602768.

FT12: We gratefully acknowledge the ongoing contribution of the participating twin families. Data collection has been supported by the National Institute of Alcohol Abuse and Alcoholism (grants AA-12502, AA-00145, and AA-09203 to RJR) and the Academy of Finland (grants 100499, 205585, 118555, 141054 and 264146 to JK). JK has been supported by the Academy of Finland (grants 308248, 312073). Data collection in FinnTwin12 has been supported by the National Institute of Alcohol Abuse and Alcoholism

GENR: This work was supported by the Netherlands Organization for Scientific Research (NWO-grant 016.VICI.170.200) to HT, and the Sophia Children's Hospital Research Foundation (research fellowship grant 921) to KB. Super computing resources were made possible through the NWO Physical Sciences Division (surfsara.nl).

The first phase of the Generation R Study is made possible by financial support from the Erasmus Medical Centre, Rotterdam; the Erasmus University Rotterdam; The Netherlands Organization for Health Research and Development (ZonMw). The authors gratefully acknowledge the contribution of all children and parents, general practitioners, hospitals, midwives, and pharmacies involved in the Generation R Study. The Generation R Study is conducted by the Erasmus Medical Center (Rotterdam) in close collaboration with the School of Law and Faculty of Social Sciences of the Erasmus University Rotterdam; the Municipal Health Service Rotterdam area, Rotterdam; the Rotterdam Homecare Foundation, Rotterdam; and the Stichting Trombosedienst & Artsenlaboratorium Rijnmond, Rotterdam.

NTR: We gratefully acknowledge the ongoing contribution of the participating familes. Data collection in the NTR was funded by the Netherlands Organization for Science (NWO): Twin-family database for behavior genetics and genomics studies (NWO-480-04-004); Gravitation program of the Dutch Ministry of Education, Culture, and Science and the Netherlands Organization for Scientific Research (NWO-024.001.003). 'Longitudinal data collection from teachers of Dutch twins and their siblings' (NWO-481-08-011); 'Twin-family study of individual differences in school achievement' (NWO-056-32-010); ZonMW "Genetic influences on stability and change in psychopathology from childhood to young adulthood" (NWO-912-10-020); "Netherlands Twin Registry Repository" (NWO-480-15-001/674). DIB would like to acknowledge the KNAW Academy Professor Award (PAH/6635).

TEDS: We gratefully acknowledge the ongoing contribution of the participants in the Twins Early Development Study (TEDS) and their families. TEDS is supported by a program grant to RP from the UK Medical Research Council (MR/M021475/1).

No funders had any involvement in the conduct of this research, including in the study design, data collection, analysis and interpretation of the data, the writing of the report, and the decision to submit to the article for publication.

**Conflicts of Interest**

J Kaprio and T Korhonen have consulted for Pfizer on nicotine dependence (2011-2015 and 2011-2017, respectively). All other authors (E Vuoksimaa, K Bolhuis, EL de Zeeuw EL, M Mauri, L Pulkkinen, K Rimfeld, RJ Rose, T CEM van Beijsterveldt, M Bartels, R Plomin, H Tiemeier, DI Boomsma) declare no conflicts of interest.

**Abstract**

Non-pathological levels of children's co-occurring externalizing/internalizing problems in the school setting are not well characterized. We assessed associations between aggressive behavior and co-occurring externalizing/internalizing problems among schoolchildren in 39,936 teacher ratings from 4 population-based cohorts (FinnTwin12 from Finland, Generation R and the Netherlands Twin Register from the Netherlands, and the Twins Early Development Study from

the UK). Correlations of aggressive behavior were high with other externalizing problems (0.47-0.80) and lower with internalizing problems (0.02-0.39). A negative association was seen with prosocial behavior, ranging -0.33 to -0.54. Although the mean levels of aggressive behavior differed significantly by gender, the correlations of aggressive behavior with co-occurring behaviors were notably more similar between the genders (e.g., aggressive-hyperactivity correlations: 0.51-0.75 boys, 0.47-0.70 girls). Comparisons are also discussed with respect to age, instrument/cohort, and parental ratings. These population-based European observations highlight the complexity of aggressive behavior at school, and can assist in creating appropriate interventions.

**Keywords:** aggression, behavioral problems, comorbidity, teacher ratings, gender differences

## Introduction

Aggressive behavior in school is a persistent topic of discussion, with concerns surrounding bullying, gender- and sexuality-based misconduct, and extreme violence (e.g., school fights, stabbings, shootings) currently garnering the most attention (UNESCO 2019; OECD 2017). Although these specific areas of violence and aggression are important to understand and manage for the safety of all in the school environment, they are but pieces in the larger puzzle of aggressive behavior in the school setting. Indeed, aggression is a heterogenous behavior that has been characterized in a variety of ways, including well-researched subtypes such as proactive (i.e., planned aggression, such as bullying or stealing), reactive (i.e., in response to some stressor, such as a bully instigating a confrontation), direct (i.e., yelling, hitting, kicking, verbal threats) and indirect (i.e., social/relational aggression, such as spreading nasty rumors or ostracizing an individual) (Pulkkinen 2018; Vitaro et al. 2006). All children express such behaviors to varying degrees, in time and amount. This complex behavior among school-aged children is associated

with relationship quality among their peers, teachers, and families, as well as their eductional attainment, risk for substance abuse and other psychiatric disorders, and risk for partipation in criminal activities (Farrington 1995; Pulkkinen 2017; Whipp et al. 2018).

In studying aggression, gender differences have played a large and complex role in our understanding of the problem behavior. Early research focused only or mainly on boys, although studies in recent decades have greatly enriched our knowledge regarding both genders. Distinctions between direct/physical aggression as being higher among boys and indirect aggression being higher among girls have generally given way to an understanding that both genders engage in both types of aggression, and that at least for indirect aggression the gender differences are negligible (Card et al. 2008; Cleverley et al. 2012). Likewise, a study by Nivette et al. (2018) indicates that in societies that experience more gender inequality, differences in physical aggression by gender are minimal.

Complicating our understanding of this behavioral problem further, children with aggressive behavior often have co-occurring problem behaviors, either externalizing, internalizing or both (August et al. 1996, Bartels et al. 2018, McNellis et al. 2017, Verhulst et al. 1993, Willner et al. 2016). Furthermore, children with aggression and co-morbidities often have more than one co-morbidity, worse prognoses and poorer outcomes (Fontaine et al. 2008; Pulkkinen 2017; Vaalamo et al. 2002; Verhulst et al. 1993). Regarding the school setting, research has generally focused on severe aggression problems (e.g., conduct disorder, oppostional-defiant disorder, bullying), however, less is known about general, sub-diagnostic levels of aggressive behavior and co-occuring behaviors.

Aggressive behavior is often characterized and studied via parental, self, and mental health professional assessments, with teacher ratings being used less often in epidemiologic or clinical research. Research from the past several decades has shown that an individual's aggressive

behavior may be situational (e.g., present at school, but not at home) and that assessments of behavioral and emotional problems do not correlate well between raters (mean correlations between raters from Achenbach 1987 range 0.22–0.28 for different raters (such as parent and teacher) and 0.60 for similar raters (such as mother and father)) (Achenbach 1987; Verhulst 1994; De Los Reyes 2005). Furthermore, studies have shown that teacher ratings of problem behaviors have often been most useful in diagnostics and predictive outcomes, including psychiatric disorders and criminality (Stanger et al. 1993; Verhulst 1994; Carbonneau et al. 2005; Hodgins et al. 2013; Whipp et al. 2018). Teachers' ability to observe children in a complex task-based setting and among peers of similar age and abilities provides them with a valuable comparison-base, making their insight unique and important.

This study's overall objective was to characterize the levels and associations of aggressive behavior and co-occurring behaviors in the school setting, using teacher ratings of children at ages 7–14. Large datasets from population-based cohorts of children from Finland, the Netherlands, and the UK were utilized. Specific aims were: 1) to report the mean levels of aggressive and other behaviors as assessed by teachers and to examine associations of aggressive behavior and co-occurring behaviors and 2) to assess gender differences in the levels and associations of aggressive behavior and co-occurring behaviors.

## Materials and Methods

The datasets for this investigation were obtained through the collaboration of the ACTION (Aggression in Children: Unravelling gene-environment interplay to inform Treatment and InterventiON strategies; http://www.action-euproject.eu/) consortium (Bartels et al. 2018). ACTION's main goal is to improve our understanding of the causes of the individual differences in childhood aggression and co-occurring behavioral and emotional problems, with a strong focus on

genetics, epigenetics, and biomarkers, as well as an integrative component concerned with tailoring prevention and intervention strategies.

Here, we utilized data from four ACTION cohorts — from Finland, the Netherlands, and the UK —with available teacher ratings to address our study aims. In total, the data included 39,936 teacher ratings on children at ages 7, 9, 10, 12, and 14 (48.9% boys). Although cross-sectional in approach, the longitudinal nature of the cohort data provides some possible overlap of the same child at two or more age time points. As in Bartels et al. (2018), in the cohorts that include twin data one twin per family was randomly selected for analysis, and individuals were excluded if they had an illness or handicap that interfered with their daily functioning, and if they were missing substantial data from one of the examined subscales from cohort-respective behavioral and emotional problem questionnaires (see also: http://www.action-euproject.eu/content/data-protocols). All data in the current investigation were collected under protocols that have been approved by the appropriate ethics committees, and studies were performed in accordance with the ethical standards established in the 1964 Declaration of Helsinki and its later amendments.

*Cohort descriptions*

FinnTwin12 (FT12)

The FT12 dataset was established in Finland from the cohort of twins born 1983–87 (Kaprio et al. 2013). It is a population-based twin study aimed at examining health-related behaviors and their precursors. Data were collected at ages 12, 14, 17, and 22. Teacher ratings on behavioral and emotional problems were collected at ages 12 and 14 using the modified Multidimensional Peer Nomination Inventory (MPNI) (Pulkkinen et al. 1999). Response rates were 93% and 94% for age 12 and 14 teacher ratings, respectively.

Generation R (GENR)

The GENR Study is a population-based birth cohort, established in the Netherlands from children born 2002–06 in the city of Rotterdam and surrounding areas (Kooijman et al. 2016), aimed at examining growth, development and health from fetal life to young adulthood. Data were collected during pregnancy, at birth, and frequently throughout childhood. Currently, data are being collected for the mean age 13 years assessment wave. Teacher ratings on behavioral and emotional problems were collected at child mean age 7 years using the Achenbach System of Empirical-Based Assessment's (ASEBA) Teacher Report Form (TRF) (Achenbach 1991); the response rate was 77%.

Netherlands Twin Register (NTR)

The NTR was established in the Netherlands in 1987. The register includes adults and young twins who were registered by their parents shortly after birth and who are followed longitudinally to examine, in particular, behavioral development and psychopathology (Van Beijsterveldt et al., 2013). Data are collected at ages 1, 2, 3, 5, 7, 9, 10, 12, 14, 16, and 18 years (at which point twins move into the adult twin register), with tools including the ASEBA to obtain information. In school year 1999–2000, the NTR began collecting teacher ratings for 7, 10 and 12 year old twins, with the ASEBA's TRF (Achenbach 1991); the average response rate was ~60%.

Twins Early Development Study (TEDS)

The TEDS study was established in the UK from the cohort of twins born 1994–96 (Haworth et al. 2013). It is a UK-representative longitudinal twin study aimed at examining language, cognitive, and behavioral development. Data were collected at ages 2, 3, 4, 7, 8, 9, 10, 12, 14, 16, 18, and 21. Teacher ratings on behavioral and emotional problems were collected at ages 7, 9, and

12 using the Strengths and Difficulties Questionnaire (SDQ) (Goodman 1997); response rates were 85%, 76%, 78%, respectively.

*Study questionnaires*

Multidimensional Peer Nomination Inventory (MPNI)

The MPNI, used by the FT12 cohort, was originally developed as a tool for rating peers on childhood social behavior, however, it has been adapted and modified to collect ratings from parents, teachers, and the twin themself (Pulkkinen et al. 1999). It is a 37-item questionnaire with the 6 subscales used here being: aggression (6 items), depression (5 items), hyperactivity–impulsivity (7 items), inattention (4 items), prosocial behavior (12 items), and social anxiety (2 items). Although further subscale scoring was not performed, the MPNI internal structure based on factor analysis indicates aggression, hyperactivity–impulsivity and inattention subscales as part of a behavioral problem/externalizing subscale, and depression and social anxiety subscales as part of an emotional problems/internalizing subscale. Each item on the teacher rating questionnaire has four response choices: 'not observed in child', 'observed sometimes, but not consistently', 'definitely observed, but is not prominent', and 'clearly observed'. Response choices are scored 0–3, and subscales are formed by taking the mean of all items in the subscale (no missing values were allowed).

Strengths and Difficulties Questionnaire (SDQ)

The SDQ, used by the TEDS cohort, is a 25-item questionnaire that measures common childhood mental health problems (Goodman 1997). The five subscales of the SDQ are conduct problems (5 items), emotional problems (5 items), hyperactivity (5 items), peer problems (5 items), and prosocial (5 items). For this study, we use 'conduct problems' as a proxy for aggressive

behavior, and ‘emotional problems’ as a proxy for anxiety problems, as in Bartels et al. (2018). Although further subscale scoring was not performed, the SDQ does recognize conduct problems and hyperactivity subscales as part of an externalizing subscale, and emotional and peer problems subscales as part of an internalizing subscale. Each item on the questionnaire has three response choices: ‘not true’, ‘somewhat true’, and ‘certainly true’. Response choices range from 0–2 and are coded so that higher scores represent greater risk for the behavior attribute, and subscale scores are derived as a scaled mean score when at least 3 of the 5 items in a subscale are non-missing.

Teacher Report Form (TRF)

ASEBA’s TRF, used by the GENR and NTR cohorts, is a 112-item questionnaire that measures childhood behavioral and emotional problems (Achenbach 1991, 2001; Lamb 2012). The eight syndrome subscales used here are aggressive behavior (20 items), anxious/depressed (16 items), attention problems (26 items), rule-breaking behavior (12 items), social problems (11 items), somatic complaints (9 items), thought problems (1 item), and withdrawn/depressed (8 items). Although further subscale scoring was not performed, the TRF recognizes aggressive behavior and rule-breaking behavior subscales as part of an externalizing problems subscale, and anxious/depressed, withdrawn/depressed and somatic complaints subscales as part of an internalizing problems subscale. In order to retain comparability across instruments, the attention problems subscale will also be considered as externalizing for the purposes of this study. Each item on the questionnaire has three response choices: ‘not true’, ‘somewhat or sometimes true’, and ‘very true or often true’. Response choices are scored 0–2, and items are summed to created sum scores of individual subscales. GENR weighted the subscale sum scores for missing values, while NTR allowed, in accordance with the ASEBA manual, a small number of missing values per

subscale (depending on the total number of items per subscale), replacing the missing values with the mean item score of that subscale.

*Analyses*

Data from the FT12, NTR, and TEDS cohorts were analyzed at one site (by AMW) using Stata version 15 (Stata Corporation, College Station, TX, USA), while data from GENR were analyzed locally (by KB) using R version 3.3.2. The analyses followed a standard operating procedure (similar to Bartels et al. 2018) to ensure uniform data handling and analysis.

First, means, standard deviations (SD) and standard errors (SE) were obtained for each subscale of the study questionnaire for each age and cohort, separately by gender. T-tests were performed for the subscales of each questionnaire to determine if gender differences existed (p-value <0.05 was considered significant), and effect sizes were calculated as Cohen's *d* values (with positive values indicating boys have larger mean levels than girls and vice versa). Effect sizes are emphasized over statistical significance since: 1) we are not hypothesis testing and 2) our sample sizes are large and thus significance is expected for many relationships. Next, Pearson correlations between the subscales were computed for each age and cohort, separately by gender since differences were identifed.

Linear regression models were then employed to formally test gender differences and examine whether gender interactions existed, as well as to assess the amount of variance in aggression that could be explained by the independent variable (using $R^2$ values). Before modeling, the subscale scores were standardized (mean=0, SD=1) in order to allow for comparisons of the same questionnaire across different ages or cohorts (where possible). Initial modeling included aggressive behavior as the dependent variable and one subscale score as the independent variable. There were three types of subscale scores utilized in the initial models: 1)

the externalizing problems subscale with the highest correlation with aggressive behavior (e.g., hyperactivity, inattention, rule breaking); 2) the internalizing problems subscale with the highest correlation with aggressive behavior (e.g., anxiety, depression, somatic); and 3) the prosocial subscale (if available). For each model, the genders were modeled separately, after testing for gender interactions. Lastly, to examine multiple co-occuring behaviors with aggression, multiple regression modeling was performed with aggressive behavior as the dependent variable and the two (or three, if available) subscale measures from the initial models simultaneously modeled as independent variables, modeled separately for gender after testing for gender interactions.

**Results**

Overall, we examined 39,936 teacher ratings on children: 3627 observations from FT12, 4512 observations from GENR, 18,569 observations from NTR, and 13,228 observations from TEDS. Regarding the spread of data at different ages, there were 17,267 observations of 7-year-olds (49.3% boys), 2762 of 9-year-olds (46.9% boys), 6582 of 10-year-olds (49.6% boys), 11,884 of 12-year-olds (48.4% boys), and 1441 of 14-year-olds (48.4% boys). Interactive results can also be found at (http://www.action-euproject.eu/TeacherRatingsChildAggression).

*Average Scores*

Across all cohorts, mean levels were higher for boys than girls in all externalizing subscale scores, with Cohen's *d* values ranging 0.31–0.69 (Figure 1 A-C, Supp Table 1 A-C).

Results regarding internalizing problems depended on the cohort, though gender effect sizes were smaller than externalizing problems and prosocial behaviors (Figure 1 A-C, Supp Table 1 A-C). In FT12, girls had slightly higher levels of internalizing problems compared to boys (Cohen's *d* range: -0.09–(-)0.23). In GENR and NTR, for all internalizing problems, boys and girls generally had

similar levels (Cohen's *d* range: -0.04–0.12 in GENR and -0.09–0.10 in NTR). In TEDS, levels of anxiety problems between boys and girls were also generally similar (Cohen's *d* range: -0.08–0.03).

Additionally, for all cohorts, mean differences between boys and girls were found in all social behaviors, with boys having slightly more social/peer problems (Cohen's *d* range: 0.10–0.24) and girls having higher prosocial scores (Cohen's *d* range: -0.37–(-)0.63).

*Co-occurrence of aggression and other problems*

Overall, the strength of the correlations of aggressive behavior with other externalizing problems ranged 0.47–0.80, for aggressive behavior and internalizing problems 0.02–0.39, and for aggressive behavior and prosocial behavior 0.33–0.54 (note: some internalizing problems and all prosocial behavior correlation values with aggressive behavior were negative)(Figure 2 A-C, Supp Table 2). In general, correlation patterns of aggressive behavior with other co-occurring behaviors were similar across cohorts and ages and between genders.

In FT12, correlations of aggressive behavior and other externalizing problems (i.e., hyperactivity–impulsivity and inattention) ranged from 0.51–0.75 for boys and 0.48-0.70 for girls (Figure 2 A-C, Supp Table 2). For aggressive behavior and depression, correlations ranged 0.16–0.29, while aggressive behavior and social anxiety correlations were negative, ranging from -0.10 to -0.02. Correlations of aggressive and prosocial behavior ranged from -0.33 to -0.54. Finally, although the relationships between aggression and co-occurring behaviors are the main focus of this study, other noteworthy correlations were hyperactivity–impulsivity with inattention (r range: 0.61–0.74), depression with social anxiety (r range: 0.53–0.62), and prosocial with inattention (r range: -0.44 to -0.61).

In GENR, correlations of aggressive behavior and other externalizing problems (i.e., rule-breaking behavior and attention problems) ranged 0.69–0.80 (Figure 2 A-C, Supp Table 2). For

correlations of aggressive and internalizing problems (i.e., anxious/depressed, somatic complaints, and withdrawn/depressed), correlations ranged 0.14–0.39, with aggressive behavior and anxious/depressed correlations being highest and aggressive behavior and somatic complaints correlations being lowest. Other noteworthy patterns were that correlations of rule-breaking behavior and internalizing problems were 1.3–1.7-fold higher in boys than girls, and that correlations of social problems with all other subscales were nearly identical for boys and girls.

In NTR, correlations of aggressive behavior and other externalizing problems (i.e., rule-breaking behavior and attention problems) ranged from 0.69–0.78 for boys and 0.62–0.67 for girls (Figure 2 A-C, Supp Table 2). For correlations of aggressive behavior and internalizing problems (i.e., anxious/depressed, somatic complaints, and withdrawn/depressed), correlations ranged 0.26–0.38 for aggressive behavior with anxious/depressed and 0.15–0.19 for aggressive behavior with withdrawn/depressed and somatic complaints. Other noteworthy patterns were that correlations of social problems with all other subscales were very similar for boys and girls (as with GENR).

In TEDS, correlations of aggressive behavior and other externalizing problems (i.e., hyperactivity) ranged from 0.51–0.58 for boys and 0.47-0.50 for girls (Figure 2 A-C, Supp Table 2). For aggressive behavior and anxiety, correlations ranged 0.10–0.21, while aggressive behavior and peer problems correlations ranged 0.26–0.42. Correlations of aggressive and prosocial behaviors were negative and ranged -0.43–(-)0.50. Finally, other noteworthy patterns were that while correlations of hyperactivity with anxiety remained similar over the ages (r range: 0.19–0.23), correlations of peer problems with both of them change at age 12, after having been similar for ages 7 and 9 (r range for boys and girls ages 7 and 9: 0.30–0.37 for both correlations; r values age 12 for peer problems with hyperactivity: 0.49 for boys, 0.45 for girls; r values age 12 for peer problems with anxiety: 0.19 for boys, 0.21 for girls).

*Linear Regression Models*

In initial modeling (i.e., aggressive behavior as the dependent variable and one subscale as the independent variable), gender was significantly associated with aggression. Additionally, all GENR and NTR models indicated significant gender interaction terms, and in the TEDS cohort all but the age-12 anxiety model resulted in a significant gender interaction (Table 1 A-C). In contrast, for FT12, the age-12 and age-14 internalizing models and age-14 externalizing model did not indicate significant gender interaction. Regarding model $R^2$ values (% of variance in dependent variable explained by independent variables), externalizing models were consistently largest (boy range: 0.26–0.63; girl range: 0.22–0.54). For internalizing models, $R^2$ values were small (boy range: 0.03–0.15; girl range: 0.01–0.11). For prosocial models in FT12, the $R^2$ values ranged 0.11–0.29 and were 1.9–4.4-fold smaller than externalizing $R^2$ values, while in TEDS, the $R^2$ values ranged 0.18–0.25 and were only 1.1–1.5-fold smaller than externalizing $R^2$ values.

In the multiple regression modeling (i.e., multiple subscales included simultaneously as independent variables), the $R^2$ values did not generally differ from values for externalizing only models (Table 2 A-C). However, the independent variables in these analyses were generally significant. All betas (effect sizes) were generally smaller in these models compared to models with only one independent variable at a time, with betas for externalizing behaviors being the largest of the independent variables. Gender interactions were also generally significant in these models (exceptions being most FT12 independent variables and anxiety for TEDS age 9 and 12 models, and prosocial for the TEDS age 12 model).

**Discussion**

This study provides a reference panel of average levels and co-occurence of aggressive behavior and other behaviors for given ages in the European school setting. Results indicate that patterns are quite similar across the different cohorts of Finland, the Netherlands, and the UK. We show that, as expected, the levels of aggressive behavior are statistically significantly different by gender, however, the effect sizes are only moderate. Generally, regarding levels of all behaviors, externalizing problems were higher among boys, internalizing problems were similar between the genders, and prosocial behavior was higher among girls; all effect sizes ranged from small to moderate. Furthermore, we see that aggression often co-occurs with other behavioral and emotional difficulties, and correlations of aggressive behavior and other co-occurring behaviors are, in practice, quite similar between the genders. Regression modeling indicated that much of the variation ($R^2$) in aggressive behavior levels was explained by other co-occurring externalizing problems, though prosocial models also had rather large $R^2$ values in initial individual models. Modeling multiple co-occurring behaviors simultaneously did not generally improve the explained variance but all modeled co-occurring behaviors, generally, remained significant in these models (but with smaller effect sizes). Thus, it seems that children with aggressive behavior often have not only one co-occurring problem behavior, but possibly multiple co-occurring problems (including low prosocial skills).

There was a very strong co-occurence of aggressive behavior with other externalizing problems. What is remarkable are the similarities in the gender patterns of aggressive behavior and other externalizing problems. Although boys did exhibit the expected statisically significantly higher levels and associations of aggression and other externalizing problems compared to girls, the mean level effect sizes were only moderate and correlations between aggression and co-occurring externalizing behaviors were very similar, the difference in size was never more than 0.1 between boys and girls. In this respect, it is important to point to another study on the NTR cohort

regarding gender differences in ADHD diagnosis and comorbidity using teacher ratings (Derks, 2007) in which boys and girls had similar comorbidity profiles and school impairment. However, girls were less likely than boys to be identified by teachers as disruptive and referred for treatment. Similarly, conduct disorder and oppositional defiant disorder have both been much more likely to be attributed to boys than girls among teacher ratings, compared to parent ratings (Offord et al. 1996, Maughan et al. 2004). While keeping situational aggressive behavior and discrepancies between raters in mind (Achenbach et al. 1987, Verhulst et al. 1994, De Los Reyes et al. 2005), we also need to consider that girls may not be as readily identified as being aggressive by teachers as boys, although their patterns of comorbid problems and need for treatment may be similar.

While the relationships between aggressive behavior and co-occurring externalizing problems are well-established, the co-occurrence with internalizing problems are less well described. Generally, there is moderate correlation between aggressive and depressive symptoms (Van den Oord 2000), but the relationship with anxiety is less well described. The behavioral questionnaires used here cannot fully clarify because internalizing problems were characterized and measured in various ways. The MPNI (used by FT12) indicated that depressiveness was moderately correlated with aggressive behavior, while social anxiety was not correlated with aggressive behavior. The TRF (used by GENR and NTR) showed higher correlations between anxious–depressed and aggressive behavior than between withdrawn–depressed or somatic problems, while the SDQ (used by TEDS) anxiety measure showed low correlation with aggressive behavior. These mixed results regarding aggressive behavior and internalizing problems could partially be explained by August et al. (1996) who found that, among schoolchildren, anxiety/mood disorders most often co-occur with multiple externalizing disorders (instead of, for example, only conduct disorder [a severe aggression problem]). However, these correlations do

indicate a relationship with aggressive behavior, and patterns were similar across genders, thus, future studies should clarify the nuances of these relationships which are perhaps less often considered in the school setting than the more overt and administratively problematic externalizing behaviors.

Prosocial behavior has been shown to be negatively associated with aggressive behavior (Eron and Huesmann, 1984), which was seen in this study as well. Correlations of prosocial behavior with aggressive behavior were moderate in strength. In regression models, prosocial behavior alone explained much of the variance in aggressive behavior, but the effect of prosocial behavior attenuated substantially (though remained significant) when adding externalizing behavior as an additional independent variable of aggressive behavior. The relationships appear complex, students exhibiting aggressive behavior are often struggling not only with other externalizing problem behaviors but also a lack of social skills. On the other hand, Kokko et al. (2000) has shown that prosocial behavior is protective in aggressive children against future unemployment. Additionally, Hämäläinen and Pulkkinen (1996) have shown that aggressiveness without other co-occurring problems did not predict future criminality, although low prosocial behavior alone did. Furthermore, although we found statistically significant gender differences, we saw that prosocial–aggressive behavior correlations were still less than 0.15 different between boys and girls, and the differences between genders in TEDS correlation values were half the size of those in FT12. This may be related to the severity of the aggression measure, since the SDQ (used in TEDS) captures conduct problems which are more serious than the general aggression captured by the MPNI (used in FT12), thus suggesting that, especially when comparing the highest levels of aggression with levels of prosocial behavior, there are no gender differences.

In our study protocol, each age from each cohort represent a cross-sectional snapshot of behavioral and emotional problems, with some children allowed to be captured in more than one

age category. Because all data are drawn in large numbers and in a population-based manner, these data do reflect developmental stages well. Over the ages presented here, we can see that age patterns do not generally differ by gender. For all cohorts, the differences in aggressive–co-occurring behavior correlations between all ages were less than 0.11. In longitudinal settings, the general trend is that population levels of aggressive behavior diminish as age increases, although there tends to be stability in rank order across certain individuals (those with high aggression and those of lower socioeconomic status) (Odgers et al. 2008, Moffitt 1993; Pulkkinen 2017; Tremblay et al., 2004).

Considering further comparison between cohorts, we have to be careful to indicate that we are actually comparing behavioral questionnaires as well. We see similar patterns in the similar subscales across cohorts/questionnaires, which is noteworthy since the cohorts represent different European countries and the questionnaires were developed for different purposes. The MPNI was developed mainly to characterize behavior and emotion regulation, the SDQ was developed as a behavioral screening tool, while the TRF can be used for both psychopathology screening and diagnostic capabilities. Furthermore, both the ASEBA system (of which the TRF is a part) and the SDQ were assessed in Achenbach et al. (2008) and generally found to be comparable across countries, with more variation found within populations than between.

Broadening our considerations and comparisons further, although direct assessment of differences in country or school systems would be inappropriate in this context, we briefly present some items to consider in the Supplemental Material (see Supplemental Text for further details on school systems and teacher rating collections in Finland, the Netherlands, and the UK). Regarding country differences, Hendricks et al. (in press) found less adolescent aggressive behaviors in countries with more policies for child and adolescent mental health (Finland had more policies and less violence compared to the UK and the Netherlands). In considering the 2015 PISA for a general

sense of exposure to bullying/violence in schools in the different countries, we see: 14.2% of schoolchildren in the UK, 9.5% in Finland, and 3.3% in the Netherlands are 'frequently' exposed to bullying (OECD average: 8.9%) (OECD, 2017). It is possible these differing bullying frequencies affect a teacher's perception of problem behaviors and the prevalence of aggressive behavior among the children themselves. However, although these differences in exposure to bullying exist across countries, the patterns of aggressive behavior and most co-occuring behaviors remain notably similar.

One additional comparison of the patterns seen here can be made to a parallel analysis among the same ACTION cohorts, using parental ratings (Bartels et al. 2018; interactive results also available at http://www.action-euproject.eu/ComorbidityChildAggression). Overall, comparing parental ratings at similar ages, teacher-based correlations of aggressive behavior with co-occuring behaviors are generally higher for externalizing problems and lower for internalizing problems and prosocial behaviors. In line with these differences, the meta-analysis of Achenbach et al. (1987) on cross-informant agreement showed only moderate agreement between teacher ratings and parent assessments (0.27). Explanations include actual situational differences in behaviors and that parents have fewer opportunities for making observations of many similar-aged children. These rater differences are also being further explored in forthcoming ACTION publications pertaining to (epi)genetic-wide association analyses.

These results can be further utilized in aggression interventions, both school-based and ones supporting the child in all environments. School-based aggression interventions which can organize learning and hobbies at school in a way that enforces social skills and which can identify children at special risk (high levels of aggression with comorbid problems) and guide their parents to seek help from specialists are of particular importance. One such intervention with positive results was performed by Greenberg et al. (2010). Their "multi-year universal social–emotional

learning program" was implemented and showed reduced levels of aggression and increased prosocial skills (results were most notable among children with higher baseline levels of aggression).

Despite the nearly 40,000 teacher-rated observations of behavioral and emotional problems among school children ages 7–14 collected across three European countries, there are limitations to consider in this study. Regarding the generalizability of these results, these data come from three higher income countries of Europe, but it is unclear if these patterns would remain in lower income countries or countries of differing cultural backgrounds. For example, a recent publication utilizing data from 63 low- and middle-income countries collected by the World Health Organization's Global School-based Student Health Survey indicates that gender differences in physical aggression are stronger as a function of greater gender equality in a society (Nivette, 2018).

Lastly, three of the four cohorts consist of twins. While twins are born on average prematurely and of lower birthweight, they are generally indistinguishable from singletons later in life on average on multiple traits and conditions. More specifically, Pulkkinen et al. (2003) compared the FT12 twins included in the present study with their 23,200 classmates, and showed that the twins are very representative of Finnish schoolchildren at ages 11–12 with respect to both internalizing and externalizing behaviors on the MPNI. In the current analysis, we had assessments of 7-year-old twins (NTR) and singletons (GENR) from the Netherlands using the same assessment instrument and found the correlations of aggression with other scales was very similar in twins and singletons. De Zeeuw et al. (2012) looked at differences between twins included in the current project and their nontwin siblings in educational achievement and found that any differences could largely be explained by birth order within the family.

We have presented a reference set of teacher-rated childhood aggressive behavior levels and associations with co-occurring behaviors in the school setting. These results indicate that aggressive behavior regularly co-occurs with other externalizing behaviors and a lack of prosocial skills, and moderately co-occurs with internalizing problems. Furthermore, although differences exist regarding levels of aggressive behavior between boys and girls, we draw attention to the relative similarities in patterns of associations between aggressive behavior and co-occurring behaviors across genders. Teachers are a valuable resource for identifying children in need of specific support. However, teacher trainings could help to reduce potential gender prejudice regarding problem behaviors, instead focusing on all children, and to recognize that those with one problem behavior likely have multiple problem behaviors they are struggling with. Additionally, school interventions for aggression need to be holistic, focusing on broad behavioral and emotional improvement.

**References**

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: implications of cross-informant correlations for situational specificity. Psychol Bull, 101(2), 213-232. doi:10.1037/0033-2909.101.2.213

Achenbach, T.M., & Achenbach, T.M. (1991). Manual for the Teacher's report form and 1991 profile. Burlington, Vt.: Dept. of Psychiatry, University of Vermont

Achenbach, T.M., & Rescorla, L.A. (2001). Manual for the ASEBA School-Age Forms & Profiles. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.

Achenbach, T. M., Becker, A., Dopfner, M., Heiervang, E., Roessner, V., Steinhausen, H. C., & Rothenberger, A. (2008). Multicultural assessment of child and adolescent psychopathology with ASEBA and SDQ instruments: research findings, applications, and future directions. J Child Psychol Psychiatry, 49(3), 251-275. doi: 10.1111/j.1469-7610.2007.01867.x

August, G. J., Realmuto, G. M., MacDonald, A. W., 3rd, Nugent, S. M., & Crosby, R. (1996). Prevalence of ADHD and comorbid disorders among elementary school children screened for disruptive behavior. J Abnorm Child Psychol, 24(5), 571-595. doi: 10.1007/BF01670101

Bartels, M., Hendriks, A., Mauri, M., Krapohl, E., Whipp, A., Bolhuis, K., . . . Boomsma, D. I. (2018). Childhood aggression and the co-occurrence of behavioural and emotional problems: results

across ages 3-16 years from multiple raters in six cohorts in the EU-ACTION project. Eur Child Adolesc Psychiatry, 27(9), 1105-1121. doi: 10.1007/s00787-018-1169-1

Carbonneau, R., Tremblay, R. E., Vitaro, F., & Saucier, J. F. (2005). Can teachers' behavior ratings be used to screen early adolescent boys for psychiatric diagnoses? Psychopathology, 38(3), 112-123. doi: 10.1159/000085773

Card, N. A., Stucky, B. D., Sawalani, G. M., & Little, T. D. (2008). Direct and indirect aggression during childhood and adolescence: a meta-analytic review of gender differences, intercorrelations, and relations to maladjustment. Child Dev, 79(5), 1185-1229. doi: 10.1111/j.1467-8624.2008.01184.x

Cleverley, K., Szatmari, P., Vaillancourt, T., Boyle, M., & Lipman, E. (2012). Developmental trajectories of physical and indirect aggression from late childhood to adolescence: sex differences and outcomes in emerging adulthood. J Am Acad Child Adolesc Psychiatry, 51(10), 1037-1051. doi: 10.1016/j.jaac.2012.07.010

De Los Reyes, A., & Kazdin, A. E. (2005). Informant discrepancies in the assessment of childhood psychopathology: a critical review, theoretical framework, and recommendations for further study. Psychol Bull, 131(4), 483-509. doi: 10.1037/0033-2909.131.4.483

de Zeeuw, E.L., van Beijsterveldt, C.E., de Geus, E.J., Boomsma, D.I. (2012). Twin specific risk factors in primary school achievements. Twin Res Hum Genet. Feb;15(1):107-15. doi: 10.1375/twin.15.1.107

Derks, E. M., Hudziak, J. J., & Boomsma, D. I. (2007). Why more boys than girls with ADHD receive treatment: a study of Dutch twins. Twin Res Hum Genet, 10(5), 765-770. doi: 10.1375/twin.10.5.765

Eron L.D., & Huesmann, L.R. (1984). The relation of prosocial behavior to the development of aggression and psychopathology. Aggress Behav, 10, 201-211. doi: 10.1002/1098-2337(1984)10:33.0.CO;2-S

Farrington, D. P. (1995). The development of offending and antisocial behaviour from childhood: key findings from the Cambridge Study in Delinquent Development. *J Child Psychol Psychiatry, 36*(6), 929-964. doi: 10.1111/j.1469-7610.1995.tb01342.x

Fontaine, N., Carbonneau, R., Barker, E. D., Vitaro, F., Hebert, M., Cote, S. M., . . . Tremblay, R. E. (2008). Girls' hyperactivity and physical aggression during childhood and adjustment problems in early adulthood: a 15-year longitudinal study. Arch Gen Psychiatry, 65(3), 320-328. doi: 10.1001/archgenpsychiatry.2007.41

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: a research note. J Child Psychol Psychiatry, 38(5), 581-586. doi: 10.1111/j.1469-7610.1997.tb01545.x

Greenberg, M. T., Bierman, K. L., Coie, J. D., Dodge, K. A., Lochman, J. E., McMahon, R. J., . . . Gr, C. P. P. R. (2010). The Effects of a Multiyear Universal Social-Emotional Learning Program: The Role of Student and School Characteristics. J Consult Clin Psychol, 78(2), 156-168. doi: 10.1037/a0018607

Hämäläinen, M., & Pulkkinen, L. (1996). Problem behavior as a precursor of male criminality. Dev Psychopathol, 8, 443-455. doi: 10.1017/S0954579400007185

Haworth, C. M., Davis, O. S., & Plomin, R. (2013). Twins Early Development Study (TEDS): a genetically sensitive investigation of cognitive and behavioral development from childhood to young adulthood. Twin Res Hum Genet, 16(1), 117-125. doi: 10.1017/thg.2012.91

Hendriks, A. M., Bartels, M., Stevens, G. W. J. M., Walsh, S. D., Torsheim, T., Elgar, F. J., & Finkenauer, C. (in press, 2019). National child and adolescent health policies as indicators of adolescent mental health: A multilevel analysis of 30 European countries. Journal of Early Adolescence. doi: 10.1177/0272431619858413

Hodgins, S., Larm, P., Ellenbogen, M., Vitaro, F., & Tremblay, R. E. (2013). Teachers' ratings of childhood behaviours predict adolescent and adult crime among 3016 males and females. Can J Psychiatry, 58(3), 143-150. doi: 10.1177/070674371305800304

Kaprio, J. (2013). The Finnish Twin Cohort Study: an update. Twin Res Hum Genet, 16(1), 157-162. doi: 10.1017/thg.2012.142

Kokko, K., & Pulkkinen, L. (2000). Aggression in childhood and long-term unemployment in adulthood: a cycle of maladaptation and some protective factors. Dev Psychol, 36(4), 463-472. doi: 10.1037/1522-3736.3.1.332a

Kooijman, M.N., Kruithof, C.J., van Duijn, C.M., Duijts, L., Franco, O.H., van IJzendoorn, M.H. … Jaddoe, V.W. (2016). The Generation R Study: design and cohort update 2017. Eur J Epidemiol, 31(12):1243-1264. doi: 10.1007/s10654-016-0224-9.

Lamb, D.J., Middeldorp, C.M., Van Beijsterveldt, C.E., Boomsma, D.I. (2012) Gene-environment interaction in teacher-rated internalizing and externalizing problem behavior in 7- to 12-year-old twins. J Child Psychol Psychiatry, 53(8), 818–25. doi: 10.1111/j.1469-7610.2011.02497.x

Maughan, B., Rowe, R., Messer, J., Goodman, R., & Meltzer, H. (2004). Conduct disorder and oppositional defiant disorder in a national sample: Developmental epidemiology. Journal of Child Psychology and Psychiatry, 45, 609–621. doi: 10.1111/j.1469-7610.2004.00250.x

McNeilis, J., Maughan, B., Goodman, R., & Rowe, R. (2017). Comparing the characteristics and outcomes of parent- and teacher-reported oppositional defiant disorder: findings from a national sample. J Child Psychol Psychiatry. doi: 10.1111/jcpp.12845

Moffitt, T. E. (1993). Adolescence-limited and life-course-persistent antisocial behavior: a developmental taxonomy. Psychol Rev, 100(4), 674-701. doi: 10.1037/0033-295X.100.4.674

Nivette, A., Sutherland, A., Eisner, M., & Murray, J. (2018). Sex differences in adolescent physical aggression: Evidence from sixty-three low-and middle-income countries. Aggress Behav. doi: 10.1002/ab.21799

Odgers, C. L., Moffitt, T. E., Broadbent, J. M., Dickson, N., Hancox, R. J., Harrington, H., . . . Caspi, A. (2008). Female and male antisocial trajectories: from childhood origins to adult outcomes. Dev Psychopathol, 20(2), 673-716. doi: 10.1017/S0954579408000333

OECD (2017), PISA 2015 Results (Volume III): Students' Well-Being, PISA, OECD Publishing, Paris. doi: 10.1787/9789264273856-en

Offord, D. R., Boyle, M. H., Racine, Y., Szatmari, P., Fleming, J. E., Sanford, M., & Lipman, E. L. (1996). Integrating assessment data from multiple informants. J Am Acad Child Adolesc Psychiatry, 35(8), 1078-1085. doi: 10.1097/00004583-199608000-00019

Pulkkinen, L., Kaprio, J., & Rose, R.J. (1999). Peers, teachers and parents as assessors of the behavioural and emotional problems of twins and their adjustment: the Multidimensional Peer Nomination Inventory. Twin Res, 2(4), 274-285. doi: 10.1375/twin.2.4.274

Pulkkinen, L., Vaalamo, I., Hietala, R., Kaprio, J., & Rose, R. J. (2003). Peer reports of adaptive behavior in twins and singletons: is twinship a risk or an advantage? Twin Res, 6(2), 106-118. doi: 10.1375/136905203321536236

Pulkkinen, L. (2017). *Human Development: From Middle Childhood to Middle Adulthood*. London: Routledge. doi: 10.4324/9781315732947

Pulkkinen, L. (2018). *Longitudinal Study of Personality and Social Development: Insights about aggression after five decades*. In A. T. Vazsonyi, D. J. Flannery, & M. DeLisi (Eds.). The Cambridge Handbook of Violent Behavior and Aggression, 2nd Edition (pp. 31-51). Cambridge, UK: Cambridge University Press. doi: 10.1017/9781316847992

Stanger, C., & Lewis, M. (1993). Agreement among Parents, Teachers, and Children on Internalizing and Externalizing Behavior Problems. Journal of Clinical Child Psychology, 22(1), 107-115. doi: 10.1207/s15374424jccp2201_11

Tremblay, R. E., Nagin, D. S., Seguin, J. R., Zoccolillo, M., Zelazo, P. D., Boivin, M., . . . Japel, C. (2004). Physical aggression during early childhood: trajectories and predictors. Pediatrics, 114(1), e43-50. Doi: 10.1542/peds.114.1.e43

United Nations Educational, Scientific and Cultural Organization (UNESCO). (2019). Behind the numbers: Ending school violence and bullying. UNESCO Publishing, Paris. ISBN 978-92-3-100306-6

Vaalamo, I., Pulkkinen, L., Kinnunen, T., Kaprio, J., & Rose, R. J. (2002). Interactive effects of internalizing and externalizing problem behaviors on recurrent pain in children. J Pediatr Psychol, 27(3), 245-257. doi:10.1093/jpepsy/27.3.245

van Beijsterveldt, C. E., Groen-Blokhuis, M., Hottenga, J. J., Franic, S., Hudziak, J. J., Lamb, D., . . . Boomsma, D. I. (2013). The Young Netherlands Twin Register (YNTR): longitudinal twin and family studies in over 70,000 children. Twin Res Hum Genet, 16(1), 252-267. doi:10.1017/thg.2012.118

Van den Oord, E.J.C.G., Boomsma, D.I., Verhulst, F.C. (2000). A study of genetic and environmental effects on the co-occurrence of problem behaviors in three-year-old twins. J Abnorm Psychol. Aug;109(3):360-372. doi: 10.1037/0021-843X.109.3.360

Verhulst, F. C., & van der Ende, J. (1993). "Comorbidity" in an epidemiological sample: a longitudinal perspective. J Child Psychol Psychiatry, 34(5), 767-783. doi: 10.1111/j.1469-7610.1993.tb01070.x

Verhulst, F. C., Koot, H. M., & Van der Ende, J. (1994). Differential predictive value of parents' and teachers' reports of children's problem behaviors: a longitudinal study. J Abnorm Child Psychol, 22(5), 531-546. doi: 10.1007/BF02168936

Vitaro, F., Brendgen, M., & Barker, E. D. (2006). Subtypes of aggressive behaviors: a developmental perspective. Int J Behav Devel, 30(1), 1-19. doi: 10.1177/0165025406059968

Whipp, A. M., Korhonen, T., Raevuori, A., Heikkila, K., Pulkkinen, L., Rose, R. J., . . . Vuoksimaa, E. (2018). Early adolescent aggression predicts antisocial personality disorder in young adults: a population-based study. Eur Child Adolesc Psychiatry. doi:10.1007/s00787-018-1198-9

Willner, C. J., Gatzke-Kopp, L. M., & Bray, B. C. (2016). The dynamics of internalizing and externalizing comorbidity across the early school years. Dev Psychopathol, 28(4pt1), 1033-1052. doi:10.1017/S0954579416000687

**Figure Legends**

Figure 1 (A-C). Means and standard errors (SEs) of behavioral scales, separated by gender, age, and questionnaire

A. MPNI questionnaire (FT12); Agg=aggression, Dep=depression, Hyp-Imp=hyperactivity-impulsivity, Inatt=inattention, Soc Anx=social anxiety

B. TRF questionnaire (GENR and NTR); AGG=aggression, ANX=anxious/depressed, AP=attention problems, RB=rule-breaking, SOC=social problems, SOM=somatic problems, TP=thought problems, WIT=withdrawn/depressed

C. SDQ questionnaire (TEDS)

Figure 2 (A-C). Pearson correlations between aggression and other behavioral/emotional problems, separated by gender, age, cohort and behavioral questionnaire

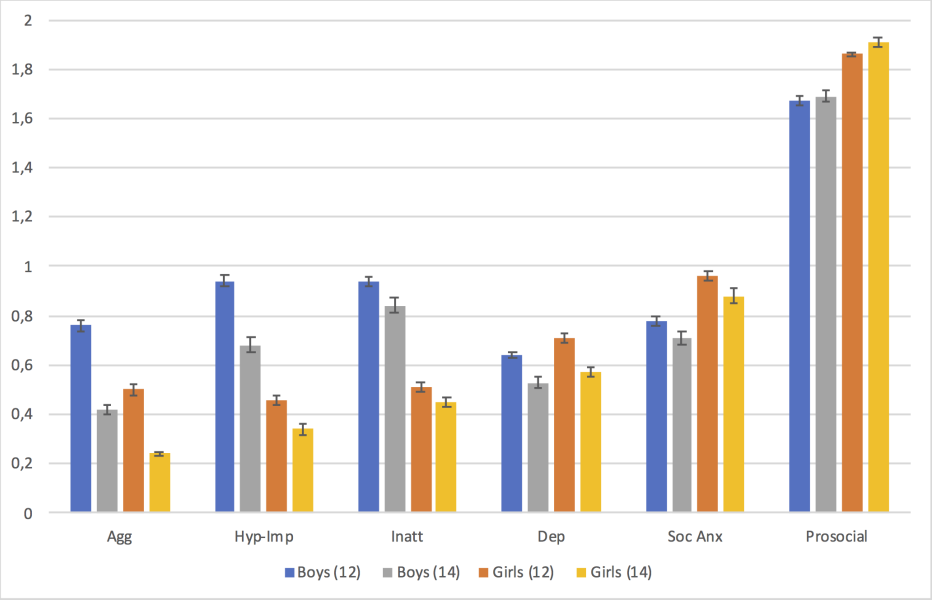
A. MPNI questionnaire (FT12); Dep=depression, Hyp-Imp=hyperactivity-impulsivity, Inatt=inattention, Soc Anx=social anxiety

B. TRF questionnaire (GENR and NTR); AGG=aggression, ANX=anxious/depressed, AP=attention problems, RB=rule-breaking, SOC=social problems, SOM=somatic problems, TP=thought problems, WIT=withdrawn/depressed

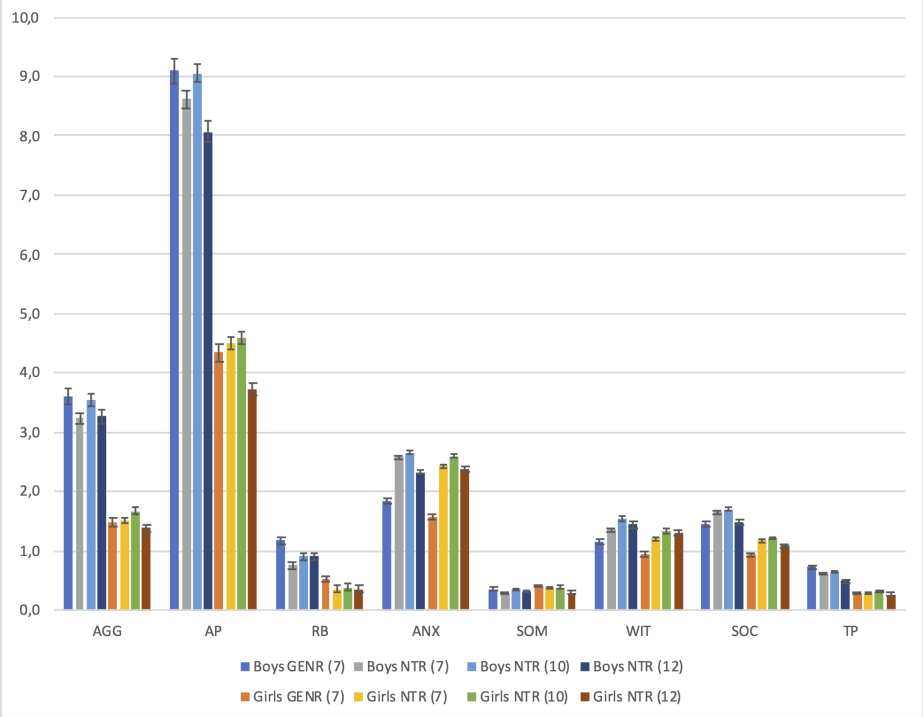
C. SDQ questionnaire (TEDS)

**Figure 1 (A-C)**

A




B




C

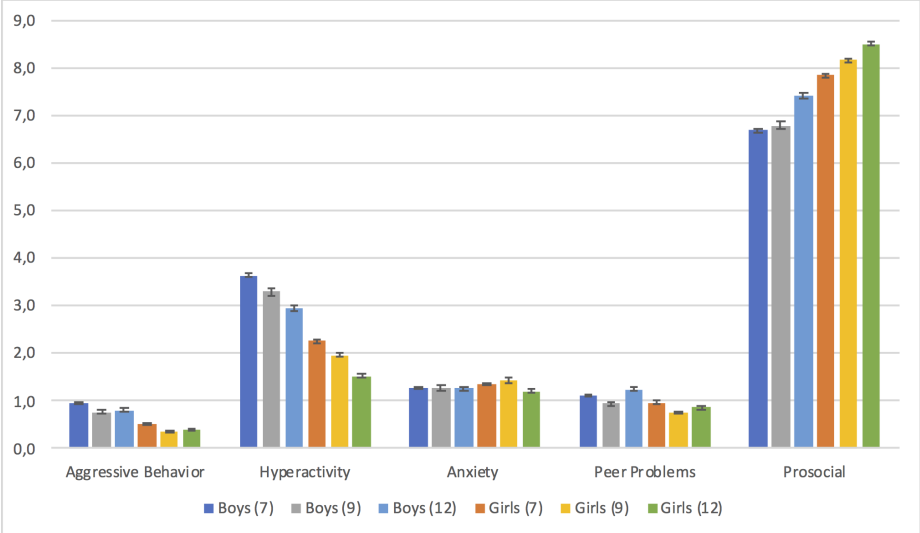
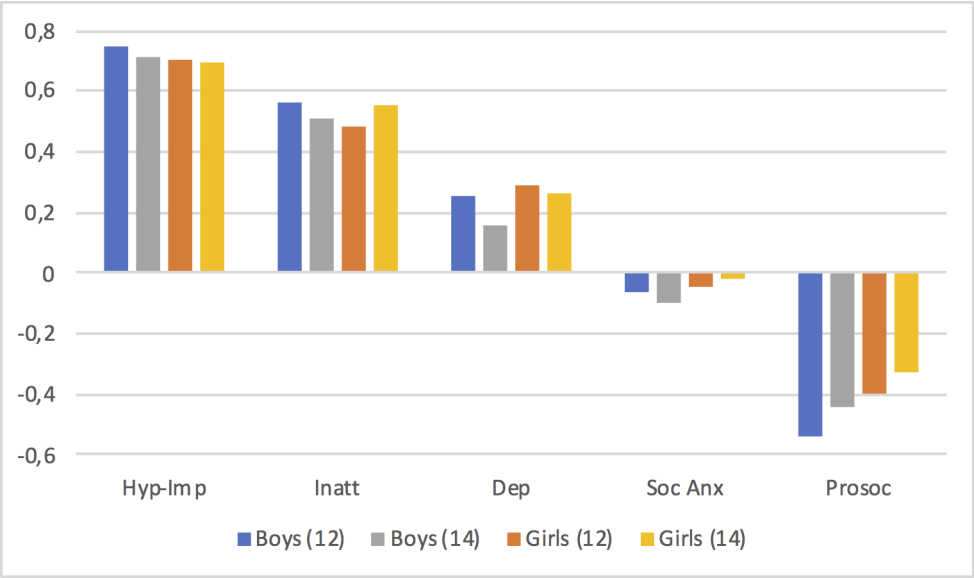
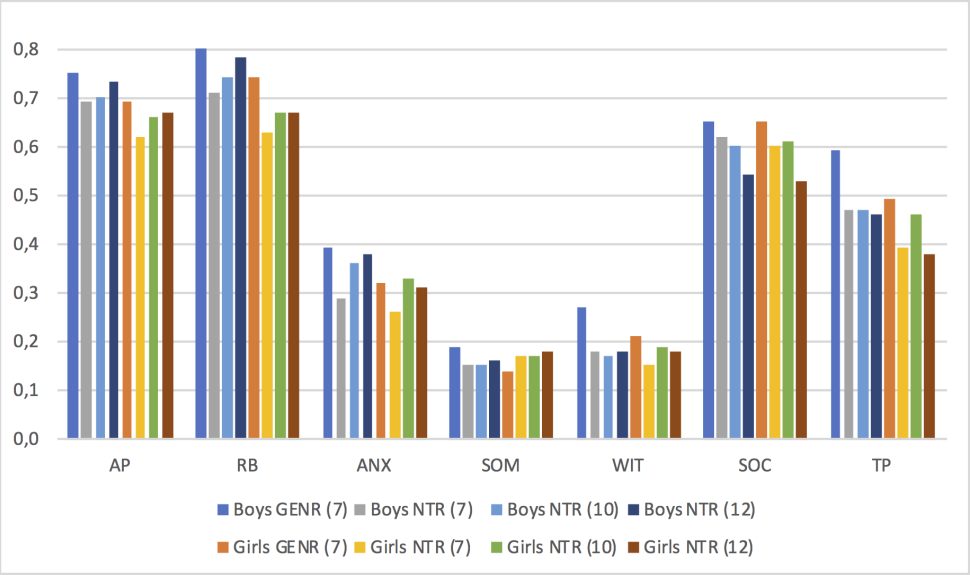


**Figure 2 (A-C)**

A




B




C

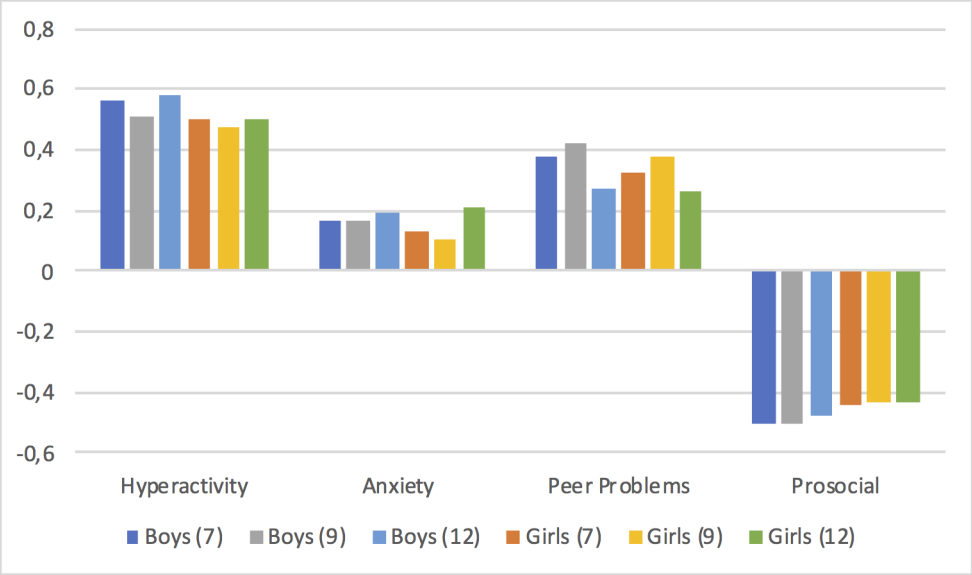


**Table 1 (A-C).** Linear regression summaries of standardized single independent variable models adjusted by age and behavioral questionnaire. Genders modeled separately due to (generally) significant gender interaction.

A. MPNI questionnaire (FT12)

| *Age 12 Boys (n=1105)* | | | | *Age 12 Girls (n=1081)* | | | | Gender interaction model | |
|---|---|---|---|---|---|---|---|---|---|
| **Model Variable** | **Agg Beta** | **95% CI** | **$R^2$** | **Model Variable** | **Agg Beta** | **95% CI** | **$R^2$** | **Interaction term p-value** | **$R^2$** |
| Hyp-Imp | 0.73 | 0.69, 0.77 | 0.56 | Hyp-Imp | 0.80 | 0.75, 0.85 | 0.48 | 0.03 | 0.55 |
| | | | | | | | | | |
| Depression | 0.27 | 0.21, 0.34 | 0.06 | Depression | 0.25 | 0.20, 0.30 | 0.08 | 0.53 | 0.11 |
| | | | | | | | | | |
| Prosocial | -0.57 | -0.62, -0.51 | 0.29 | Prosocial | -0.38 | -0.43, -0.32 | 0.16 | 0.00 | 0.27 |
| | | | | | | | | | |
| *Age 14 Boys (n=697)* | | | | *Age 14 Girls (n=744)* | | | | Gender interaction model | |
| **Model Variable** | **Agg Beta** | **95% CI** | **$R^2$** | **Model Variable** | **Agg Beta** | **95% CI** | **$R^2$** | **Interaction term p-value** | **$R^2$** |
| Hyp-Imp | 0.73 | 0.67, 0.78 | 0.50 | Hyp-Imp | 0.70 | 0.64, 0.75 | 0.48 | 0.46 | 0.51 |
| | | | | | | | | | |
| Depression | 0.20 | 0.11, 0.28 | 0.03 | Depression | 0.21 | 0.16, 0.27 | 0.07 | 0.73 | 0.07 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | |
| Prosocial | -0.50 | -0.57, -0.42 | 0.19 | Prosocial | -0.29 | -0.35, -0.23 | 0.11 | 0.00 | 0.19 |

Abbreviations: Agg=aggression, CI=confidence interval, Hyp-Imp=hyperactivity–impulsivity

B. TRF questionnaire (GENR and NTR)[a]

| **GEN-R** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| *Age 7 Boys (n=2270)* | | | | *Age 7 Girls (n=2242)* | | | | Gender interaction model | |
| **Model Variable** | **Agg Beta** | **95% CI** | **$R^2$** | **Model Variable** | **Agg Beta** | **95% CI** | **$R^2$** | **Interaction term p-value** | **$R^2$** |
| Attention Problems | 0.80 | 0.77, 0.83 | 0.56 | Attention Problems | 0.62 | 0.59, 0.65 | 0.48 | <0.001 | 0.56 |
| | | | | | | | | | |
| Rule-Breaking | 0.81 | 0.78, 0.83 | 0.63 | Rule-Breaking | 0.68 | 0.66, 0.71 | 0.54 | <0.001 | 0.63 |
| | | | | | | | | | |
| Anxious/Depressed | 0.45 | 0.40, 0.49 | 0.15 | Anxious/Depressed | 0.22 | 0.19, 0.25 | 0.10 | <0.001 | 0.17 |
| | | | | | | | | | |
| **NTR** | | | | | | | | | |
| *Age 7 Boys (n=3416)* | | | | *Age 7 Girls (n=3518)* | | | | Gender interaction model | |
| **Model Variable** | **Agg Beta** | **95% CI** | **$R^2$** | **Model Variable** | **Agg Beta** | **95% CI** | **$R^2$** | **Interaction term p-value** | **$R^2$** |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Attention Problems | 0.74 | 0.71, 0.76 | 0.48 | Attention Problems | 0.56 | 0.54, 0.58 | 0.38 | <0.001 | 0.48 |
| | | | | | | | | | |
| Rule-Breaking | 0.72 | 0.70, 0.75 | 0.51 | Rule-Breaking | 0.57 | 0.55, 0.60 | 0.40 | <0.001 | 0.50 |
| | | | | | | | | | |
| Anxious/Depressed | 0.34 | 0.30, 0.37 | 0.09 | Anxious/Depressed | 0.19 | 0.17, 0.21 | 0.07 | <0.001 | 0.12 |
| | | | | | | | | | |
| *Age 10 Boys (n=3264)* | | | | *Age 10 Girls (n=3318)* | | | | Gender interaction model | |
| **Model Variable** | **Agg Beta** | **95% CI** | **$R^2$** | **Model Variable** | **Agg Beta** | **95% CI** | **$R^2$** | **Interaction term p-value** | **$R^2$** |
| Attention Problems | 0.73 | 0.71, 0.76 | 0.48 | Attention Problems | 0.62 | 0.59, 0.64 | 0.43 | <0.001 | 0.49 |
| | | | | | | | | | |
| Rule-Breaking | 0.73 | 0.71, 0.75 | 0.55 | Rule-Breaking | 0.67 | 0.65, 0.70 | 0.45 | 0.004 | 0.54 |
| | | | | | | | | | |
| Anxious/Depressed | 0.43 | 0.39, 0.46 | 0.13 | Anxious/Depressed | 0.24 | 0.22, 0.26 | 0.11 | <0.001 | 0.16 |
| | | | | | | | | | |
| *Age 12 Boys (n=2477)* | | | | *Age 12 Girls (n=2576)* | | | | Gender interaction model | |
| **Model Variable** | **Agg Beta** | **95% CI** | **$R^2$** | **Model Variable** | **Agg Beta** | **95% CI** | **$R^2$** | **Interaction term p-value** | **$R^2$** |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Attention Problems | 0.77 | 0.75, 0.80 | 0.54 | Attention Problems | 0.62 | 0.59, 0.65 | 0.45 | <0.001 | 0.54 |
| | | | | | | | | | |
| Rule-Breaking | 0.77 | 0.75, 0.80 | 0.61 | Rule-Breaking | 0.67 | 0.64, 0.70 | 0.45 | <0.001 | 0.59 |
| | | | | | | | | | |
| Anxious/Depressed | 0.47 | 0.42, 0.51 | 0.14 | Anxious/Depressed | 0.2 | 0.18, 0.23 | 0.10 | <0.001 | 0.17 |

[a]Both 'Attention Problems' and 'Rule-Breaking' are modeled due to 'Attention Problems' being more comparable to the other cohorts' modeled externalizing behavior measure, although 'Rule-Breaking' was the externalizing behavior most highly correlated with aggression (a specified 'rule' in the methods for inclusion in linear regression)

Abbreviations: Agg=aggression, CI=confidence interval

C. SDQ questionnaire (TEDS)

| *Age 7 Boys (n=2834)* | | | | *Age 7 Girls (n=2987)* | | | | Gender interaction model | |
|---|---|---|---|---|---|---|---|---|---|
| **Model Variable** | **Agg Beta** | **95% CI** | **$R^2$** | **Model Variable** | **Agg Beta** | **95% CI** | **$R^2$** | **Interaction term p-value** | **$R^2$** |
| Hyperactivity | 0.59 | 0.56, 0.63 | 0.31 | Hyperactivity | 0.49 | 0.46, 0.52 | 0.25 | <0.001 | 0.31 |
| | | | | | | | | | |
| Anxiety | 0.20 | 0.16, 0.25 | 0.03 | Anxiety | 0.11 | 0.08, 0.14 | 0.02 | <0.001 | 0.05 |
| | | | | | | | | | |
| Prosocial | -0.56 | -0.59, -0.52 | 0.25 | Prosocial | -0.40 | -0.43, -0.37 | 0.19 | <0.001 | 0.25 |
| | | | | | | | | | |

| *Age 9 Boys (n=1295)* | | | | *Age 9 Girls (n=1467)* | | | | Gender interaction model | |
|---|---|---|---|---|---|---|---|---|---|
| **Model Variable** | **Agg Beta** | **95% CI** | **$R^2$** | **Model Variable** | **Agg Beta** | **95% CI** | **$R^2$** | **Interaction term p-value** | **$R^2$** |
| Hyperactivity | 0.57 | 0.52, 0.62 | 0.26 | Hyperactivity | 0.42 | 0.37, 0.46 | 0.22 | <0.001 | 0.27 |
| | | | | | | | | | |
| Anxiety | 0.21 | 0.14, 0.27 | 0.03 | Anxiety | 0.07 | 0.04, 0.11 | 0.01 | <0.001 | 0.05 |
| | | | | | | | | | |
| Prosocial | -0.56 | -0.62, -0.51 | 0.25 | Prosocial | -0.36 | -0.40, -0.32 | 0.18 | <0.001 | 0.25 |
| | | | | | | | | | |
| *Age 12 Boys (n=2168)* | | | | *Age 12 Girls (n=2477)* | | | | Gender interaction model | |
| **Model Variable** | **Agg Beta** | **95% CI** | **$R^2$** | **Model Variable** | **Agg Beta** | **95% CI** | **$R^2$** | **Interaction term p-value** | **$R^2$** |
| Hyperactivity | 0.59 | 0.56, 0.63 | 0.34 | Hyperactivity | 0.50 | 0.46, 0.53 | 0.25 | <0.001 | 0.33 |
| | | | | | | | | | |
| Anxiety | 0.21 | 0.17, 0.26 | 0.04 | Anxiety | 0.17 | 0.14, 0.20 | 0.05 | 0.092 | 0.07 |
| | | | | | | | | | |
| Prosocial | -0.52 | -0.56, -0.48 | 0.23 | Prosocial | -0.38 | -0.41, -0.35 | 0.19 | <0.001 | 0.24 |

Abbreviations: Agg=aggression, CI=confidence interval

**Table 2 (A-C).** Linear regression summaries of standardized multiple independent variable models by age and behavioral questionnaire. Genders modeled separately due to (generally) significant gender interaction.

A. MPNI questionnaire (FT12)

| *Age 12 Boys (n=1105)* | | | | *Age 12 Girls (n=1081)* | | | | Gender interaction model | |
|---|---|---|---|---|---|---|---|---|---|
| **Model Variable** | **Agg Beta** | **95% CI** | **$R^2$** | **Model Variable** | **Agg Beta** | **95% CI** | **$R^2$** | **Interaction term p-value** | **$R^2$** |
| Hyp-Imp | 0.61 | 0.57, 0.65 | 0.61 | Hyp-Imp | 0.71 | 0.67, 0.76 | 0.54 | 0.00 | 0.60 |
| Depression | 0.10 | 0.06, 0.14 | | Depression | 0.10 | 0.07, 0.13 | | 0.97 | |
| Prosocial | -0.22 | -0.27, -0.18 | | Prosocial | -0.19 | -0.22, -0.16 | | 0.06 | |
| | | | | | | | | | |
| *Age 14 Boys (n=697)* | | | | *Age 14 Girls (n=744)* | | | | Gender interaction model | |
| **Model Variable** | **Agg Beta** | **95% CI** | **$R^2$** | **Model Variable** | **Agg Beta** | **95% CI** | **$R^2$** | **Interaction term p-value** | **$R^2$** |
| Hyp-Imp | 0.66 | 0.60, 0.72 | 0.53 | Hyp-Imp | 0.64 | 0.59, 0.70 | 0.51 | 0.74 | 0.54 |
| Depression | 0.12 | 0.05, 0.18 | | Depression | 0.11 | 0.06, 0.15 | | 0.80 | |
| Prosocial | -0.15 | -0.22, -0.09 | | Prosocial | -0.10 | -0.15, -0.05 | | 0.19 | |

Abbreviations: Agg=aggression, CI=confidence interval, Hyp-Imp=hyperactivity–impulsivity

B. TRF questionnaire (GENR and NTR)

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **GEN-R** | | | | | | | | | |
| *Age 7 Boys (n=2270)* | | | | *Age 7 Girls (n=2242)* | | | | Gender interaction model | |
| **Model Variable** | **Agg Beta** | **95% CI** | **$R^2$** | **Model Variable** | **Agg Beta** | **95% CI** | **$R^2$** | **Interaction term p-value** | **$R^2$** |
| Attention Problems | 0.75 | 0.72, 0.78 | 0.57 | Attention Problems | 0.59 | 0.57, 0.62 | 0.49 | <0.001 | 0.57 |
| Anxious/Depressed | 0.15 | 0.11, 0.18 | | Anxious/Depressed | 0.06 | 0.04, 0.08 | | <0.001 | |
| | | | | | | | | | |
| **NTR** | | | | | | | | | |
| *Age 7 Boys (n=3416)* | | | | *Age 7 Girls (n=3518)* | | | | Gender interaction model | |
| **Model Variable** | **Agg Beta** | **95% CI** | **$R^2$** | **Model Variable** | **Agg Beta** | **95% CI** | **$R^2$** | **Interaction term p-value** | **$R^2$** |
| Attention Problems | 0.70 | 0.68, 0.73 | 0.50 | Attention Problems | 0.53 | 0.51, 0.56 | 0.40 | <0.001 | 0.49 |
| Anxious/Depressed | 0.16 | 0.13, 0.19 | | Anxious/Depressed | 0.09 | 0.07, 0.11 | | <0.001 | |
| | | | | | | | | | |
| *Age 10 Boys (n=3264)* | | | | *Age 10 Girls (n=3318)* | | | | Gender interaction model | |
| **Model Variable** | **Agg Beta** | **95% CI** | **$R^2$** | **Model Variable** | **Agg Beta** | **95% CI** | **$R^2$** | **Interaction term** | **$R^2$** |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | **p-value** | |
| Attention Problems | 0.69 | 0.66, 0.71 | 0.50 | Attention Problems | 0.58 | 0.55, 0.60 | 0.45 | <0.001 | 0.51 |
| Anxious/Depressed | 0.17 | 0.14, 0.20 | | Anxious/Depressed | 0.10 | 0.08, 0.12 | | <0.001 | |
| | | | | | | | | | |
| *Age 12 Boys (n=2477)* | | | | *Age 12 Girls (n=2576)* | | | | Gender interaction model | |
| **Model Variable** | **Agg Beta** | **95% CI** | **$R^2$** | **Model Variable** | **Agg Beta** | **95% CI** | **$R^2$** | **Interaction term p-value** | **$R^2$** |
| Attention Problems | 0.73 | 0.70, 0.76 | 0.55 | Attention Problems | 0.59 | 0.56, 0.62 | 0.46 | <0.001 | 0.55 |
| Anxious/Depressed | 0.15 | 0.11, 0.18 | | Anxious/Depressed | 0.07 | 0.05, 0.09 | | <0.001 | |

Abbreviations: Agg=aggression, CI=confidence interval

C. SDQ questionnaire (TEDS)

| *Age 7 Boys (n=2834)* | | | | *Age 7 Girls (n=2987)* | | | | Gender interaction model | |
|---|---|---|---|---|---|---|---|---|---|
| **Model Variable** | **Agg Beta** | **95% CI** | **$R^2$** | **Model Variable** | **Agg Beta** | **95% CI** | **$R^2$** | **Interaction term p-value** | **$R^2$** |
| Hyperactivity | 0.43 | 0.40, 0.47 | 0.39 | Hyperactivity | 0.38 | 0.35, 0.41 | 0.32 | 0.027 | 0.39 |
| Anxiety | 0.07 | 0.04, 0.11 | | Anxiety | 0.01 | -0.01, 0.04 | | 0.005 | |
| Prosocial | -0.35 | -0.39, -0.32 | | Prosocial | -0.26 | -0.29, -0.23 | | <0.001 | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| *Age 9 Boys (n=1295)* | | | | *Age 9 Girls (n=1467)* | | | | Gender interaction model | |
| **Model Variable** | **Agg Beta** | **95% CI** | **$R^2$** | **Model Variable** | **Agg Beta** | **95% CI** | **$R^2$** | **Interaction term p-value** | **$R^2$** |
| Hyperactivity | 0.40 | 0.34, 0.45 | 0.36 | Hyperactivity | 0.31 | 0.26, 0.35 | 0.28 | 0.014 | 0.36 |
| Anxiety | 0.07 | 0.01, 0.12 | | Anxiety | 0.01 | -0.03, 0.04 | | 0.060 | |
| Prosocial | -0.38 | -0.44, -0.33 | | Prosocial | -0.23 | -0.27, -0.19 | | <0.001 | |
| | | | | | | | | | |
| *Age 12 Boys (n=2168)* | | | | *Age 12 Girls (n=2477)* | | | | Gender interaction model | |
| **Model Variable** | **Agg Beta** | **95% CI** | **$R^2$** | **Model Variable** | **Agg Beta** | **95% CI** | **$R^2$** | **Interaction term p-value** | **$R^2$** |
| Hyperactivity | 0.45 | 0.41, 0.49 | 0.40 | Hyperactivity | 0.37 | 0.33, 0.40 | 0.32 | 0.003 | 0.39 |
| Anxiety | 0.09 | 0.06, 0.13 | | Anxiety | 0.07 | 0.04, 0.09 | | 0.291 | |
| Prosocial | -0.28 | -0.32, -0.24 | | Prosocial | -0.23 | -0.26, -0.20 | | 0.091 | |

Abbreviations: Agg=aggression, CI=confidence interval