Running head: TEACHER–STUDENT GENDER MATCH

Are boys better off with male and girls with female teachers?

A multilevel investigation of measurement invariance and gender match in teacher-student relationship quality

Jantine L. Spilt

VU University Amsterdam

Helma M. Y. Koomen

University of Amsterdam

Suzanne Jak

University of Amsterdam

Correspondence to: Jantine Spilt, Department of Developmental Psychology, VU University Amsterdam, Van der Boechorststraat 1, 1081 BT Amsterdam, The Netherlands. Email: jl.spilt@psy.vu.nl

This is an Author's Accepted Manuscript of an article published in Journal of School Psychology, 2012, 50, 363-378. © Elsevier, available online at http://dx.doi.org/10.1016/j.jsp.2011.12.002

This is an Author's Accepted Manuscript of an article published in Journal of School Psychology, 2012, 50, 363-378. © Elsevier, available online at http://dx.doi.org/10.1016/j.jsp.2011.12.002

Abstract

Although research consistently points to poorer teacher–student relationships for boys than girls, there are no studies that take into account the effects of teacher gender and control for possible measurement non-invariance across student and teacher gender. This study addressed both issues. The sample included 649 primary school teachers (182 men) and 1493 students (685 boys). Teachers completed a slightly adapted version of the Student–Teacher Relationship Scale. The results indicated limited measurement non-invariance in teacher reports. Female teachers reported better (i.e., more close, less conflictual, and less dependent) relationships with students than male teachers. In addition, both male and female teachers reported more conflictual relationships with boys than with girls, and female teachers also reported less close relationships with boys than with girls. The findings challenge society's presumption that male teachers have better relationships with boys than women teachers.

Keywords: teacher–student relationship quality, student gender, teacher gender, gender match, measurement invariance

This is an Author's Accepted Manuscript of an article published in Journal of School Psychology, 2012, 50, 363-378. © Elsevier, available online at http://dx.doi.org/10.1016/j.jsp.2011.12.002

Are boys better off with male and girls with female teachers?

A multilevel investigation of measurement invariance and gender match in teacher–student relationship quality

Affective relationships between students and teachers are a key factor in students' school adjustment. Research demonstrates pervasive effects of relationship quality on students' academic achievement for both boys and girls (e.g., Hughes, Luo, Kwok, & Loyd, 2008; Stipek & Miles, 2008). When mean levels of relationship quality are examined, research consistently indicates that teachers have poorer relationships with boys than girls (Baker, 2006; Hamre & Pianta, 2001; O'Connor, 2010). Importantly though, current evidence is limited to student gender only as a factor in teacher–student relationships. To our knowledge, the possible additional effects of teacher gender or teacher–student gender match have not yet been studied, although people wonder whether boys perhaps would do better with male teachers. This omission is probably due to the fact that most research samples include too few male teachers. The current paper addresses this gap in the literature by studying the combined effects of student and teacher gender in a Dutch sample of primary school teachers with 28% males. However, before gender effects on teacher–student relationship quality can be appropriately studied, equivalence in the measurement of teacher–student relationship quality across gender should be established first. For valid group comparisons and the detection of "true" differences, it is crucial that largely similar constructs are measured that are not biased by either student gender or teacher gender or both. This study aimed to examine differences in male versus female teachers' relationships with boys versus girls, while controlling for possible measurement non-invariance.

Teacher–Student Relationships and Student Gender

This is an Author's Accepted Manuscript of an article published in Journal of School Psychology, 2012, 50, 363-378. © Elsevier, available online at http://dx.doi.org/10.1016/j.jsp.2011.12.002

The way teachers relate to individual students is a fundamental aspect of teaching. Guided by ecological models of development, teacher–student relationships have increasingly gained attention as proximal processes that influence students' learning experiences and achievement. Students with positive relationships are better adjusted and more engaged in learning activities (Hughes et al., 2008; Thijs & Koomen, 2008), whereas students with negative relationships are at risk of a wide array of adjustment problems and academic underachievement (Ladd, Birch, & Buhs, 1999; Spilt, Hughes, Wu, & Kwok, in press). Teacher–student relationship quality is typically measured by teacher report on the dimensions closeness and conflict and sometimes dependency (Ang, 2005; Koomen, Verschueren, van Schooten, Jak, & Pianta, 2011; Pianta, 2001). Close relationships are characterized by warmth, trust, and open communication. In line with the extended attachment perspective and self-system theories of motivation, warm and open relationships have been shown to foster students' motivation for learning and positive task behaviors, and are predictive of gains in academic achievement (Furrer & Skinner, 2003; Hughes, et al., 2008; Roorda, Koomen, Spilt, & Oort, 2011). Conversely, conflictual relationships are characterized by discordance and mistrust, and can seriously undermine students' school career (Hamre & Pianta, 2001; O'Connor & McCartney, 2007; Stipek & Miles, 2008). Especially behaviorally-challenging students tend to have poor relationships with teachers, which exacerbates the risk of school failure (Doumen et al., 2008; Ladd & Burgess, 2001; Stipek & Miles, 2008). The extant literature also shows that teacher–student relationships can buffer students against risks associated with behavior problems and poor self regulation skills. Close relationships are associated with decreases in behavior problems and increases in positive behaviors (Meehan, Hughes, & Cavell, 2003; Silver, Measelle, Armstrong, & Esgender, 2005).

This is an Author's Accepted Manuscript of an article published in Journal of School Psychology, 2012, 50, 363-378. © Elsevier, available online at http://dx.doi.org/10.1016/j.jsp.2011.12.002

The vulnerability hypothesis posits that at-risk students are more susceptible to contextual influences than non-risk students. Likewise, boys are believed to be more influenced by the quality of relationships with teachers than girls because boys are more likely to have socio-behavioral and academic difficulties throughout primary school (Coley, 2001; Hamre & Pianta, 2001; Matthews, Ponitz, & Morrison, 2009; Pomerantz, Altermatt, & Saxon, 2002). Thus, compared to girls, boys may be more hindered by poor relationships with teachers but are also believed to profit most from supportive relationships. However, primary school teachers consistently report poorer relationships with boys than girls (Baker, 2006; Hamre & Pianta, 2001; O'Connor, 2010). For instance, recent longitudinal research indicates that boys are more likely to have chronically-high levels of conflict with their teachers throughout elementary school than girls (Spilt et al., in press). In addition, studies using self-reports of elementary school students also indicate more conflict and less closeness for boys (Koepke & Harkins, 2008; Koomen & Jellesma, 2011; Skinner & Belmont, 1993).

The poorer relationship quality of boys with school teachers is not an isolated phenomenon, but part of a broader picture of a gender imbalance in schooling. Boys do far worse than girls academically and behaviorally across the entire school career and concerns about this gender imbalance are rising (e.g., Entwisle, Alexander, & Olson, 2007). To advance the understanding of gender inequality, it is crucial to examine gender differences in different domains of schooling and in basic qualities of students' school experiences including their relationships with teachers.

Theoretical Notions about Gender Differences

The effects of student gender on teacher–student relationship quality are typically attributed to differences in biological traits across gender on the one hand and sociological factors

This is an Author's Accepted Manuscript of an article published in Journal of School Psychology, 2012, 50, 363-378. © Elsevier, available online at http://dx.doi.org/10.1016/j.jsp.2011.12.002

on the other. Boys are considered to be biologically disposed to more assertive, energetic, and aggressive behavior than girls, and therefore, they would be expected to have more conflictual relationships with teachers (Carlson, Tamm, & Gaub, 1997; Zahn-Waxler, Shirtcliff, & Marceau, 2008). According to gender socialization theory, gender-typical behaviors and attitudes are reinforced in students through differential treatment of boys and girls (for review see Koch, 2003). Teachers tend to be more tolerant of dominant and autonomous behaviors in boys, whereas they respond more favorably to calm and submissive behaviors in girls (Basow, 2004; Meece & Scantlebury, 2006). In addition, affiliation seeking behavior and emotional disclosure may be more reinforced in girls than boys (cf. Zahn-Waxler et al., 2008). Thus, girls may seek more nurturing relationships with teachers, whereas boys are more inclined to present themselves as autonomous, resulting in less close teacher–student relationships for boys in comparison to girls (Ewing, 2009; Ewing & Taylor, 2009).

In addition, based on gender schema theory (Bem, 1981), it is believed that boys would more easily identify with male teachers and girls with female teachers, suggesting positive effects of a same-gender teacher. Male teachers could present boys with more well-adjusted forms of masculinity and as a result promote more positive behavioral attitudes toward school. Therefore, it is widely believed that male teachers could inspire and handle boys better. Although this latter idea has been decidedly criticized as an oversimplification of gender theory (Cushman 2010; Skelton, 2003), it is a common assumption that boys do worse in primary school because the workforce is predominantly female. Policy makers worldwide have expressed concerns about the feminization of education (cf. Cushman, 2010; Driessen, 2007; Francis et al., 2006). They advocate the need to increase the number of male teachers in primary schools to counteract the educational delay of boys (for an overview, see Cushman, 2007). However, the empirical

This is an Author's Accepted Manuscript of an article published in Journal of School Psychology, 2012, 50, 363-378. © Elsevier, available online at http://dx.doi.org/10.1016/j.jsp.2011.12.002

evidence does not bolster this argument. Instead, research findings argue strongly against the presumed beneficial effects of same-gender teachers on students' behavioral and cognitive competencies; for both boys and girls, favorable effects appear trivial (Driessen, 2007; Marsh, Martin, & Cheng, 2008; Martin & Marsh, 2005; Neugebauer, Helbig, & Landmann, 2010). There is even some evidence that both male and female students have more positive social and academic self-views in classes of female than male teachers (Hopf & Hatzichristou, 1999).

Taken together, research suggests that the impact of same-gender teachers on children's behavioral and cognitive development is largely negligible, but it is possible that there is a positive effect of same-gender teachers on children's interpersonal functioning. It is conceivable that male teachers may have better interpersonal relationships with boys than female teachers. If, on the other hand, it is found that boys have worse relationships with both male and female teachers than girls, this finding could advance our understanding of factors that add to the educational delay of boys. To our knowledge, Ewing's dissertation study (2009) was the first to explore the combined effects of both teacher and student gender on teacher–student relationship qualities. This study indicates that girls have closer relationships with female teachers than male teachers. Boys, however, were not rated higher than girls on closeness by male teachers (Ewing, 2009). Thus, gender match may be important for female students only, possibly because female teachers and female students are both socialized to cultivate nurturing relationships. Relationships with boys were rated as more conflictual by both male and female teachers. However, before such findings can be interpreted, it is pivotal to verify that the measurement of the constructs is invariant across teacher and student gender (Mellenbergh, 1989; Meredith, 1993). To our knowledge, no studies have yet tested measurement invariance of teacher–student relationship quality across both teacher and student gender.

This is an Author's Accepted Manuscript of an article published in Journal of School Psychology, 2012, 50, 363-378. © Elsevier, available online at http://dx.doi.org/10.1016/j.jsp.2011.12.002

Measurement Invariance across Gender in Teacher Questionnaires

Different interpretations of questionnaire items between male and female teachers for boys versus girls may cause single items to be stronger indicators of the latent construct in one group than the other (metric non-invariance) or could produce artificially inflated scores (scalar non-invariance). Gender differences then are due to unequal measurement between groups and cannot be attributed to true differences in the underlying trait (Mellenbergh, 1989; Meredith, 1993).

Although research typically relies on teacher reports of teacher–student relationship quality in early to middle childhood, the vast majority of empirical investigations neglect the issue of measurement invariance in teacher questionnaires. Recently, Koomen et al. (2011) studied measurement invariance of teacher–student relationship quality across student gender in a large sample of students (ages 3-12). The results indicated only minor measurement non-invariance. Similar constructs were measured for boys and girls (demonstrating metric invariance), but constructs were measured on slightly different scales (demonstrating scalar non-invariance). Similar findings of minor measurement non-invariance across student gender have been detected in teacher reports of student behavior scales including the Preschool Behavior Questionnaire (Spilt, Koomen, Thijs, Stoel, & van der Leij, 2010; Tremblay, Desmarais-Gervais, Gagnon, & Charlebois, 1987), the Conners Teacher Rating Scale-Revised Short Version (Derks, Dolan, Hudziak, Neale, & Boomsma, 2007), and the Strengths and Difficulties Questionnaire (d'Acremont & Van der Linden, 2008). No research has yet explored measurement invariance across teacher gender in teacher-report measures of teacher–student relationship quality.

Current Study

This is an Author's Accepted Manuscript of an article published in Journal of School Psychology, 2012, 50, 363-378. © Elsevier, available online at http://dx.doi.org/10.1016/j.jsp.2011.12.002

Although it is consistently shown that teachers report poorer relationships with boys than with girls in primary education, this result is difficult to interpret for two reasons. First, findings are primarily based on reports from female teachers, whereas the effects of teacher gender are virtually unknown. Second, we cannot draw firm conclusions on gender issues as long as the prerequisite of measurement invariance across student and across teacher gender is not properly studied. The current study therefore examined the combined effects of student and teacher gender. The key question was: Do female and male teachers report mean differences in teacher–student relationship qualities for boys and girls when possible measurement non-invariance across teacher- and student gender is accounted for?

This study presents follow-up research of Koomen and colleagues (2011) by investigating measurement non-invariance across both student and teacher gender (28% male teachers). This study includes a subsample of primary school students and teachers of the original sample used by Koomen et al. Unlike Koomen et al., we will investigate measurement invariance, while explicitly modeling the multilevel structure in the data. This modeling of the multilevel structure is important because about 15% to 33% of the variance in teacher reports of relationship quality with individual students can be attributed to differences between teachers and not to differences between individual students (Mashburn, Hamre, Downer, & Pianta, 2006). It is thus possible that two students with identical observed test scores but rated by different teachers actually differ in scores on the underlying construct or trait. If teacher reports do not measure the same construct across clusters (i.e., between teachers), this would indicate the presence of *cluster bias* (Jak, Oort, & Dolan, 2011b). Findings of cluster bias would indicate that at least a part of the variance in teacher reports of relationship quality can be attributed to teacher characteristics (e.g., teacher gender). Conversely, relationship quality on the student level reflects primarily child-driven

This is an Author's Accepted Manuscript of an article published in Journal of School Psychology, 2012, 50, 363-378. © Elsevier, available online at http://dx.doi.org/10.1016/j.jsp.2011.12.002

aspects of relationship quality that can be explained by differences between children (e.g., student gender or child problem behavior).

Using multilevel modeling has the consequence that on the student level, metric invariance is assumed whereas scalar invariance can be tested. However, Koomen et al. (2011), already established metric invariance for the items of the adapted Student–Teacher Relationship Scale across student gender, albeit in a single-level analysis. On the teacher level, both metric invariance and scalar invariance is directly tested in the current study.

Based on different theoretical perspectives, partly competing hypotheses were stated. In line with gender socialization theory and the widely-assumed feminization of education (i.e., growing congruence between school environment and female stereotype), we predicted poorer relationships for boys than girls (i.e., less closeness and more conflict). No hypothesis was stated with respect to dependency. Drawing from gender schema theory, we expected positive effects of gender match on teacher–student relationships. That is, boys were anticipated to have better relationships with male teachers than with female teachers, and vice versa for girls.

Method

Sample

The sample included 649 primary school teachers (182 men) from 92 regular elementary schools in the Netherlands (grades 1 to 6). Male teachers reported on 242 boys and 227 girls; female teachers reported on 463 boys and 561 girls (N = 1,493). Twenty-one cases (7 teachers) were excluded because teacher gender was unknown. Students were equally distributed across grades 1 to 6.

Male teachers had significantly more years of experience in education (M = 19.6; SD = 12.24) than female teachers, M = 13.2; SD = 10.16; $F(1,633) = 45.390$, $p < .001$. Male teachers

This is an Author's Accepted Manuscript of an article published in Journal of School Psychology, 2012, 50, 363-378. © Elsevier, available online at http://dx.doi.org/10.1016/j.jsp.2011.12.002

were underrepresented in the lower elementary grades (8%, 14%, 23%, 33%, 49%, and 49% of male teachers in grades 1 to 6, respectively). A fairly similar male–female distribution across grades and difference in experience between male and female teachers in elementary schools has been found in another Dutch nationally representative sample (N = 60,000; Driessen, 2007). Students were on average 9.7 years old (SD = 1.86) and 94.1% were of Dutch origin. Students had spent at least 6 weeks in the classroom of the teachers who reported on their relationship quality.

The current sample was drawn from a larger Dutch sample (N = 2,235) of 3- to 12-year old students that was selected to validate an adapted version of the Student–Teacher Relationship Scale (Koomen et al., 2011). The larger sample included random samples of schools that were drawn from lists of all schools for each geographical region in the Netherlands and contacted by letter, phone, or both methods. Participation rate was 20% at the school level. The distribution across geographical regions (North, East, West, and South) and school types (regular versus special elementary schools) was fairly similar to that of the general Dutch youth population (Koomen, Verschueren, & Pianta, 2007). In participating elementary schools, children were randomly selected from student registers per class and balanced across student gender, age groups, and teachers. About 65% of the parents gave written permission to participate, resulting in an average participation rate of two to three students from each grade per school (for more details about the sample, see Koomen et al., 2011).

Questionnaires

Teacher–Student Relationship. An authorized Dutch translated and slightly adapted version of the Student–Teacher Relationship Scale (STRS; Koomen et al., 2007; Koomen et al., 2011;

This is an Author's Accepted Manuscript of an article published in Journal of School Psychology, 2012, 50, 363-378. © Elsevier, available online at http://dx.doi.org/10.1016/j.jsp.2011.12.002

Pianta, 2001) was employed to measure teacher perceptions of the affective relationship with individual students. Three constructs are measured: Closeness, Conflict, and Dependency. Closeness refers to the degree of warmth and open communication; Conflict refers to negative and coercive teacher–student interactions, and Dependency refers to overly clingy student behavior. Teachers rated items on a 5-point scale ranging from 1 (*definitely does not apply*) to 5 (*definitely does apply*) for each individual child participating in the study (see Table 2 for items).

In comparison to the original scales, the adapted Conflict and Closeness scales show minor alterations in the form of, respectively, removal of one and substitution of two items (see Table 2). The adapted Dependency scale was altered more thoroughly (e.g., 2 of the 6 items were newly developed) because of its mediocre internal consistency in previous research (Cronbach's α ranging from .40-.74; e.g., Doumen et al., 2009; Ewing & Taylor, 2009; Hamre & Pianta, 2001; Rey, Smith, Yoon, Soomers, & Barnett, 2007). Cronbach's alpha coefficients have ranged from .88-.93 for Closeness, .88-.91 for Conflict, and .75-.82 for Dependency in the adapted Student–Teacher Relationship Scale (Doumen, Koomen, Buyse, Wouters, & Verschueren, in press; Koomen et al., 2007; Koomen et al., 2011). In the present study, Cronbach's alpha coefficients were .90, .91, and .74 for respectively Closeness, Conflict, and Dependency. Previous research has also provided evidence for validity by confirming the assumed three-dimensional structure of the adapted Student–Teacher Relationship Scale (Doumen et al., in press; Koomen et al., 2007; Koomen et al., 2011). In addition, these studies have reported significant associations with observer-rated teacher–child interactions, with measures of problem behavior and prosocial behavior (via teacher and parents reports), with school engagement (via teacher and observer reports), and with teaching stress.

Analyses

This is an Author's Accepted Manuscript of an article published in Journal of School Psychology, 2012, 50, 363-378. © Elsevier, available online at http://dx.doi.org/10.1016/j.jsp.2011.12.002

Multilevel structural equation modeling (MSEM[1]) was used to test for measurement invariance with respect to Teacher Gender and Student Gender. MSEM enables researchers to fit models to the covariance matrix of within teacher differences (within-clusters level) and between teacher differences (between-clusters level). We used robust maximum likelihood estimation (MLR) in Mplus (Muthén & Muthén, 2007) to obtain parameter estimates. This estimation method provides a test statistic that is asymptotically equivalent to the Yuan-Bentler test statistic (Yuan & Bentler, 2000) and standard errors that are robust for non-normality. A correction factor for the chi-squares is used to calculate chi-square differences between nested models (Satorra & Bentler, 2001). In addition to the adjusted χ^2 statistic, the root mean squared error of approximation (RMSEA) and the comparative fit index (CFI) were used as measures of overall goodness-of-fit. RMSEA values smaller than .08 are satisfactory (Browne & Cudeck, 1992); values smaller than .05 indicate close fit, CFI values over .90 indicate reasonably good fit (Hu & Bentler, 1999).

We used restricted factor analysis (RFA; Oort, 1992, 1998) to investigate measurement invariance with respect to Student Gender. With the RFA method, Student Gender is entered as a covariate, which is correlated with the common factor but not with item scores. Direct effects of Student Gender on one or more item scores correspond to scalar non-invariance. Direct effects were added when the modification index was significant at a Bonferroni corrected level of significance (α = .05 /number of possible effects) and the standardized parameter change was larger than .10. In addition, the content of these items was reviewed. Decisions to add direct effects on item scores were thus based on both empirical and theoretical grounds. The RFA

[1] Results from a two-level model are reported (students nested in teachers) because school-level variance was very low (ICCs .02-.07). Moreover, virtually the same results were found in the two-level model as in the three-level model.

This is an Author's Accepted Manuscript of an article published in Journal of School Psychology, 2012, 50, 363-378. © Elsevier, available online at http://dx.doi.org/10.1016/j.jsp.2011.12.002

method allows for a direct test of scalar invariance (equality of intercepts) but not metric invariance (equality of factor loadings). To our knowledge, a method to explicitly evaluate metric invariance at the within-clusters level (i.e., student level) has not yet been evaluated in multilevel models. Multigroup factor analysis is not applicable in multilevel models when the interest is in a level 1 variable (such as Student Gender), because it is not possible to analyze boys and girls separately without accounting for the multilevel structure (i.e., some boys and girls share the same teacher).

Multigroup factor analysis (MGFA) was used to test measurement invariance (both scalar and metric invariance) across Teacher Gender. Measurement invariance is tested by the equality of measurement parameters across groups. We started with the relevant parameters (i.e., intercepts and factor loadings) equal over Gender, and freed parameters based on modification indices. Modification indices were again evaluated at a Bonferroni corrected level of significance (α = .05 /number of constrained intercepts or .05 /number of constrained factor loadings).

Modeling Procedure. The following steps (based on Jak, Oort, & Dolan, 2011a) were taken for each relationship dimension separately. In order to decide whether multilevel modeling is necessary, intraclass correlations (ICCs) were calculated in Step 1, and the significance of between-clusters level variance and covariance was tested. In Step 2, the measurement model for within-clusters level variance was established. In Step 3, scalar measurement invariance across Student Gender was investigated at the within-clusters level. In Step 4, cluster bias (see Jak et al., 2011b) was examined to determine whether the same construct was measured across teachers. In Step 5, measurement invariance across Teacher Gender was tested at the between-clusters level. In Step 6, latent mean differences between male and female teachers in relationship constructs were tested, and their relationships with male versus female students were compared. In case of

This is an Author's Accepted Manuscript of an article published in Journal of School Psychology, 2012, 50, 363-378. © Elsevier, available online at http://dx.doi.org/10.1016/j.jsp.2011.12.002

possible measurement non-invariance, latent mean differences were tested under partial measurement invariance, that is: the intercepts and/or factor loadings of non-invariant items are freely estimated so that the latent means are not affected by the bias (Wicherts & Dolan, 2010; Byrne, Shavelson & Muthén, 1989).

Step 1 requires calculation of the ICC for each item. These ICCs are calculated as the ratio of between-clusters level variance over the total variance. The null model (specifying no variance at the between-clusters level) and independence model (specifying no covariance between the items at the between-clusters level) are fitted at the between-clusters level. Fitting the null and independence model is done with an unrestricted model (all estimable parameters are free) at the within-clusters level, so that misfit stems from the between-clusters part of the model only. In Step 2, a one-factor solution is fitted and evaluated. In Step 3, the effects of Student Gender on the item scores are evaluated (i.e., scalar invariance). Step 2 and 3 are to be performed with an unrestricted model at the between-clusters level so that misfit stems from the within-clusters part of the model only. Step 4 requires fitting the same model at the between-clusters level as at the within-clusters level, namely the measurement model from Step 2. In order to test for cluster bias (Step 4), the factor loadings are constrained to be equal across levels, and the residual variance at the between-clusters level is fixed at 0 (Jak et al., 2011b). These constraints ensure that differences between the cluster means are exclusively attributable to differences in the common factor means. Tenability of these constraints can be tested with chi-square difference tests. Cluster bias shows up as residual variance at the between-clusters level and indicates that factors other than the common factor are causing differences between the teachers' mean scores on the items. Only if cluster bias is present, Step 5 will be completed to investigate measurement non-invariance with respect to specific between level variables (Teacher Gender), while

This is an Author's Accepted Manuscript of an article published in Journal of School Psychology, 2012, 50, 363-378. © Elsevier, available online at http://dx.doi.org/10.1016/j.jsp.2011.12.002

controlling for the aggregated level 1 variables (i.e., proportion of boys in the class room[2]). If measurement non-invariance is absent, or accounted for, we can test the substantive hypotheses (are there differences between male and female teachers in their Closeness, Conflict, or Dependency with boys versus girls?).

For scaling the latent variables, the factor variance is fixed at 1 in Step 2 and 3. From Step 4, the variance of the between-clusters level common factor is identified by the equality constraints on the factor loadings and can be freely estimated. In Step 5 and 6, the within-clusters factor variance for the male teachers is fixed at 1, but for female teachers it is freely estimated.

The multilevel factor model of Closeness is presented in Figure 1 to illustrate the followed steps.

Results

The observed means and standard deviations for Closeness, Conflict, and Dependency are reported in Table 1. Table 2 presents the items and unstandardized factor loadings of the final models.

Closeness. The intraclass correlations (ICCs) for the Closeness items varied between .17 (for Item 1) and .28 (for Item 4). The between-clusters level variance and covariance was significant, indicated by a significant $\Delta\chi^2$ for the null model, $\Delta\chi^2$ (66) = 702.16, $p < .001$, and for the independence model, $\Delta\chi^2$ (55) = 178.35, $p < .001$, compared with a model with covariances on the within-clusters and between-clusters levels. We therefore concluded that multilevel modeling was necessary.

[2] Although proportion of boys in the class room was included, results concerning the proportion of boys are not reported, because this variable had no significant variance (i.e. the proportion of boys could be assumed equal across clusters).

This is an Author's Accepted Manuscript of an article published in Journal of School Psychology, 2012, 50, 363-378. © Elsevier, available online at http://dx.doi.org/10.1016/j.jsp.2011.12.002

A one-factor model was fitted to the within-clusters level covariance matrix (in Step 2). The model showed an adequate fit, χ^2 (44) = 111.15, $p < .001$, RMSEA = .032, and CFI = .99. Adding Student Gender as a covariate (in Step 3) in the model resulted in a well fitting model, χ^2 (54) = 174.91, $p < .001$, RMSEA = .039, and CFI = .98. However, modification indices suggested direct effects of Student Gender on Item 3 and Item 4. Adding these direct effects significantly improved model fit, $\Delta\chi^2 = 34.96$, $\Delta df = 2$, $p < .001$. The standardized direct effects on Item 3 "seeking comfort when upset" and Item 4 "uncomfortable with physical affection" (reversed scored) were both positive ($\beta = .10$ and $\beta = .10$). So for these items, when the level of Closeness is equal, girls received somewhat higher scores (demonstrating scalar non-invariance). The correlation between the factors Closeness and Student Gender was positive and significant ($r = .25$, $p < .001$). Because boys were scored 0 and girls 1, this correlation means that teachers reported more Closeness with girls than boys.

We then examined cluster bias (in Step 4). The fit of the model with equal factor loadings on the within-clusters and between-clusters level, and no residual variance on the between-clusters level, was unsatisfactory, χ^2 (109) = 831.67, $p < .001$, RMSEA = .067, and CFI = .85. One-by-one freeing of the residual variance for the items associated with the highest modification indices resulted in a model with all between-clusters level residual variance estimated. This model fitted well, χ^2 (98) = 322.77, $p < .001$, RMSEA = .039, and CFI = .95, and significantly better than the model without residual variance, $\Delta\chi^2$ (11) = 574.62, $p < .001$. The presence of cluster bias in all Closeness items indicates that part of the variance is caused by differences between teachers.

For three items, the factor loadings could not be considered equal across the within-clusters level (student level) and between-clusters level (teacher level). Therefore, the factor

This is an Author's Accepted Manuscript of an article published in Journal of School Psychology, 2012, 50, 363-378. © Elsevier, available online at http://dx.doi.org/10.1016/j.jsp.2011.12.002

loadings of Item 7 (“beams with pride when praised”), Item 21 (“allows him/herself to be encouraged”), and Item 27 (“openly shares feelings/experiences”) were freely estimated. These model modifications resulted in a model with excellent fit, χ^2 (95) = 275.23, $p < .001$, RMSEA = .036, and CFI = .96. Items 7 and Item 21 were more indicative (i.e., higher factor loadings) of Closeness at the teacher level, which suggests that the variance in these items reflects differences between teachers rather than differences between students. Item 27 was more indicative of Closeness at the student level, so this item reflects differences between students more than differences between teachers. Therefore, the between-clusters level common factor cannot directly be interpreted as the aggregated version of the within-clusters factor.

The cluster invariance model also allows the calculation of the ICC for the common factor. The ICC is calculated as the ratio of the common between variance over the total common variance. For Closeness, 25% of the total variance can be attributed to differences between teachers.

Multigroup factor analysis was used to fit the model to male teachers and female teachers separately (in Step 5). A model with factor loadings and intercepts constrained to be equal across Teacher Gender, fitted the data well, χ^2 (249) = 515.82, $p < .001$, RMSEA = .038, and CFI = .95. However, modification indices indicated that the intercepts of Item 3 “seeking comfort when upset” and Item 4 “uncomfortable with physical affection” (reversed scored) were not equal across Teacher Gender (both higher for female teachers), indicating scalar non-invariance. Freeing these intercepts improved the model fit significantly, $\Delta\chi^2$ (2) = 21.36, $p < .001$.

This is an Author's Accepted Manuscript of an article published in Journal of School Psychology, 2012, 50, 363-378. © Elsevier, available online at http://dx.doi.org/10.1016/j.jsp.2011.12.002

The standardized difference in the common factor mean[3] between male and female teachers was large (κ= .93, $p < .001$). This finding means that overall, female teachers reported more Closeness with students (both boys and girls) than male teachers (in Step 6). Female teachers also reported more Closeness with girls than boys, indicated by a positive correlation between Student Gender and student-level Closeness ($r = .30$, $p < .001$), but this gender difference was not found for male teachers ($r = .09$, $p = .130$). The significant deteriorating in fit, when the covariance is constrained to be equal across Teacher Gender ($\Delta\chi^2$ (1) = 7.67, $p = .006$), indicates that the interaction between Teacher Gender and Student Gender is significant. Female teachers reported more Closeness for girls, whereas male teachers reported no difference between boys and girls (see Figure 2).

Conflict. For Conflict, the significance of between-clusters level variance and covariance was tested as well (in Step 1). ICCs varied between .08 (Item 20) and .27 (Item 16). Both the independence model, $\Delta\chi^2$ (55) = 106.10, $p < .001$, and the null model, $\Delta\chi^2$ (66) = 281.59, $p < .001$, fitted significantly worse than a model with covariances on both levels. The fit of the one-factor model (in Step 2) was satisfactory, χ^2 (44) = 349.03, $p < .001$, RMSEA = .068, and CFI = .94. However, based on the modification indices, residual covariances between Item 2 and 20 and between Item 11 and 18 were allowed, which significantly improved model fit, $\Delta\chi^2$ (2) = 69.83, $p < .001$. Closer inspection of Item 2 and 20 disclosed that both items refer to a constant exhausting struggle between teacher and child. Item 11 and 18 both denote the child being angry with the teacher.

In Step 3, Student Gender was added to the model, χ^2 (52) = 355.75, $p < .001$, RMSEA = .063, and CFI = .95. A significant direct effect on Item 25 “whines or cries” (β = -.17) was found,

[3] κ is interpretable as Cohen’s *d* (Cohen, 1992), so that κ’s of .20, .50 and .80 are associated with small, medium, and large mean differences, respectively

This is an Author's Accepted Manuscript of an article published in Journal of School Psychology, 2012, 50, 363-378. © Elsevier, available online at http://dx.doi.org/10.1016/j.jsp.2011.12.002

$\Delta\chi^2$ (1) = 25.52, $p < .001$. Thus, when levels of Conflict were equal, girls received lower scores on Item 25 than boys (demonstrating scalar non-invariance).

The overall fit of the cluster invariance model (in Step 4) was acceptable, χ^2 (105) = 508.49, $p < .001$, RMSEA = .051, and CFI = .92. However, cluster bias was found in Items 16, 20, 22, 24, 25, and 26. Allowing between-clusters level residual variance for these items improved model fit in comparison to a cluster invariance model, $\Delta\chi^2$ (6) = 79.78, $p < .001$. Additionally, the factor loading of Item 25 could not be considered equal across levels, $\Delta\chi^2$ (1) = 7.09, $p = .008$. This item was more indicative of the between-clusters level construct than of the within-clusters level construct. Fourteen percent of the common variance (i.e., ICC) in Conflict could be attributed to differences between teachers.

The multigroup model with factor loadings and intercepts constrained to be equal across Teacher Gender (in Step 5), was then tested. This model showed adequate fit, χ^2 (243) = 658.47, $p < .001$, RMSEA = .048, and CFI = .93. Modification indices did not point out items with measurement non-invariance across Teacher Gender.

Finally, mean differences in Conflict between male and female teachers were examined (in Step 6). Female teachers reported considerably less Conflict than male teachers (κ = -.65, p = .013). Both female (r = -.16, $p < .001$) and male teachers (r = -.19, $p < .001$) reported less Conflict for girls than for boys. There was no interaction between Teacher Gender and Student Gender as indexed by non-significant change in fit when this covariance was constrained to be equal, $\Delta\chi^2$ (1) = .94, p = .67. The latent mean differences in Conflict are depicted in Figure 3.

Dependency. The ICCs of the Dependency items ranged between .09 (Item 14) and .39 (Item 8). Both the null model, $\Delta\chi^2$ (21) = 385.30, $p < .001$, and the independence model, $\Delta\chi^2$ (15) = 123.57, $p < .001$, fitted the data significantly worse than a model with covariance on both

This is an Author's Accepted Manuscript of an article published in Journal of School Psychology, 2012, 50, 363-378. © Elsevier, available online at http://dx.doi.org/10.1016/j.jsp.2011.12.002

levels, indicating that multilevel modeling is needed (in Step 1). The fit of the one-factor model of Step 2 was excellent, χ^2 (9) = 22.19, p = .008, RMSEA = .031, and CFI = .99. The fit was still adequate when Student Gender was added in Step 3, χ^2 (14) = 36.81, p = .001, RMSEA = .033, and CFI = .99. Modification indices did not suggest measurement non-invariance across Student Gender.

The cluster invariance model (in Step 4) did not fit satisfactorily based on the CFI, χ^2 (29) = 288.94, p < .001, RMSEA = .077, and CFI = .82. Cluster bias was found in all items, and the factor loadings of Items 8 (“reacts strongly to separation”), 14 (“asks for help when not needed”), 19 (“needs to be continually confirmed”) could not be considered equal across levels. Items 14 and 19 were more indicative of between-clusters level differences (teacher level), and Item 8 was more indicative of within-clusters level differences (student level). Allowing between-clusters level variance for all items and freely estimating these loadings improved model fit in comparison to a cluster invariance model, $\Delta\chi^2$ (9) = 234.86, p < .001. Forty-three percent of the common variance in Dependency could be attributed to differences between teachers.

The multigroup model (in Step 5) with factor loadings and intercepts constrained across Teacher Gender, fitted the data well, χ^2 (73) = 99.53, p = .021, RMSEA = .022, and CFI = .99. The absence of significant modification indices indicates no measurement non-invariance across Teacher Gender.

A small mean difference between male and female reports of Dependency was found in Step 6 (κ = -.32, p = .018), indicating that female teachers experienced less Dependency from students than male teachers. There was no interaction between Teacher Gender and Student Gender, $\Delta\chi^2$ (1) = 2.57, p = .109. Figure 4 presents the latent mean differences for Dependency.

Additional analyses

This is an Author's Accepted Manuscript of an article published in Journal of School Psychology, 2012, 50, 363-378. © Elsevier, available online at http://dx.doi.org/10.1016/j.jsp.2011.12.002

Because male teachers were more present in the higher grades, we tested whether Teacher and Student Gender effects on Closeness, Conflict, and Dependency were moderated by Grade Level (grades 1-3 versus grades 4-6). To this end, a multigroup model with four groups was fitted, namely male teachers in grades 1 through 3, male teachers in grades 4 through 6, female teachers in grades 1 through 3, and female teachers in grades 4 through 6. For male teachers, there were no differences between the two Grade Level groups in the effects of Student Gender on the common factor for any of the three factors, for Closeness, $\Delta\chi^2$ (1) = 0.53, p = .47; for Conflict, $\Delta\chi^2$ (1) = 0.40; p = .53, and for Dependency, $\Delta\chi^2$ (1) = 0.47, p = .49. For female teachers, there were also no differences between the Grade Level groups in the effects of Student Gender, for Closeness, $\Delta\chi^2$ (1) = 0.08, p = .78; for Conflict, $\Delta\chi^2$ (1) = 0.44, p = .51, and for Dependency, $\Delta\chi^2$ (1) = 1.21, p = .27. Thus, Grade Level was not a significant moderator of the effects of Teacher and Student Gender on relationship qualities, indicating similar results in the lower and upper grades.

Discussion

Although research has consistently demonstrated that boys have poorer affective relationships with teachers than girls, the meaning of these findings is not clear because almost all studies have relied on reports of predominantly female teachers and have failed to account for possible measurement non-invariance across teacher gender and student gender. The current study included an acceptable number of male teachers and examined measurement invariance across both teacher and student gender to provide empirically strong evidence for gender differences in teacher–student relationship quality. The results indicated effects of both student and teacher gender.

This is an Author's Accepted Manuscript of an article published in Journal of School Psychology, 2012, 50, 363-378. © Elsevier, available online at http://dx.doi.org/10.1016/j.jsp.2011.12.002

Concerns about the feminization of education are widespread and have been linked to relatively meager educational outcomes for boys in comparison to girls. These concerns prompted us to explore gender differences and gender match in teacher–student relationship quality because relationships with teachers are at the heart of students' school experiences and have a long-term impact on children's social and academic school adjustment (e.g., Hughes et al., 2008; O'Connor & McCartney, 2007). We found that teachers reported overall poorer relationships with boys than with girls. More specifically, both male and female teachers reported more conflictual relationships with boys than girls, and female teachers also reported substantially less close relationships with boys. Thus, both female and male teachers seemed to experience special difficulties in relating to boys.

There are two main explanations for the finding that teachers perceive their relationships with boys as more conflictual. First, boys tend to have poorer self-regulation skills and exhibit more antisocial behavior than girls (Carlson, Tamm, & Gaub, 1997; Gaub, & Carlson, 1996; Zahn-Waxler, Shirtcliff, & Marceau, 2008). Second, as a result of the feminization of education, it could be that both male and female teachers value female-typical behaviors such as cooperation, compliance, and attentiveness more than stereotypical male behaviors such as liveliness, impulsivity, and autonomy. It has been suggested for a long time that the school environment has become more congruent with the female than the male stereotype (Brophy & Good, 1974; Fagot, 1981). Thus, qualities that make a "good student" may be more typical for girls.

Intriguingly, the most conflictual relationships were found between boys and male teachers. This finding runs counter to the expectation that a same-gender match results in more positive relationships, as was based on gender schema and role model theories. It is in sharp

This is an Author's Accepted Manuscript of an article published in Journal of School Psychology, 2012, 50, 363-378. © Elsevier, available online at http://dx.doi.org/10.1016/j.jsp.2011.12.002

contrast with the common belief among policy makers that boys would have better relationships with male teachers, because male teachers are better able to empathize with boys' feelings and interests and are more tolerant of boy-typical interactive behavior (e.g., Cushman, 2007). Research in Greece has also demonstrated that male teachers tend to rate boys higher on interpersonal discordance than female teachers, whereas no differentiation was detected in such reports of girls' negative interpersonal behaviors (Hopf & Hatzichristou, 1999). Male teachers may have more power struggles with male students because both are biologically disposed and socialized to strive for autonomy and dominance, an interpretation that converges with the gender role socialization framework.

Female teachers reported less close relationships with boys than girls, whereas no such differences were found for male teachers. Female teachers appear to share the most favorable relationships with girls as indexed by the highest levels of Closeness and lowest levels of Conflict. The mutual inclination to seek nurturing and emotionally-rewarding relationships with others may result in a positive female to female match. Ewing's dissertation study (2009), albeit not controlling for measurement invariance, also reported a positive match between girls and female teachers. We did not, however, find a positive match between boys and male teachers. Thus, the presumption that boys identify more with male teachers than female teachers and therefore share more affection with them seems not tenable based on teacher reports.

The belief is widespread that the high number of female teachers in primary education is a factor that perpetuates the inequality in social and academic functioning between boys and girls. However, whereas previous investigations could not confirm the presumed positive effects of same-gender teachers on students' cognitive adjustment and performance (Driessen, 2007; Marsh et al., 2008; Martin & Marsh, 2005; Neugebauer et al., 2010), the current results disconfirm

This is an Author's Accepted Manuscript of an article published in Journal of School Psychology, 2012, 50, 363-378. © Elsevier, available online at http://dx.doi.org/10.1016/j.jsp.2011.12.002

similar presumed effects on a social indicator of school adjustment, i.e., teacher–student relationships. Overall, in convergence with existing research, the current study suggests that increasing the number of male teachers in primary education, although perhaps warranted for other reasons, may not attenuate the social and academic difficulties of boys.

Importantly, this study also revealed significant differences between male and female teachers. Female teachers reported better relationships with their students than male teachers, all along the line. Large, medium, and small differences between male and female teachers were found for Closeness, Conflict, and Dependency, respectively. The most plausible explanation for these findings is that females are more socialized to cultivate nurturing relationships with others and more accepting of students' misbehavior and comfort-seeking behavior.

Non-invariance in the Measurement of Relationship Quality across Gender

Although overall sufficient measurement invariance was found for valid group comparisons (Byrne et al., 1989), the constructs were not entirely free from gender non-invariance. Two of the eleven Closeness items, one of the eleven Conflict items, and none of the six Dependency items demonstrated scalar non-invariance across student gender, meaning that Closeness and Conflict were measured on slightly different scales for boys and girls. In addition, the same two Closeness items, but none of the Conflict and Dependency items, showed scalar non-invariance (but not metric non-invariance) across teacher gender.

It appeared that girls are judged to be more comfortable with physical affection and more inclined to seek comfort from teachers when upset whereas boys are more easily judged as whiny when they want something. When levels of Closeness were similar, female teachers tended to report more comfort seeking behavior from students and judge students as more comfortable with physical affection, probably because these items reflect more nurturing, female-typical (or

This is an Author's Accepted Manuscript of an article published in Journal of School Psychology, 2012, 50, 363-378. © Elsevier, available online at http://dx.doi.org/10.1016/j.jsp.2011.12.002

stereotypical) attitudes (cf. Eagly, 2009). Koomen et al. (2011) suggest that girls' interactive behavior may be overrepresented in the closeness construct as measured by the adapted Student–Teacher Relationship Scale. Our research takes this suggestion one step further by suggesting that it is typical female interactive behavior that seems prevalent in the closeness construct. Researchers are encouraged to pursue a conceptualization of teacher–child relationships that is really gender sensitive. Furthermore, because this research is the first to examine measurement invariance across teacher and student gender in a Dutch sample, researchers both in the Netherlands and elsewhere should verify measurement invariance in their own data.

Limitations, Implications, and Conclusions

Some qualifications should be recognized. First, metric invariance (i.e., equality of factor loadings) across student gender was assumed but could not be tested in a multilevel model. This fact limits the interpretation of scalar invariance (i.e., the equality of intercepts) because differences in factor loadings can produce differences in intercepts as an artifact. However, previous single-level analyses revealed metric invariance of the three factor-model (Closeness, Conflict, and Dependency) across student gender (Koomen et al., 2011). Second, the results were obtained in a Dutch sample and may not generalize to other countries. Gender-role differentiation could be larger in other cultures, especially in more traditional societies. Although this study demonstrated largely similar teacher and student gender effects as found in a Greek sample (Hopf & Hatzichristou, 1999), an example of a more traditional society, cross-cultural replication is warranted. Third, this study examined teacher perceptions of relationship quality. To obtain a more comprehensive picture of gender differences in teacher–student relationships, it is necessary to investigate teacher and gender effects in student reports and reports from independent observers. Lastly, the study does not address the implications of poorer teacher–child

This is an Author's Accepted Manuscript of an article published in Journal of School Psychology, 2012, 50, 363-378. © Elsevier, available online at http://dx.doi.org/10.1016/j.jsp.2011.12.002

relationships for boys in relation to their future school development. For instance, relatively high conflict levels with male teachers may be less harmful for boys' school adjustment than high conflict levels with female teachers, because male-to-male conflict could result from male-typical (i.e., gender normative) behaviors and attitudes. Similarly, boys may have a less need for close relationships with female teachers than girls.

Despite these limitations, our study contributes insight into gender differences in teacher–student relationships. Teachers and policy-makers should take note of empirical findings that challenge widespread ideas about the importance of male teachers for boys. The picture that emerges from the current study is sobering: male teachers do not report better relationships and even tend to have more conflictual relationships with boys than female teachers. Male teachers may differ from female teachers in their way of relating to students, which is not necessarily better or worse. However, judging from the effects of teacher–student relationship quality that have been amply demonstrated for students' school adjustment, male teachers should perhaps be encouraged to purposefully invest in emotionally-close relationships and be supported to help their students deal with personal and emotional issues. Considering the relatively high levels of male-to-male conflict, it seems also fruitful for male teachers to reflect on male-typical issues such as the desire for autonomy and dominance in order to prevent or adequately cope with power struggles and conflicts with boys. Next to differences between male and female teachers, this study showed that boys overall tend to have poorer relationships with teachers than girls. Although beyond the scope of this study, it could be that both male and female teachers need to develop or re-develop a view of the "good student" that is more in line with male gender-role standards.

This is an Author's Accepted Manuscript of an article published in Journal of School Psychology, 2012, 50, 363-378. © Elsevier, available online at http://dx.doi.org/10.1016/j.jsp.2011.12.002

In the context of psychological assessments, school psychologists should recognize that there are slight differences in average mean levels across student and teacher gender (see also Koomen et al., 2011). Although the differences seem small, gender-specific norms may be required for individual assessments. Gender- and age-specific norms for Dutch boys and girls are available for the adapted Student–Teacher Relationship Scale, but these norms are not adjusted for systematic differences between female and male teacher reports about boys and girls (Koomen et al., 2007).

In conclusion, this study examined teacher and student gender effects on teacher–student relationship quality while controlling for measurement non-invariance. Both male and female teachers reported more conflictual relationships with boys than with girls, and female teachers also reported less close relationships with boys. Furthermore, female teachers tended to have overall better relationships with their students than male teachers. The current findings do not dispute the need for male teachers in primary education but further challenge the idea that simply increasing the number of male teachers can resolve or attenuate "boys' problems" in schooling.

This is an Author's Accepted Manuscript of an article published in Journal of School Psychology, 2012, 50, 363-378. © Elsevier, available online at http://dx.doi.org/10.1016/j.jsp.2011.12.002

References

Ang, R. (2005). Development and validation of the teacher-student relationship inventory using exploratory and confirmatory factor analysis. *Journal of Experimental Education, 74*, 55-74.

Baker, J. A. (2006). Contributions of teacher-student relationships to positive school adjustment during elementary school. *Journal of School Psychology, 44*, 211-229.

Basow, S. (2004). The hidden curriculum: Gender in the classroom. In M. A. Paludi (Ed.), *Praeger guide to the psychology of gender* (pp. 117-131). Westport: Praeger.

Bem, S. (1981). Gender schema theory: A cognitive account of sex-typing. *Psychological Review, 88,* 354-364.

Brophy, J. E., & Good, T. L. (1974). *Teacher-student relationships: Causes and consequences*. New York: Holt, Rinehart and Winston.

Browne, M. W. & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research, 21,* 230-258.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105*, 456-466.

Carlson, C. L., Tamm, L., & Gaub, M. (1997). Gender differences in children with ADHD, ODD and co-occurring ADHD/ODD identified in a school population. *Journal of the American Academy of Child & Adolescent Psychiatry, 36*, 1706-1714.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159.

Coley, R. J. (2001). Differences in the gender gap: Comparisons across racial/ethnic groups in education and work. Princeton, NJ: Educational Testing Service.

This is an Author's Accepted Manuscript of an article published in Journal of School Psychology, 2012, 50, 363-378. © Elsevier, available online at http://dx.doi.org/10.1016/j.jsp.2011.12.002

Cushman, P. (2007). The male teacher shortage: A synthesis of research and worldwide strategies for addressing the shortage. *KEDI Journal of Educational Policy, 4*, 79-98.

Cushman, P. (2010). Male primary school teachers: Helping or hindering a move to gender equity? *Teaching and Teacher Education, 26*, 1211-1218.

d'Acremont, M., & Van der Linden, M. (2008). Confirmatory factor analysis of the strengths and difficulties questionnaire in a community sample of French-speaking adolescents. *European Journal of Psychological Assessment, 24*, 1-8.

Derks, E., Dolan, C., Hudziak, J., Neale, M., & Boomsma, D. (2007). Assessment and etiology of attention deficit hyperactivity disorder and oppositional defiant disorder in boys and girls. *Behavior Genetics, 37*, 559-566.

Doumen, S., Koomen, H. M. Y., Buyse, E., Wouters, S., & Verschueren, K. (2011). Teacher and observer views on student-teacher relationships: Convergence across kindergarten and relations with student engagement. *Journal of School Psychology, in press*.

Doumen, S., Verschueren, K., Buyse, E., De Munter, S., Max, K., & Moens, L. (2009). Further examination of the convergent and discriminant validity of the Student-Teacher Relationship Scale. *Infant and Child Development, 18*, 502-520.

Doumen, S., Verschueren, K., Buyse, E., Germeijs, V., Luyckx, K., & Soenens, B. (2008). Reciprocal relations between teacher-child conflict and aggressive behavior in kindergarten: A three-wave longitudinal study. *Journal of Clinical Child and Adolescent Psychology*, *37*, 588-599.

Driessen, G. (2007). The feminization of primary education: Effects of teachers' gender on pupil achievement, attitudes, and behaviour. *International Review of Education, 53*, 183-203.

This is an Author's Accepted Manuscript of an article published in Journal of School Psychology, 2012, 50, 363-378. © Elsevier, available online at http://dx.doi.org/10.1016/j.jsp.2011.12.002

Eagly, A. H. (2009). The his and hers of prosocial behavior: An examination of the social psychology of gender. *American Psychologist, 64*, 644-658

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.

Ewing, A. (2009). *Teacher-child relationship quality and children's school outcomes: Exploring gender differences across elementary school grades.* Ph.D. thesis, The University of Arizona.

Ewing, A. R., & Taylor, A. R. (2009). The role of child gender and ethnicity in teacher-child relationship quality and children's behavioral adjustment in preschool. *Early Childhood Research Quarterly, 24*, 92-105.

Fagot, B. I. (1981). Male and female teachers: Do they treat boys and girls differentely? *Gender Roles, 7,* 263-271.

Francis, B., Skelton, C., Carrington, B., Hutchings, M., Read, B., & Hall, I. (2006). *A perfect match? Pupils' and teachers' views of the impact of matching educators and learners by gender*. Paper presented at the British Educational Research Association Annual Conference, University of Warwick.

Furrer, C., & Skinner, E. (2003). Sense of relatedness as a factor in children's academic engagement and performance. *Journal of Educational Psychology, 95,* 148-162.

Gaub, M., & Carlson, C. L. (1996). Gender differences in ADHD. A meta-analysis and critical review. *Journal of the American Academy of Child & Adolescent Psychiatry, 36*, 1036-1045.

Hamre, B. K., & Pianta, R. C. (2001). Early teacher-child relationships and the trajectory of children's school outcomes through eighth grade. *Child Development, 72*, 625-638.

This is an Author's Accepted Manuscript of an article published in Journal of School Psychology, 2012, 50, 363-378. © Elsevier, available online at http://dx.doi.org/10.1016/j.jsp.2011.12.002

Hopf, D., & Hatzichristou, C. (1999). Teacher gender-related influences in Greek schools. *British Journal of Educational Psychology, 68,* 1-18.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indices in covariance structure analysis: Conventional versus new alternatives. *Structural Equation Modeling, 6,* 1-55.

Hughes, J. N., Luo, W., Kwok, O. M., & Loyd, L. K. (2008). Teacher–student support, effortful engagement, and achievement: A 3-year longitudinal study. *Journal of Educational Psychology, 100,* 1-14.

Jak, S., Oort, F. J. & Dolan, C.V. (2011a). Measurement bias in multilevel data. *Manuscript submitted for publication.*

Jak, S., Oort, F.J. & Dolan, C.V. (2011b). A test for cluster bias: Detecting violations of measurement invariance across clusters in multilevel data. *Manuscript submitted for publication.*

Koepke, M. F., & Harkins, D. A. (2008). Conflict in the classroom: Gender differences in the teacher-child relationship. *Early Education & Development, 19,* 843-864.

Koomen, H. M. Y., & Jellesma, F. C. (2011). Can closeness, conflict, and dependency be used to characterize students' perceptions of relationships with their teachers? Testing a new child measure in middle childhood. *Manuscript submitted for publication.*

Koomen, H. M. Y., Verschueren, K., & Pianta, R. C. (2007). *Leerling-Leerkracht Relatie Vragenlijst (LLRV): Handleiding. [Student-Teacher Relationship Scale: Manual]*. Houten: Bohn Stafleu van Loghum.

Koomen, H. M. Y., Verschueren, K., van Schooten, E., Jak, S., & Pianta, R. C. (2011). Validating the Student-Teacher Relationship Scale: Testing factor structure and

This is an Author's Accepted Manuscript of an article published in Journal of School Psychology, 2012, 50, 363-378. © Elsevier, available online at http://dx.doi.org/10.1016/j.jsp.2011.12.002

measurement invariance across child gender and age in a Dutch sample. *Journal of School Psychology*. doi: 10.1016/j.jsp.2011.09.001

Ladd, G. W., Birch, S. H., & Buhs, E. S. (1999). Children's social and scholastic lives in kindergarten: Related spheres of influence? *Child Development, 70*, 1373-1400.

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*, 1579-1601.

Marsh, H. W., Martin, A. J., & Cheng, J. H. S. (2008). A multilevel perspective on gender in classroom motivation and climate: Potential benefits of male teachers for boys? *Journal of Educational Psychology, 100*, 78-95.

Martin, A., & Marsh, H. W. (2005). Motivating boys and motivating girls: Does teacher gender really make a difference? *Australian Journal of Education, 49*, 320-334.

Mashburn, A. J., Hamre, B. K., Downer, J. T., & Pianta, R. C. (2006). Teacher and classroom characteristics associated with teachers' ratings of pre-kindergartners' relationships and behaviors. *Journal of Psychoeducational Assessment, 24*, 367-380.

Matthews, J. S., Ponitz, C. C., & Morrison, F. J. (2009). Early gender differences in self-regulation and academic achievement. *Journal of Educational Psychology, 101*, 689-704.

Meece, J. L., & Scantlebury, K. (2006). Gender and schooling: Progress and persistent barriers. In J. Worell & C. D. Goodheart (Eds.), *Handbook of girls' and women's psychological health* (pp. 283-291). New York: Oxford University Press.

Meehan, B. T., Hughes, J. N., & Cavell, T. A. (2003). Teacher-student relationships as compensatory resources for aggressive children. *Child Development, 74*, 1145-1157.

Mellenbergh, G. J. (1989). Item bias and item response theory. *International Journal of*

This is an Author's Accepted Manuscript of an article published in Journal of School Psychology, 2012, 50, 363-378. © Elsevier, available online at http://dx.doi.org/10.1016/j.jsp.2011.12.002

Educational Research, 13, 127-143.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*, 525-543.

Muthén, L. K., & Muthén, B. O. (1998-2007). Mplus (Version 5). Los Angeles: Muthén & Muthén.

Neugebauer, M., Helbig, M., & Landmann, A. (2010). Unmasking the myth of the same-sex teacher advantage. *European Sociological Review*. doi: 10.1093/esr/jcq038

O'Connor, E. (2010). Teacher-child relationships as dynamic systems. *Journal of School Psychology*, *48*, 187-218.

O'Connor, E., & McCartney, K. (2007). Examining teacher-child relationships and achievement as part of an ecological model of development. *American Educational Research Journal, 44*, 340-369.

Oort, F. J. (1992). Using restricted factor analysis to detect item bias. *Methodika, 6*, 150-166.

Oort, F. J. (1998). Simulation study of item bias detection with restricted factor analysis. *Structural Equation Modeling, 5*, 107-124.

Pianta, R. C. (2001). *Student-Teacher Relationship Scale. Professional Manual*. Lutz, Florida: Psychological Assessment Resources.

Pomerantz, E. M., Altermatt, E. R., & Saxon, J. L. (2002). Making the grade but feeling distressed: Gender differences in academic performance and internal distress. *Journal of Educational Psychology, 94*, 396–404.

Rey, R. B., Smith, A. L., Yoon, J., Somers, C., & Barnett, D. (2007). Relationships between teachers and urban African American children: The role of informant. *School Psychology International, 28*, 346-364.

This is an Author's Accepted Manuscript of an article published in Journal of School Psychology, 2012, 50, 363-378. © Elsevier, available online at http://dx.doi.org/10.1016/j.jsp.2011.12.002

Roorda, D. L., Koomen, H. M. Y., Spilt, J. L., & Oort, F. J. (2011). The influence of affective teacher-student relationships on students' school engagement and achievement: a meta-analytic approach. *Review of Educational Research, 81, 493-529.*

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika, 66*, 507-514.

Silver, R. B., Measelle, J. R., Armstrong, J. M., & Esgender, M. J. (2005). Trajectories of classroom externalizing behavior: Contributions of child characteristics, family characteristics, and the teacher-child relationship during the school transition. *Journal of School Psychology, 43*, 39-60.

Skelton, C. (2003). Male primary teachers and perceptions of masculinity. *Educational Review, 55*, 195-209.

Skinner, E. A., & Belmont, M. J. (1993). Motivation in the classroom: Reciprocal effects of teacher behavior and student engagement across the school year. *Journal of Educational Psychology, 85*, 571-581.

Spilt, J. L., Hughes, J. N., Wu, J.-Y., & Kwok, O. (2011). Dynamics of teacher-student relationships: Stability and change across elementary school and the influence on children's academic success. *Child Development, in press*.

Spilt, J. L., Koomen, H. M. Y., Thijs, J. T., Stoel, R. D., & van der Leij, A. (2010). Teachers' assessment of antisocial behavior in kindergarten: Physical aggression and measurement bias across gender. *Journal of Psychoeducational Assessment, 28*, 129-138.

Stipek, D., & Miles, S. (2008). Effects of aggression on achievement: Does conflict with the teacher make it worse? *Child Development, 79*, 1721-1735.

This is an Author's Accepted Manuscript of an article published in Journal of School Psychology, 2012, 50, 363-378. © Elsevier, available online at http://dx.doi.org/10.1016/j.jsp.2011.12.002

Thijs, J. T., & Koomen, H. M. Y. (2008). Task-related interactions between kindergarten children and their teachers: the role of emotional security. *Infant and Child Development, 17*, 181-197.

Tremblay, R. E., Desmarais-Gervais, L., Gagnon, C., & Charlebois, P. (1987). The Preschool Behaviour Questionnaire: Stability of its factor structure between cultures, genderes, ages and socioeconomic classes. *International Journal of Behavioral Development, 10*, 467-484.

Wicherts, J. M. & Dolan, C. V. (2010). Measurement invariance in confirmatory factor analysis; An illustration using IQ test performance of minorities. *Educational Measurement: Issues and Practice, 29*, 39-47.

Yuan, K. H. & Bentler, P. M. (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. In M. E. Sobel & M. P. Becker (Eds.), *Sociological Methodology 2000* (pp. 165-200). Washington, D.C.: American Sociological Association.

Zahn-Waxler, C., Park, J. H., Usher, B., Belouad, F., Cole, P., & Gruber, R. (2008). Young children's representations of conflict and distress: A longitudinal study of boys and girls with disruptive behavior problems. *Development and Psychopathology, 20*, 99-119.

Zahn-Waxler, C., Shirtcliff, E. A., & Marceau, K. (2008). Disorders of childhood and adolescence: Gender and psychopathology. *Annual Review of Clinical Psychology, 4*, 275-303

This is an Author's Accepted Manuscript of an article published in Journal of School Psychology, 2012, 50, 363-378. © Elsevier, available online at http://dx.doi.org/10.1016/j.jsp.2011.12.002

Table 1.
Observed means and standard deviations for Closeness, Conflict, and Dependency

Teacher gender	Student gender	*n*	Closeness		Conflict		Dependency	
			M	*SD*	*M*	*SD*	*M*	*SD*
Female	Boy	463	3.89	0.62	1.49	0.64	1.75	0.67
	Girl	561	4.24	0.54	1.36	0.56	1.91	0.74
Male	Boy	242	3.71	0.65	1.71	0.79	1.88	0.69
	Girl	227	3.89	0.67	1.54	0.68	1.93	0.74

Note. These statistics do not take possible measurement non-invariance into account.

Table 2.
Factor loadings of the adapted items from the Student–Teacher Relationship Scale (STRS; Pianta, 2001) representing unstandardized factor loadings at the student (within-clusters) level and teacher (between-clusters) level.

	Items	**Student level**	**Teacher level**
	Closeness		
1.	I share an affectionate, warm relationship with this child.	.57	
3.	If upset, this child will seek comfort from me. [ab]	.55	
4.	This child is uncomfortable with physical affection or touch from me. (reversed scored) [ab]	.32	
5.	This child values his/her relationship with me.	.45	
7.	When I praise this child, he/she beams with pride.	.19	.51
9.	This child seems to feel secure with me. *	.44	
12.	This child tries to please me.	.52	
15.	It is easy to be in tune with what this child is feeling.	.51	
21.	This child allows himself/herself to be encouraged by me. *	.35	.52
27.	This child openly shares his/her feelings and experiences with me.	.78	
28.	My interactions with this child make me feel effective and confident.	.65	
	Conflict		
2.	This child and I always seem to be struggling with each other.	.60	
11.	This child easily becomes angry with me.	.61	
13.	This child feels that I treat him/her unfairly.	.71	
16.	This child sees me as a source of punishment and criticism.	.59	
18.	This child remains angry or is resistant after being disciplined.	.75	
20.	Dealing with this child drains my energy.	.84	
22.	When this child is in a bad mood, I know we're in for a long and difficult day.	.67	
23.	This child's feelings toward me can be unpredictable or can change suddenly.	.73	
24.	Despite my best efforts, I'm uncomfortable with how this child and I get along.	.49	
25.	This child whines or cries when he/she wants something from me. [a]	.36	.61
26.	This child is sneaky or manipulative with me.	.58	
	Dependency		
6.	This child fixes his/her attention on me the whole day long. *	.45	
8.	This child reacts strongly to separation from me.	.26	.49
10.	This child is overly dependent on me.	.47	
14.	This child asks for my help when he/she really does not need help.	.66	.26
17.	This child expresses hurt or jealousy when I spend time with other children.	.29	
19.	This child needs to be continually confirmed by me. *	.67	.37

Notes. Level 2 factor loadings are reported when significant differences were found between the student level and teacher level.
*items that are not in the original STRS (Pianta, 2001)
[a] measurement non-invariance across Student Gender
[b] measurement non-invariance across Teacher Gender

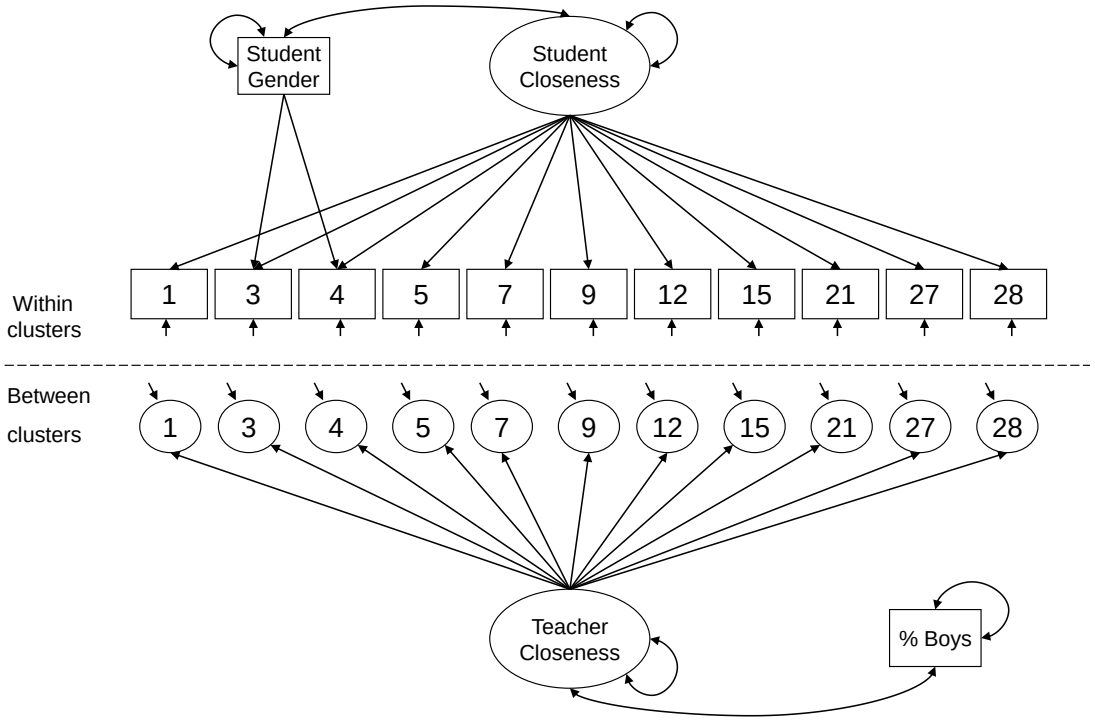

Figure 1. Multilevel factor model of Closeness with effects of Student Gender.

Note. Above the dotted line pertains to the within-clusters (student) level, and below the dotted line pertains to the between-clusters (teacher) level. Large ellipses denote (latent) common factors, small ellipses denote the between-clusters level part of the item scores. Squares represent observed variables (i.e., questionnaire items). Two-sided arrows represent a variance or covariance. One-sided arrows pointing from the common factor to the items denote factor loadings. The small one-sided arrows denote residual variance. In Step 2, a one-factor model was fitted at the student level. In Step 3, Student Gender was added as a covariate. The direct effects of Student Gender on Items 3 and 4 indicate scalar non-invariance in these items. In Step 4, the same model was fitted at the teacher level as well, with equal factor loadings across levels. Residual variance at the teacher level indicates cluster bias. This model was compared between female and male teachers in a multigroup analysis (in Step 5 & 6). Similar models were evaluated for Conflict and Dependency.

This is an Author's Accepted Manuscript of an article published in Journal of School Psychology, 2012, 50, 363-378. © Elsevier, available online at http://dx.doi.org/10.1016/j.jsp.2011.12.002

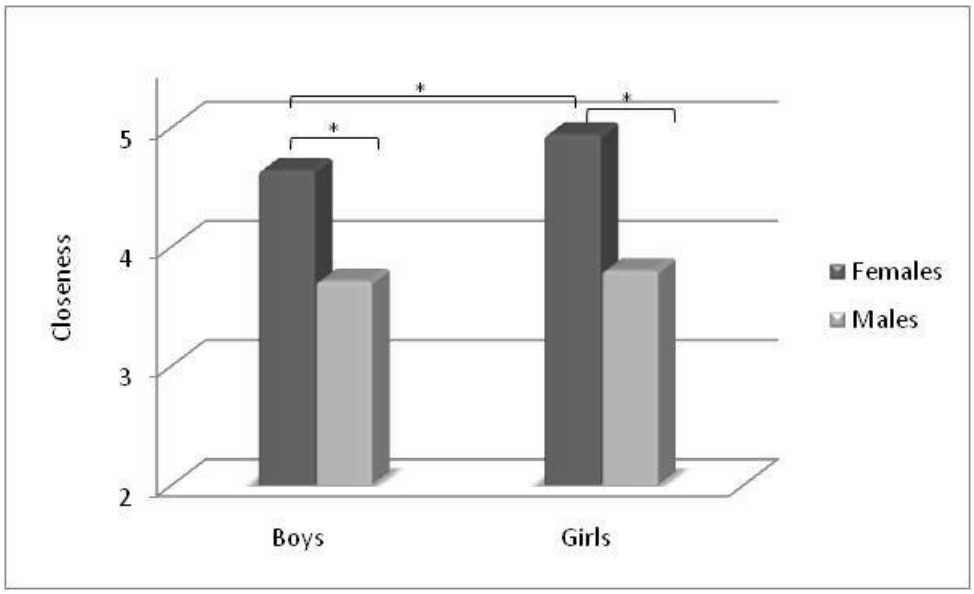

Figure 2. Estimated latent mean differences in Closeness controlled for measurement non-invariance across Student Gender and Teacher Gender

This is an Author's Accepted Manuscript of an article published in Journal of School Psychology, 2012, 50, 363-378. © Elsevier, available online at http://dx.doi.org/10.1016/j.jsp.2011.12.002

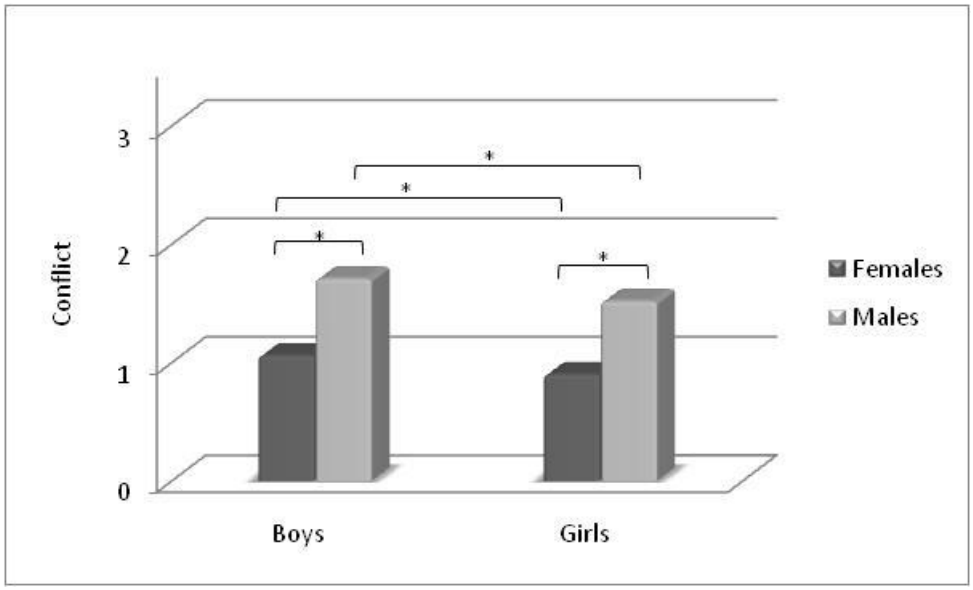

Figure 3. Estimated latent mean differences in Conflict controlled for measurement non-invariance across Student Gender and Teacher Gender

This is an Author's Accepted Manuscript of an article published in Journal of School Psychology, 2012, 50, 363-378. © Elsevier, available online at http://dx.doi.org/10.1016/j.jsp.2011.12.002

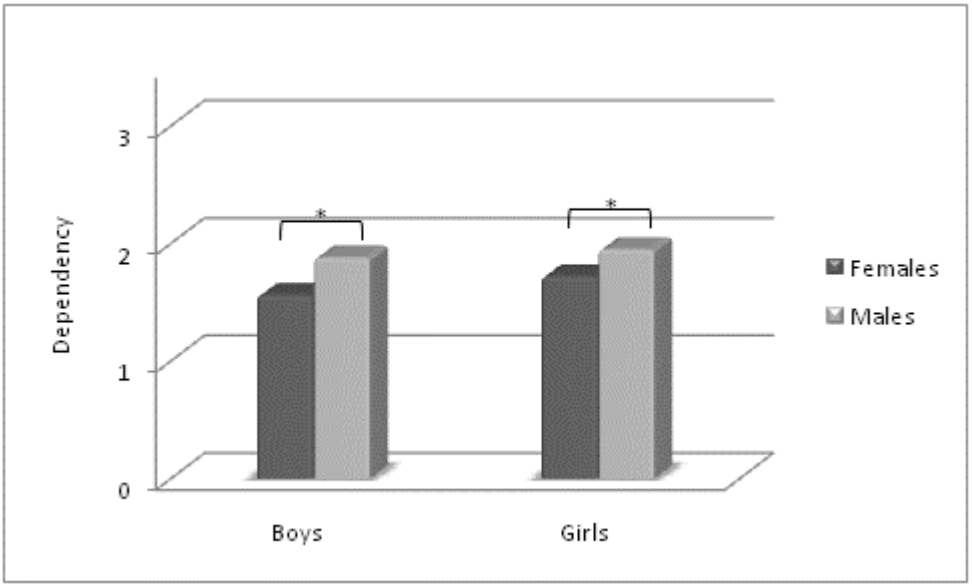

Figure 4. Estimated latent mean differences in Dependency controlled for measurement non-invariance across Student Gender and Teacher Gender

This is an Author's Accepted Manuscript of an article published in Journal of School Psychology, 2012, 50, 363-378. © Elsevier, available online at http://dx.doi.org/10.1016/j.jsp.2011.12.002