Teacher’s Assessment of Gender Differences in School Performance, Social Skills, and Externalizing Behavior From Fourth Through Seventh Grade

Terje Ogden, Asgeir Olseth, Mari-Anne Sørlie, and Silje Hukkelberg

Abstract
Gender differences in teacher ratings of academic performance, social skills, and externalizing behavior were examined from fourth through seventh grade in a group of 1,023 students from 65 primary schools. Cohen’s $d$ and $t$ tests and were used to report the magnitude of gender differences, and structural equation modeling was used to evaluate cross-domain influences in latent cascade models. Girls received significantly more positive teacher ratings than boys, but the differences leveled more out in the social and behavioral domain than in the academic. Teacher assessments of academic performance predicted future social skills to a larger extent than vice versa for both genders.

Keywords
gender differences, primary school, academic achievement, social skills, externalizing behavior

Gender differences in academic achievement in elementary school favoring girls have been a consistent finding in international research for decades. Several reviews have summarized research on how and why gender differences occur in school and also how these are influenced by individual and contextual characteristics (Voyer & Voyer, 2014; Wernersson, 2010). In recent years, gender differences in the relationships between school performance, social skills and behavioral problems, has received increased attention. And whether these are different for girls and boys. In general, non-cognitive variables like social skills have proved to have a significant, albeit moderate influence on academic achievement (Bjorklund-Young & Ronda, 2017; Organisation for Economic Co-Operation and Development [OECD], 2015). Of particular interest to this study is the identification of changes in teacher perceptions of gender differences in academic performance across fourth to seventh grade, and also the mutual relationships between teacher assessments of academic achievement, social skills, and externalizing behavior in boys and girls.

The current study was conducted in Norway, where schooling is mandatory for all children aged 6 to 16 in a unified compulsory school with one single national curriculum. Educational provisions are seen in a context of a coherent and continuous course of 13 school years. Primary school includes first to seventh grade and secondary lower education includes eighth to 10th grade.

Gender Differences in Academic Achievement

Studies of student performance in elementary school show that girls outperform boys in most school subjects. In OECD countries, the gender gap favoring girls is particularly large in reading skills and language-based subjects (Borgonovi et al., 2018). Girls have demonstrated better reading skills than boys as early as in kindergarten (Bleses et al., 2008) and the gender gap seems to persist and even increase throughout elementary school (U.S. Department of Education, 2006). Researchers disagree, however, on how much the gender differences increase during primary school. Some claim that differences at school entry are significantly amplified during elementary school (DiPrete & Jennings, 2012; Entwisle et al., 2007), while others have found that gender differences at school entry increase only minimally over the next grades (Arnot et al., 1998). Most studies that have analyze gender differences in school performance have been cross-sectional and did not capture changes over time. To the best of our knowledge, the current study is therefore among

1Norwegian Center for Child Behavioral Development, Oslo, Norway

Corresponding Author:
Terje Ogden, Norwegian Center for Child Behavioral Development, Essendropsgate 3, Oslo 0368, Norway.
Email: terje.ogden@nubu.no
the few that examines how gender differences develop in the same group of students from fourth through seventh grade.

**Gender Differences in Social Skills**

Although longitudinal studies indicate that the social skills development is similar for boys and girls, there is evidence of gender differences in the level of social skills, that is, regardless of informant, boys seem to receive significantly lower mean ratings than girls (Sørlie et al., 2020). The influence of social skills on future academic achievement seems to be moderate (Jacob, 2002) and the magnitude depends on the emphasis that teachers and schools put on social and cooperative aspects of learning. The contribution of social skills to academic performance seems mostly to be indirect through students’ behavior in the classroom (Farrington et al., 2012). Social skills are also important ingredients in social and emotional learning programs especially at lower grade levels (Domitrovich et al., 2017; Durlak et al., 2011). Using data from the Early Childhood Longitudinal Study, Sung and Chang (2010) found a significant longitudinal relationship between social skills and academic achievement from kindergarten to fifth grade. DiPrete and Jennings (2012) also analyzed data from the Early Child Longitudinal Study and found that social and behavioral skills had substantially important effects on academic outcomes from kindergarten through fifth grade. They also found that gender differences in the development of social and behavioral competence explained a great part of gender differences in reading and mathematics (Jennings & DiPrete, 2010). In the present study, teacher’s assessments of student social skills were used as an observable indicator of their social and emotional competence.

**Gender Differences in Externalizing Behavior**

Externalizing problems are more common among boys than girls (Skogen & Torvik, 2013) and such behavior has been found to interfere with effective schoolwork (Kristoffersen et al., 2015). Students, and particularly boys with externalizing behavior, often struggle with learning and social relations in school (Breslau et al., 2011). Studies investigating effects of externalizing behavior on academic achievement have generally found negative effects, controlling for other variables (Farmer et al., 2002; Halonen et al., 2006; Ladd & Burgess, 2001; McLeod & Fettes, 2007; McLeod & Kaiser, 2004).

**Cross-Domain Influences**

Teacher assessments of academic achievement at any grade level is expected to be influenced by the students’ achievement at earlier levels, but also by past and current characteristics like social skills and externalizing behavior. Academic achievement belongs to the cognitive domain, while social skills and externalizing behavior represent the non-cognitive domain because they are not measured by commonly administered cognitive tests such as IQ tests or academic examinations (Farrington et al., 2012). These influences are referred to as cross-domain influences and may be examined in developmental cascade studies (Masten & Cicchetti, 2010). One example is a study by Masten and colleagues (2005) in which they tested 205 children aged 8 to 12 years and again after 7, 10, and 20 years. Of relevance in this context was the finding that childhood externalizing problems undermined academic competence by adolescence for both boys and girls. But influences were also found in the opposite direction, indicating that good academic performance could prevent externalizing behavior. Pointing in another direction was an intervention study demonstrating that the promotion of academic skills reduced the risk of developing problem behaviors (e.g., Hawkins et al., 1999). Duncan et al. (2007) found that measures of socioemotional behaviors, including externalizing problems and social skills, were generally insignificant predictors of later academic performance, even among children with relatively high levels of problem behavior. So even if teacher ratings of social skills and externalizing behavior show a large difference in favor of girls (Gutman & Schoon, 2013), the direction and strength of these mutual influences are debated.

Few studies have examined the combined impact of social skills and externalizing behavior on academic achievement, and most longitudinal studies have focused on behavioral rather than academic outcomes (Farrington et al., 2013; Gutman & Schoon, 2013). There are exceptions, however, and Durlak et al. (2011) found positive effects of social and emotional interventions on academic achievement with an average effects size of 0.33 on student grades and 0.27 on achievement test scores. Wentzel (1993) found that prosocial behavior and antisocial behavior of sixth- and seventh-grade students independently and significantly predicted their Grade Point Average (GPA), but only prosocial behavior predicted achievement test scores. In a study of social skills and problem behaviors in third and fourth graders in an urban district, Malecki and Elliott (2002) found that student social skills (measured by Social Skills Rating Scale [SSRS]; Gresham & Elliott, 1990) were positively correlated, while problem behavior was negatively correlated with concurrent grades. Overall, externalizing behavior and social skills seem to have an opposite effect on academic achievement, but the negative effect of externalizing behavior on academic achievement applies mainly to boys (Gutman & Schoon, 2013; OECD 2015).

Teacher assessments of girls’ academic skills are in general more positive than those of boys, even if no significant gender differences in terms of general intelligence have been
found (Roth et al., 2015). Some attribute this to the finding that girls in general adapt better to the schools’ social rules and the teachers’ expectations. Girls may simply have better social skills than boys and show less acting-out behavior. In line with this argument, Cornwell et al. (2013) claim that the difference between test scores and teacher grades could be explained by gender differences in non-cognitive competence. This is also consistent with the results from a study by Gustavsen (2017) who found that teachers’ higher ratings of social skills in girls compared with boys, could explain much of the variance in students’ academic skills. A similar conclusion was reached in a literature review by Wollscheid et al. (2018) when they examined grading and assessment practices as a possible causal factor for gender differences in school performance. This raises the question of how gender differences in teacher ratings of social skills and externalizing behavior affects their assessments of student’s academic performance. The current study is based solely on teacher assessments, and therefore, the results will give a fairly good picture of how teachers assess the cross-domain influences from fourth through seventh grade. Specifically, the research questions were the following:

**Research Question 1:** How large are gender differences in teacher assessments of academic achievement in selected school subjects, social competence, and externalizing behavior from Grades 4 through 7 in primary school? Do gender differences vary across subjects and grades?

**Research Question 2:** What are the cross-domain influences between teacher ratings of academic achievement, social skills, and externalizing behavior?

**Research Question 3:** Are the cross-domain influences different for boys and girls?

### Measure

**Participants**

The present study was conducted between 2007 and 2012 using four data collection points from fourth through seventh grade in 65 primary schools in Norway. The dropout rate was reasonably high because of an open cohort design in which several students moved to another school at the end of a school year, and new students were included at the beginning of the next. Students who dropped out of the study from fourth to seventh grade were less likely to have Norwegian as the first language, but had the same gender proportion and attended schools of approximately the same size. Included in the study were 1,023 students for whom teacher assessments could be tracked from fourth through seventh grade. The data were originally collected to test the effectiveness of a schoolwide intervention, but none of the variables used in the present study was applied in that study.

Neither was the aim of the present study to examine intervention effects. Independent sample *t* tests revealed no significant differences between the intervention (*n* = 45) and comparison schools (*n* = 20) on any of the outcome variables at pre- or post-assessment. Thus, all 65 schools were included in the analyses. Measurement invariance tests were conducted to examine construct validity and ensure valid comparisons over time. Some students were however rated by different teachers from fourth to seventh grade, but our design and data did not allow us to examine effects of such changes further. The study from which the study variables were derived was approved by the Regional Committee for Medical and Health Research Ethics (REK South-East).

**Measures**

The present study is based on teacher ratings of academic achievement in selected subjects, social skills and externalizing behavior.

**Academic performance.** At each grade level from fourth to seventh grade, students stayed together as a group and had the same teacher in most subjects. Their academic performance was rated at the end of each grade level by their main teacher (who knew the students best) in five subjects: home language (which is the national majority language), English as second language, mathematics, social studies, and physical education. Data were collected from the Teacher Report Form (TRF; Achenbach, 1991) that had five response alternatives: “far below average” (1), “below average” (2), “average” (3), “over average” (4) and “far above average” (5). A sum score was calculated for the five subjects (Norwegian, English, mathematics, social studies, and physical education) mentioned under the heading of “Academic performance.” Cronbach’s alpha over time ranged from .87 to .88.

**Externalizing behavior** was assessed via teacher reports (TRF; Achenbach, 1991), and the sum score of the subscales of delinquency (nine items) and aggression (five items) were used as indicators at each time point. Each item was evaluated according to a three-category response format: “not true” (0), “somewhat and sometimes true” (1), and “very true or often true” (2). Cronbach’s alpha over time ranged from .48 to .52.

**SSRS—teacher form.** Social skills were measured with the SSRS (Gresham & Elliott, 1990) adding the subscale composites of three underlying dimensions: assertion (10 items), self-control (10 items), and cooperation (10 items). Mean sum scores were calculated based on teacher ratings of how often students practiced the 30 social skills. Each item was scored on a 4-point scale from 1 (never) to 4 (almost all the time). Cronbach’s alpha over time ranged from .83 to .86.
Analytic Approach

The gender differences at each grade were calculated with Cohen’s $d$, which is an effect size indicating the standardized difference between means (Cohen, 1988). The percentage of missing data varied by indicator and time, ranging from 0% to 16% with a mean of 8% over all indicators and times. Latent models were tested using Mplus 7.3 (Muthén & Muthén, 2012), with the robust maximum likelihood (MLR) estimator that provides standard errors and fit indexes robust to the continuous nature of the items. Full information maximum likelihood (FIML) was used to generate estimates for missing data (Enders, 2010). Since data had a multilevel nature (students are clustered in schools), we used the command Type = COMPLEX in Mplus to take into account the dependency among observations. A change in comparative fit index (CFI) $\geq -0.01$ was used as an indication of non-invariance across nested models (Cheung & Rensvold, 2002).

Results

Descriptives

Gender differences in academic performance. The first research question dealt with gender differences in teacher assessments of academic achievement (summarized across five school subjects), in reading and mathematics, in social competence, and externalizing behavior (see Table 1). A small and decreasing gender difference was detected in academic performance in fourth and fifth grade ($d = 0.15$ and $0.12$), but increased again during sixth and seventh grade ($d = 0.17$ and $0.20$) (see Table 1). Cohen (1988) suggested that $d = 0.2$ be considered a “small” (or even trivial) effect size, 0.5 a “medium” effect size, and 0.8 a “large” effect size. However, when assessing an effect size, it may be a good idea to look at previous research within the same topic and see how large effect sizes others have found (Cohen, 1988). In reading, the gender difference in favor of girls increased from $d = 0.23$ to $d = 0.30$, which was lower than the other variables.

| Variable          | Boys       |                  | Girls       |                  | Cohen’s $d$ | 95% CI       |
|-------------------|------------|------------------|------------|------------------|-------------|--------------|
|                   | $n$        | $M$              | $SD$       | $n$              | $M$         | $SD$         |             |             |
| ACAD 4th grade    | 954        | 24.52            | 5.88       | 973              | 25.38       | 5.32         | -0.154      | [-0.243, -0.065] |
| ACAD 5th grade    | 979        | 24.68            | 6.53       | 959              | 25.42       | 6.02         | -0.119      | [-0.208, -0.03]  |
| ACAD 6th grade    | 937        | 24.70            | 6.52       | 927              | 25.75       | 5.94         | -0.168      | [-0.259, -0.077] |
| ACAD 7th grade    | 922        | 25.02            | 6.67       | 912              | 26.28       | 6.05         | -0.198      | [-0.29, -0.106]  |
| READ 4th grade    | 1007       | 3.06             | 1.01       | 1023             | 3.28        | 0.94         | -0.228      | [-0.315, -0.141] |
| READ 5th grade    | 991        | 3.02             | 1.01       | 972              | 3.26        | 0.95         | -0.246      | [-0.335, -0.157] |
| READ 6th grade    | 953        | 3.07             | 1.01       | 940              | 3.35        | 0.92         | -0.294      | [-0.384, -0.203] |
| READ 7th grade    | 928        | 3.11             | 0.98       | 921              | 3.39        | 0.90         | -0.304      | [-0.396, -0.212] |
| MATH 4th grade    | 1007       | 3.28             | 0.96       | 1023             | 3.14        | 0.86         | 0.154       | [0.067, 0.241]  |
| MATH 5th grade    | 991        | 3.25             | 1.00       | 972              | 3.15        | 0.94         | 0.103       | [0.014, 0.191]  |
| MATH 6th grade    | 953        | 3.20             | 1.03       | 940              | 3.18        | 0.95         | 0.023       | [-0.067, 0.113] |
| MATH 7th grade    | 928        | 3.23             | 1.06       | 921              | 3.23        | 0.97         | 0.000       | [-0.091, 0.092] |
| SOSC 4th grade    | 1007       | 80.52            | 13.52      | 1023             | 88.69       | 13.45        | -0.606      | [-0.695, -0.517] |
| SOSC 5th grade    | 991        | 82.44            | 14.66      | 972              | 89.69       | 13.91        | -0.507      | [-0.597, -0.417] |
| SOSC 6th grade    | 953        | 83.21            | 15.44      | 939              | 90.10       | 14.32        | -0.462      | [-0.554, -0.371] |
| SOSC 7th grade    | 926        | 85.14            | 16.66      | 919              | 91.93       | 15.67        | -0.420      | [-0.512, -0.327] |
| EXT 4th grade     | 995        | 7.45             | 10.27      | 1008             | 2.45        | 5.13         | 0.618       | [0.528, 0.708]  |
| EXT 5th grade     | 991        | 6.95             | 10.20      | 972              | 2.50        | 5.35         | 0.546       | [0.455, 0.636]  |
| EXT 6th grade     | 953        | 6.56             | 9.93       | 939              | 2.48        | 6.23         | 0.492       | [0.400, 0.583]  |
| EXT 7th grade     | 927        | 6.37             | 10.40      | 921              | 2.33        | 5.66         | 0.482       | [0.390, 0.575]  |

Note. ACAD = academic performance; READ = reading; MATH = mathematics; SOSC = social skills; EXT = externalizing behavior; CI = confidence interval.
the effect size of 0.5 found in the 2016 OECD PIRLS (Progress in International Reading Literacy Study) reading assessment in Norway (Borgonovi et al., 2018). The trend was different in mathematics, where boys received higher ratings than girls in fourth grade, but the effect size of 0.15 gradually diminished, and at the end of seventh grade, the mean assessment scores were identical for girls and boys. An identical trend was found in the OECD TIMSS (Trends in International Mathematics and Science Study) studies, although based on cross-sectional data: In the 2007 survey, boys outperformed girls in Grade 4, but between 2011 and 2015, girls’ scores remained stable, while boys’ decreased, thereby closing the gap in mathematics (as cited in Borgonovi et al., 2018).

Gender differences in externalizing behavior were considerably higher than those found for academic performance, and as expected, boys scored considerably higher at each measurement point. But the gender gap decreased over time: In fourth grade, Cohen’s $d$ was 0.62, and in seventh grade, it was slightly reduced to 0.48. The gender gap in teacher rated social skills was about the same size as that of externalizing behavior, but the effect size diminished from $d = 0.61$ in fourth grade to $d = 0.42$ in seventh grade (see Table 1).

**Academic Achievement, Social Skills, and Externalizing**

Table 2 shows means and standard deviations for academic achievement, social skills, and externalizing behaviors for girls and boys, separately. Academic achievement was quite stable within gender over time, whereas social skills steadily increased and externalizing behavior decreased over time, especially among boys. For academic achievement, independent sample $t$ test showed a significant gender effect in favor of girls, at Grade 6 and Grade 7. Girls also showed significant higher mean levels of social skills across all time points ($p < .001$). Boys showed significant higher mean levels on externalizing behaviors at all five time points compared with girls ($p < .001$). Noteworthy, boys showed considerable more variation in externalizing behaviors compared with girls.

Table 3 shows the correlations among the attenuated composite variables over time. Correlations across domains within same time point were overall moderate ($r = -.20$ to $.48$). Within each construct, academic achievement was highly correlated over time ($r = .68$), and so was also externalizing behavior ($r = .56$). Social skills, on the other hand, showed more variation in the correlation over time ($r = .38$ to .75), indicating that teacher ratings of this construct is more unstable, and probably more malleable.

**Testing Cross-Domain Effects**

This part of the study relates to the second and third research questions which address the cross-domain influences and gender differences in teacher ratings of academic achievement, social skills, and externalizing behavior. In this analysis, academic achievement included Norwegian, English, mathematics, social science, and physical education, which

---

**Table 2.** Mean Scores and Standard Deviations (SD) of Teacher Ratings and Independent Sample $t$ Test for Differences Between Girls and Boys.

| Variable               | Girls      | Boys      | t test |
|------------------------|------------|-----------|--------|
|                        | $M$        | SD        | $M$    | SD        |          |
| Academic achievement   |            |           |        |           |          |
| ACAD 4th grade         | 15.83      | 3.58      | 15.52  | 3.81      | 1.79     |
| ACAD 5th grade         | 15.82      | 3.62      | 15.58  | 3.92      | 1.40     |
| ACAD 6th grade         | 15.96      | 3.57      | 15.55  | 3.92      | 2.32*    |
| ACAD 7th grade         | 16.34      | 3.68      | 15.74  | 4.05      | 3.33**   |
| Social skills          |            |           |        |           |          |
| SOS 4th grade          | 89.20      | 14.19     | 80.73  | 14.28     | 13.17**  |
| SOS 5th grade          | 89.69      | 13.91     | 82.44  | 14.66     | 11.24**  |
| SOS 6th grade          | 90.10      | 14.32     | 83.21  | 15.44     | 10.06**  |
| SOS 7th grade          | 91.92      | 15.67     | 85.13  | 16.66     | 9.02**   |
| Externalizing behavior |            |           |        |           |          |
| EXT 4th grade          | 2.56       | 5.64      | 7.31   | 10.54     | −12.44** |
| EXT 5th grade          | 2.50       | 5.35      | 6.95   | 10.20     | −12.15** |
| EXT 6th grade          | 2.48       | 6.23      | 6.56   | 9.93      | −10.73** |
| EXT 7th grade          | 2.32       | 5.66      | 6.36   | 10.40     | −10.39** |

Note. ACAD = academic performance; SOS = social skills; EXT = externalizing.

*p ≤ .05. **p ≤ .001.
all loaded significantly from $\lambda = .43$ to $\lambda = .89$. The language indicators showed the strongest standardized factor loadings, followed by social science, mathematics, and last, physical education, across all five time points. All factor loadings in social skills loaded significantly from $\lambda = .68$ to $\lambda = .90$. The strongest indicator was self-control across all time points. Externalizing problems included the two composite indicators delinquency and aggression which showed significant standardized factor loadings at $\lambda = .81$ and $\lambda = .93$. Aggression consistently showed the strongest loading across all time points. The model fitted the data at an acceptable, $F^2(614) = 3380.26$, CFI = .94, TLI = .92, RMSEA = .046 [90% confidence interval (CI): .044, .047].

### Invariance tests

A prerequisite to make valid comparisons over time and gender is measurement invariance. The results of the longitudinal invariance testing are shown in Table 4. As seen, there was no difference in CFI between the configural (I) and the metric model (II). Next, between the metric model (II) and the scalar model (III), there was no difference in CFI, indicating that constructs showed invariance across the four time points. Next, we investigated measurement invariance across boys and girls, within the longitudinal invariant model. Initially, baseline models were established for boys and girls. As seen, the model fit was acceptable for both groups, but somewhat better for girls than for boys (Table 5). When comparing the metric (II) and the configural model (I), results showed $\Delta$CFI > .010, which indicates the presence of non-invariant factor loadings across the groups. Inspections of modification indices (MIs) showed that inequality among girls and boys was related to the social skill construct, and the indicator assertion. This item was therefore allowed to vary among girls and boys. Analysis of the partial invariant metric model resulted in an improved model fit, and acceptable $\Delta$CFI. Last, we considered intercept invariance by comparing Model IIa with Model III. Results revealed a $\Delta$CFI of −.017, above the recommended benchmark value. In sum, the results showed measurement invariance across time points, but partial invariance across boys and girls, due to some differences in factor loadings and intercepts.

### Structural models

The longitudinal model was specified with lagged effects within each construct, cross-lagged effects over time, and correlations between constructs at within same time point. In addition, residuals between same indicators were allowed to correlate over time.

### Driven models

Since the unique contributions of externalizing behavior, social skills, and academic achievement...
were of interest, three models were tested (Figure 1). The externalizing-driven model included cross-lagged paths from the externalizing constructs on academic achievement and social skills (bold lines). The model showed good fit, $\chi^2(1396) = 4690.45$, $CFI = 0.94$, $TLI = 0.93$, $RMSEA = 0.047$ [90% CI: .041, .046], and significant cross-lagged relations from externalizing behavior on both social skills and academic achievement, for both girls (four time points) and boys (five time points) (Table 6). As seen, externalizing behavior related negatively to both social skills and academic achievement, for both girls (four time points) and boys (five time points) (Table 6). As seen, externalizing behavior related negatively to both social skills and academic achievement, and externalizing behavior had a greater impact on subsequent behavior among boys than among girls. The social skill–driven model included cross-lagged paths from social skills on academic achievement and academic achievement (whole lines). The model showed acceptable model fit, $\chi^2(1396) = 4754.81$, $CFI = 0.94$, $TLI = 0.93$, $RMSEA = 0.047$ [90% CI: .046, .049], but significant cross-lagged paths at only two time points each for girls and boys. This indicates that teacher-reported social skills are to a great extent unrelated to subsequent student academic achievement and social skills. For boys, there were significant paths from social skills on both externalizing and academic achievement, but for girls, significant paths only emerged from social skill on externalizing behavior. Social skills were as expected negatively related to externalizing behavior and positively related to academic achievement.

The academic-driven model included cross-lagged paths from academic achievement on externalizing and social skills (stippled lines). The fit of this model was acceptable, $\chi^2(1396) = 4719.63$, $CFI = 0.94$, $TLI = 0.93$, $RMSEA = 0.047$ [90% CI: .046, .049]. The results showed significant cross-lagged paths from academic achievement on both social skills and externalizing behavior, among boys and

Table 5. Test of Gender Invariance Within an Invariant Longitudinal Framework.

| Baseline models and factor loadings | $\chi^2 (df)$ | RMSEA (90% CI) | CFI | TLI | $\Delta$CFI |
|-----------------------------------|---------------|----------------|-----|-----|-------------|
| Girls ($n = 1120$)                | 2007.95* (658) | .044 (.041, .046) | .943 | .933 | –           |
| Boys ($n = 1103$)                 | 2281.25* (656) | .048 (.046, .050) | .942 | .930 | –           |
| I. Factor loadings and intercepts are freely estimated | 4892.91* (1318) | .050 (.049, .052) | .931 | .918 | –           |
| II. Factor loadings invariant across groups, intercepts freely estimated | 5851.53* (1330) | .056 (.055, .056) | .912 | .897 | –.019       |
| II a. Non-invariant loadings       | 5168.54* (1326) | .052 (.050, .053) | .925 | .912 | –.006       |
| III. Factor loadings and intercepts invariant across groups | 6079.19* (1366) | .057 (.055, .058) | .908 | .895 | –.017       |
| III a. Non-invariant intercepts    | 5667.78* (1357) | .054 (.053, .056) | .916 | .904 | –.009       |

Note. A difference in CFI $\leq .010$ indicates invariance. *$p \leq .05$.
girls. Academic achievement related positively with social skills and negatively to externalizing behavior.

Testing the full model. The full model, including all the hypothesized paths, resulted in acceptable model fit, \( \chi^2(1370) = 4458.62, \text{CFI} = 0.94, \text{TLI} = 0.93, \text{RMSEA} = 0.046 \) [90 % CI: .044 , .047] (Table 7). A large MI indicated an additional path from ACAD4 on ACAD6 (MI = 80.85 for girls, MI = 102.03 for boys), which was included into the model (\( \beta = .38 \) for girls, \( \beta = .41 \) for boys, both \( p < .05 \)). No significant cross-lagged paths appeared for social skills on neither academic achievement nor externalizing behavior. Externalizing behavior showed cross-lagged effects on social skills among boys but was also related to academic achievement among girls \( (\text{EXT5} \rightarrow \text{ACAD6}) \). Cross-lagged paths from academic achievement predicted social skills at all time points for boys \( (\beta = .12-.20) \), and the same appeared for girls, but only at two time points \( (\beta = .13-.21) \).

Discussion

The current longitudinal study supports findings from previous cross-sectional studies that gender differences in academic performance increasingly favor girls from fourth through seventh grade. The largest difference, in favor of girls, was found for reading, which also matches international findings (Borgonovi et al., 2018). In mathematics, the trend was a bit different, showing that boys did better than girls in fourth grade, but then the gender difference decreased over the years and leveled out in seventh grade. The gender differences in social skills and externalizing behavior was generally higher than the gender differences in academic achievement. While girls received higher ratings in social skills, boys were rated higher in terms of externalizing behavior, and the gender difference far exceeded those reported for academic achievement. Contrary to the development in academic performance, however, both the assessment scores and the gender difference decreased over time. The explanation for this may be that boys are increasingly exposed to a formal pressure of reducing externalizing behavior and informal expectations to adapt to the social demands at school. Taken together, gender differences in favor of girls were most evident in reading and in non-cognitive variables like social skills and externalizing behavior. The ratings of general academic performance showed a more moderate difference in favor of girls. While ratings of girls’ behavior were rather stable over the years, boys seemed to improve by getting increasingly more positive ratings of behavior. Gender differences were least evident in mathematics indicating that boys and girls tended to perform more equally when assessments were based on numeracy skills. In line with

Table 6. Standardized Paths for Boys and Girls (Driven Models).

| Models | Girls \( \beta \) (p) | Boys \( \beta \) (p) |
|--------|---------------------|---------------------|
| Externalizing driven | | |
| EXT4 \( \rightarrow \) ACAD5 | -.04** | - |
| EXT4 \( \rightarrow \) SOS5 | -.17** | -.28** |
| EXT5 \( \rightarrow \) ACAD6 | - | -.09** |
| EXT5 \( \rightarrow \) SOS6 | - | -.17** |
| EXT7 \( \rightarrow \) ACAD7 | -.09** | -.05** |
| EXT7 \( \rightarrow \) SOS7 | -.12** | -.15** |
| Social skills driven | | |
| SOS4 \( \rightarrow \) ACAD5 | - | - |
| SOS4 \( \rightarrow \) EXT5 | -.12** | -.12** |
| SOS5 \( \rightarrow \) ACAD6 | - | -.09** |
| SOS5 \( \rightarrow \) EXT6 | - | - |
| SOS6 \( \rightarrow \) ACAD7 | - | - |
| SOS6 \( \rightarrow \) EXT7 | -.09** | - |
| Academic driven | | |
| ACAD4 \( \rightarrow \) SOS5 | .16** | - |
| ACAD4 \( \rightarrow \) EXT5 | -.07** | -.05** |
| ACAD5 \( \rightarrow \) SOS6 | - | -.10** |
| ACAD5 \( \rightarrow \) EXT6 | - | - |
| ACAD6 \( \rightarrow \) SOS7 0.126 | .13** | .14** |
| ACAD6 \( \rightarrow \) EXT7 | -.05** | -.04, p = .06 |

Note. ACAD = academic performance; SOS = social skills; EXT = externalizing. **p \leq .001

Table 7. Standardized Paths for Boys and Girls (Full Model).

| Models | Girls \( \beta \) (p) | Boys \( \beta \) (p) |
|--------|---------------------|---------------------|
| EXT4 \( \rightarrow \) ACAD5 | - | - |
| EXT4 \( \rightarrow \) SOS5 | -.20** | -.35** |
| EXT5 \( \rightarrow \) ACAD6 | - | - |
| EXT5 \( \rightarrow \) SOS6 | - | -.21** |
| EXT6 \( \rightarrow \) ACAD7 | -.10** | - |
| EXT6 \( \rightarrow \) SOS7 | -.11, p=.06 | -.18** |
| SOS4 \( \rightarrow \) ACAD5 | - | - |
| SOS4 \( \rightarrow \) EXT5 | - | - |
| SOS5 \( \rightarrow \) ACAD6 | - | - |
| SOS5 \( \rightarrow \) EXT6 | - | - |
| SOS6 \( \rightarrow \) ACAD7 | - | - |
| SOS6 \( \rightarrow \) EXT7 | -.06, p = .054 | - |
| ACAD4 \( \rightarrow \) SOS5 | .21** | .20** |
| ACAD4 \( \rightarrow \) EXT5 | - | - |
| ACAD5 \( \rightarrow \) SOS6 | - | .12** |
| ACAD5 \( \rightarrow \) EXT6 | - | - |
| ACAD6 \( \rightarrow \) SOS7 | .13** | .17** |
| ACAD6 \( \rightarrow \) EXT7 | - | - |
| ACAD5 \( \rightarrow \) ACAD7 | .38** | .41** |

Note. ACAD = academic performance; SOS = social skills; EXT = externalizing; \( \beta \) = standardized path coefficients; all regressions coefficients are \( p \leq .05 \), if nothing else is written. Non-significant regression coefficients are not included in the table.
data from previous research, the outcomes indicate that gender differences in academic performance are present in the early stages of schooling, and increase during primary school, particularly in Grade 6 (Borgonovi et al., 2018; DiPrete & Jennings, 2012; Entwisle et al., 2007). Regarding the stability of the gender gap through primary school, the results indicate that the gender differences in fourth grade favoring girls increase only moderately over the next three grades (Arnot et al., 1998). The gender differences in teacher ratings of social skills and externalizing behavior, on the other hand, tend to decrease over time due to improvements in boys’ behavior.

This is the first study to examine longitudinal associations and cross-lagged effects among primary school students in Norway. Contrary to expectations, social skills turned out to be poor predictors of children’s subsequent academic performance (DiPrete & Jennings, 2012; Malecki & Elliott, 2002). More in line with previous research was the finding that externalizing behavior ratings by teachers predicted lower ratings of academic performance and social skills, especially among boys. But the predictions were more accurate in the externalizing-driven model than in the full model (Tables 6 and 7).

Even if the current study supports the finding that promotion of academic competence may reduce the risk of externalizing behavior (Hawkins et al., 1999), it also supports the finding that externalizing behavior problems may interfere with, or undermine academic achievement, especially for boys (Kristoffersen et al., 2015; Masten et al., 2005). The negative effects of externalizing behavior on academic performance and social skills were supported by the externalizing-driven model, particularly for boys in the upper grades of primary school. But when paths from social skills and academic achievement were included into the model, the effects from externalizing on academic disappeared, indicating that the unique effects of externalizing behavior are weak (Table 7).

There is considerable stability in all three constructs examined, but hardly any evidence of influences of teacher assessments in fourth grade on outcomes in seventh grade. The best predictors are usually the same variable measured at the previous grade level. An exception to this was academic performance in fourth grade which showed strong cross-lagged effects on academic performance of both genders in seventh grade. That implies that there is considerable stability in academics, while social skills and externalizing behavior seem to be more malleable through environmental influences. Taken together and as visualized in Figure 1, the outcomes support the assumption of mutual influences among academic achievement, social skills, and externalizing behavior. Even if there is more support for the externalizing-driven and academic-driven models than for the social skills-driven model, the strength of the predictors are modest and there are clear indications of mutual influences.

A limitation of the study is the use of only one informant group, which may cause method overlap and the overestimation of relationships and predictions. This may particularly have influenced the associations found between academic achievement, social skills, and externalizing behavior. Nevertheless, teacher ratings of academic achievements were better predictors of social skills than the other way around. This may explain why teachers seem to be more concerned with academic performance, and that compared with test results, teacher assessment and grading practice take into consideration the students’ learning relevant behavior and social skills. Teachers’ grading practices are sometimes criticized for being biased, but it is not clear from the literature how much teachers’ grades may be influenced by student behavior. It could be that they reward learning relevant social skills with higher grades and punish externalizing problems in the classroom with lower grades (Farrington et al., 2012). The gender gap in teacher assessments and grades may therefore be influenced by other factors than subject-relevant knowledge and skills.

The gender differences probably occur at an even earlier stage than demonstrated in this study. According to Programme for International Student Assessment (PISA, 2015, cited in Borgonovi et al., 2018), the gender gap increases substantially from around age 10 (fourth grade) to age 15, after which the differences decrease. The gender gap in teacher ratings of academic achievement found in this study matches those found in international studies, but the differences tend to be smaller (mean \(d = 0.15\) in fourth grade as compared with \(d = 0.5\) in the national OECD PIRLS study, 2016). Cognitive competence as measured by academic achievement seems to be rather stable throughout primary school, in comparison with non-cognitive competence which seems to be more malleable over time. The gender differences in teacher ratings of social skills and externalizing behavior seem to decrease earlier than gender differences in academic performance, and may be the first signal of leveling out the gender gap in academic performance. The larger and more changeable gender differences in non-cognitive factors may indicate that these factors should be prioritized in efforts to reduce unintended gender differences and to increase academic achievement in both girls and boys. On the other hand, this study demonstrates that teacher assessments of academic performance predicted students’ future social and behavioral skills better than the other way around, a finding which indicates that from their perspective, academic competence is more basic than social and behavioral competence. Still, the influences across the academic, social, and behavioral domains also point to the importance of promoting several forms of competence to increase the probability of school success for both boys and girls.
Declaration of Conflicting Interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD
Terje Ogden https://orcid.org/0000-0002-2353-0798

References
Achenbach, T. M. (1991). Manual for the teacher’s report form and 1991 profile. Univ Vermont/Department Psychiatry.
Arnot, M., Gray, J., James, M., & Rudduck, J. (1998). Recent research on gender and educational performance. Office for Standards in Education.
Bjorklund-Young, A. V., & Ronda, V. (2017). The multidimensionality of teacher quality: Teaching skills and students’ noncognitive skills (Working paper, John Hopkins University).
Bleses, D., Vach, W., Slott, M., Webberg, S., Thomsen, P., Madsen, T. O., & Basboll, H. (2008). The Danish Communicative Developmental Inventories: Validity and main developmental trends. Journal of Child Language, 35(3), 651–669.
Borgonovi, F., Ferrara, A., & Maghnouj, S. (2018). The gender gap in educational outcomes in Norway (OECD Education working paper no. 183).
Breslau, J., Miller, E., Joanie Chung, W. J., & Schweitzer, J. B. (2011). Childhood and adolescent onset psychiatric disorders, substance use, and failure to graduate high school on time. Journal of Psychiatric Research, 45(3), 295–301.
Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. Structural Equation Modeling, 9, 233–255.
Cohen, J. (1988). Statistical power analysis for the behavioural sciences. Erlbaum.
Cornwell, C., Mustard, D. B., & Van Parys, J. (2013). Noncognitive skills and the gender disparities in test scores and teacher assessments: Evidence from primary school. Journal of Human Resources, 48(1), 236–264.
DiPrete, T. A., & Jennings, J. L. (2012). Social and behavioral skills and the gender gap in early educational achievement. Social Science Research, 41(1), 1–15. https://doi.org/10.1016/j.ssresearch.2011.09.001
Domitrovich, C. E., Durlak, J. A., Staley, K. C., & Weissberg, R. P. (2017). Social-emotional competence: An essential factor for promoting positive adjustment and reducing risk in school children. Child Development, 88(2), 408–416. https://doi.org/10.1111/cdev.12739
Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L. S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K., & Japel, C. (2007). School readiness and later achievement. Developmental Psychology, 43, 1428–1446.
Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students’ social and emotional learning: A meta-analysis of school-based universal interventions. Child Development, 82(1), 405–432.
Enders, C. K. (2010). Applied missing data analysis. Guiford Press.
Entwisle, D. R., Alexander, K. L., & Olson, L. S. (2007). Early schooling: The handicap of being poor and male. Sociology of Education, 80, 114–138.
Farmer, A. D., Jr., Bierman, K. L., & Conduct Problems Prevention Research Group. (2002). Predictors and consequences of aggressive-withdrawn problem profiles in early grade school. Journal of Clinical Child and Adolescent Psychology, 31(3), 299–311. https://doi.org/10.1207/S15374424JCCP3103_02
Farrington, C. A., Roderick, M., Allensworth, E., Nagaoka, J., Keyes, T. S., Johnson, D. W., & Beechum, N. O. (2012). Teaching adolescents to become learners. The role of noncognitive factors in shaping school performance: A critical literature review. Chicago, University of Chicago Consortium on Chicago School Research.
Gresham, F. M., & Elliott, S. N. (1990). Social skills rating system: Manual. American Guidance Service.
Gustavsen, A. M. (2017). Longitudinal relationship between social skills and academic achievement in a gender perspective. Cogent Education, 4, Article 1411035, https://doi.org/10.1080/2331186x.2017.1411035
Gutman, L. M., & Schoon, I. (2013). The impact of non-cognitive skills on outcomes for young people. Education Endowment Foundation, 59 (Institute of Education).
Halonen, A., Aunola, K., Ahonen, T., & Nurmi, J. E. (2006). The role of learning to read in the development of problem behaviour: A cross-lagged longitudinal study. British Journal of Educational Psychology, 76(3), 517–534.
Hawkins, J. D., Catalano, R. F., Kosterman, R., Abbott, R., & Hill, K. G. (1999). Preventing adolescent health-risk behaviors by strengthening protection during childhood. Archives of Pediatrics & Adolescent Medicine, 153, 226–234.
Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal, 6, 1–55.
Jacob, B. A. (2002). Where the boys aren’t: Noncognitive skills, returns to school and the gender gap in higher education. Economics of Education Review, 21, 589–598.
Jennings, J. L., & DiPrete, T. A. (2010). Teacher effects on social and behavioral skills in early elementary school. Sociology of Education, 83, 135–159.
Kristoffersen, J. H. G., Obel, C., & Smith, N. (2015). Gender differences in behavioral problems and school outcomes. Journal of Economic Behavior & Organization, 115, 75–93. https://doi.org/10.1016/j.jebo.2014.10.006
Ladd, G. W., & Burgess, K. B. (2001). Do relational risks and protective factors moderate the linkages between childhood...
aggression and early psychological and school adjustment? *Child Development, 72*(5), 1579–1601.

Malecki, C. K., & Elliott, S. N. (2002). Children’s social behaviors as predictors of academic achievement: A longitudinal analysis. *School Psychology Quarterly, 17*, 1–23.

Masten, A. S., & Cicchetti, D. (2010). Developmental cascades. *Development and Psychopathology, 22*, 491–495.

Masten, A. S., Roisman, G. I., Long, J. D., Burt, K. B., Obradović, J., Riley, J. R., Boelck-Stennes, K., & Tellegen, A. (2005). Developmental cascades: Linking academic achievement and externalizing and internalizing symptoms over 20 years. *Developmental Psychology, 41*, 733–746.

McLeod, J. D., & Fettes, D. L. (2007). Trajectories of failure: The educational careers of children with mental health problems. *American Journal of Sociology (AJS), 113*(3), 653–701.

McLeod, J. D., & Kaiser, K. (2004). Childhood emotional and behavioral problems and educational attainment. *American Sociological Review, 69*(5), 636–658.

Muthén, B. O., & Muthén, L. (2012). *Mplus computer program version 7.3*. Mplus.

Organisation for Economic Co-Operation and Development. (2015). *Skills for social progress: The power of social and emotional skills* (OECD skills studies).

Roth, B., Becker, N., Romeyke, S., Schäfer, S., Domnick, F., & Spinath, F. M. (2015). Intelligence and school grades. A meta-analysis. *Intelligence, 53*, 118–137.

Skogen, J. C., & Torvik, F. A. (2013). *Aftersforstyrrelser blant barn og unge i Norge: Beregnet forekomst og bruk av hjel-
petiltak* [Behavioral disorders among children and young people in (NAME OF COUNTRY): Estimated incidence and use of support measures] (FHI report 2013:4). https://www.fhi.no/publ/2013/aftersforstyrrelser-blant-barn-og/.

Sørlie, M.-A., Hagen, K. A., & Nordahl, K. B. (2020). Development of social skills during middle childhood: Growth trajectories and school-related predictors. *International Journal of School and Educational Psychology*. Advance online publication. https://doi.org/10.1080/21683603.2020.1744492

Sung, Y. Y., & Chang, M. (2010). Which social skills predict academic performance of elementary school students. *Journal on Educational Psychology, 3*, 23–34.

U.S. Department of Education. (2006). *Condition of education* (Technical report).

Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin, 140*, 1174–1204.

Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review, 14*(5), 779–804.

Wentzel, K. R. (1993). Does being good make the grade? Social behavior and academic competence in middle school. *Journal of Educational Psychology, 85*(2), 357–364.

Wollscheid, S., Hjetland, H. N., Rogde, K., & og Skjelbred, S.-E. (2018). *Årsaker til og tiltak mot kjønnsforskjeller i skoleprestasjoner: En kunnskapsoversikt* [Causes and measures against gender differences in school performance. A knowledge review] (NIFU-report 2018:25). https://www.nifu.no/publications/1619889/