

Teachers' Relationships with Children in the Finnish Early Childhood Education Context: A Validation Study

Journal of Psychoeducational Assessment
2021, Vol. 0(0) 1–13
© The Author(s) 2021



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/07342829211019150

journals.sagepub.com/home/jpa



Wenwen Yang , Eero Laakkonen, and Maarit Silvén

Abstract

This study examined the factorial validity and measurement invariance of the Student–Teacher Relationship Scale–Short Form (STRS-SF), modified by Whitaker et al. (2015), in the Finnish Early Childhood Education (ECE) context. Confirmatory factor analysis (CFA) supported the two-factor model of Closeness and Conflict after some item modifications and partial strong measurement invariance across ECE student teachers and two qualified ECE teacher groups. In general, the participants perceived high levels of closeness and low levels of conflicts with children. The qualified teachers who voluntarily enrolled in training to improve their professional competence perceived their relationships with children as more conflictual, reported less working experience, and had younger children in their classrooms, compared to the other qualified teachers. The student teachers perceived less closeness than the teachers but reported fewer conflicts than the teachers enrolled in training. This study extended the application of the STRS-SF in teacher education and research to a Nordic cultural context.

Keywords

Student–Teacher Relationship Scale, measurement invariance, confirmatory factor analysis, early childhood education

Drawing on attachment theory and bioecological systems theory, the quality of adult–child interactions in various contexts, such as the family, daycare, and school settings, is critical to children's secure relationships and socialization (Bowlby, 1969; Bronfenbrenner, 1986; Pianta, 1999; Sroufe, 1988). Similar to the development of trusting parent–child relationships in the early years, there are bidirectional influences between the characteristics of individuals and the quality of teacher–child interactions in the classroom. Many empirical studies have reported that warm and trusting teacher–child relationships are positively associated with future indicators of school

Department of Teacher Education, University of Turku, Finland

Corresponding Author:

Wenwen Yang, Department of Teacher Education, Faculty of Education, University of Turku, Assistentinkatu 5, Turku 20500, Finland.

Email: wenyan@utu.fi

adjustment and academic development (Downer et al., 2012; Hamre et al., 2008; Meehan et al., 2003; Pianta & Steinberg, 1992; Pianta & Stuhlman, 2004; Stipek & Byler, 2004).

Since collecting observation or interview data is time-consuming and expensive, determining the most suitable and effective self-report method is very important. In this regard, the Student–Teacher Relationship Scale (STRS) developed by Pianta (2001a), designed to operationalize Bowlby’s construct of internal working models of relationships (Pianta & Steinberg, 1992), has gained widespread international acceptance. The STRS assesses a teacher’s perceptions of their relationships with individual children aged 4–8 years. The original 28-item STRS is composed of three constructs showing good internal consistency. Closeness ($\alpha = .86$) represents a positive and warm relationship with a specific child; conflict ($\alpha = .92$) indicates negative and conflictual feelings toward the child; and dependency ($\alpha = .64$) concerns the child’s overdependence on the teacher. Moreover, the STRS-SF (Pianta, 2001b), a short form with 15 items drawn from the original longer form, represents the constructs of Closeness and Conflict. In the current study, we used a modified version of the STRS-SF (Whitaker et al., 2015), which was translated and adapted into Finnish.

A growing number of studies on children’s developmental outcomes suggest that higher quality teacher–child relationships, as assessed by the STRS at the individual level, predict children’s better cognitive, social, and behavioral outcomes, including less risk of school failure in the long term (Baker et al., 2008; Birch & Ladd, 1997; Hamre & Pianta, 2001; Howes, 2000; Jerome et al., 2009). Other studies have explored the characteristics of teachers, such as workplace stress, self-efficacy, and depression, or educational contexts and have estimated teacher–child relationships at the classroom level by averaging STRS item scores based on a sample of four or six randomly selected children (Hamre et al., 2008; Kesner, 2000) or all the children in the classroom (Howes, 2000). In order to improve cost-effectiveness, Whitaker et al. (2015) modified the STRS-SF instructions and the 15 items, mainly replacing the singular forms into plural forms, for example, “child” into “children.” Because the aggregated version focuses on teachers’ perceptions of their relationships with all children in the classroom, it estimates a more generalized construct of teachers’ professional competence than the STRS at the individual level. The current study is the first to explore the psychometric properties of the modified version using a Finnish sample of early childhood education (ECE) teachers at various career stages, from the undergraduate level onward.

A substantial number of studies in the United States and Europe have examined the factorial validity of the long and short form of STRS across cultures, languages, and educational levels, using exploratory or confirmatory factor analysis (EFA or CFA). The three-factor structure has been confirmed in US studies of teachers of 4- to 8-year olds, with the total number of original scale items (Pianta, 2001a), and teachers of 5-year-olds, with a reduced number of items (Webb & Neuharth-Pritchett, 2011). In Europe, the original three-factor structure has shown an acceptable fit in samples of teachers of preschool, kindergarten, and/or school-aged children after adjustments to account for the cultural and linguistic contexts. This holds true for studies conducted in Greece (Gregoriadis & Tsigilis, 2008), Norway (Solheim et al., 2012), Germany and Austria (Milatz et al., 2014), the Netherlands (Koomen et al., 2012), and Italy (Fraire et al., 2013). In general, the European findings suggest better-fitting three-factor models when certain STRS items are removed.

Surprisingly, few studies have validated the hypothesized two-factor structure of the short form of STRS in the ECE setting (for school-aged children in Norway and Portugal, see Drugli & Hjemdal, 2013 and Patrício et al., 2015, respectively). A two-factor model with negatively correlated Closeness and Conflict constructs was valid and reliable for a Greek kindergarten sample (Tsigilis & Gregoriadis, 2008) and for a Ghanaian preschool sample (Aboagye et al., 2019). At present, however, it is not clear whether the factor structure of STRS-SF would generalize across other cultural and linguistic contexts.

Regarding other psychometric properties, some studies have explored whether the constructs of STRS, Closeness and Conflict, as well as Dependency, share their meanings among ECE teachers by testing measurement invariance (also called equivalence) across child and contextual characteristics, such as gender, age, and educational level. The teachers seemed to interpret the meaning of STRS items similarly when assessing boys and girls (Koomen et al., 2012; Milatz et al., 2014; Tsigilis & Gregoriadis, 2008) and children of varying ages and in different educational contexts (Aboagye et al., 2019; Koomen et al., 2012; Milatz et al., 2014). However, the three constructs of the STRS did not function uniformly for ECE teachers of ethnically different groups of children (Webb & Neuharth-Pritchett, 2011). So far, no studies have explored how ECE teacher characteristics, such as age, educational background, and professional experience, influence the measurement invariance of STRS. The same holds true for the aggregated STRS-SF (Whitaker et al., 2015), which may be useful in teacher training to provide feedback about participants' pedagogical competence, as well as to estimate the effect of training.

The Present Study

There is a need to expand the assessment of relational processes, from the teachers' perspectives, from individual children to the classroom level in various cultural contexts (Sabol & Pianta, 2012). The current study represents a Northern European early childhood education and care (ECEC) context, with affordable high-quality childcare services for all families. In Finland, the daycare centers serve families who have 1- to 6-year-old children (most infants are cared for at home). All children under school age have a subjective right to ECEC, which comprises education and care to support their well-being, balanced development, and learning.

In Finland, ECEC is under the mandate of the Ministry of Education and Culture. The national core curriculum for ECEC (Finnish National Agency for Education, 2018), established by the *Act on ECEC (2018)* and determined by the Finnish National Agency for Education, steers the systematic evaluation, development, and implementation of high quality and equal ECEC throughout the entire country. The national core curriculum for ECEC (Finnish National Agency for Education, 2018) provides a common basis and instructions for developing the ECEC plan for each child in collaboration with their guardians, and the staff is encouraged to become aware of children's needs and to create an interactive and stable relationship to support each child's development.

Based on our literature review, there is little scientific evidence concerning the factorial structure of the STRS-SF, and no studies have yet explored measurement invariance across ECE teacher characteristics. To explore relational processes perceived by Finnish teachers from the undergraduate level onward, we recruited first-year ECE student teachers at a university and qualified ECE teachers at different stages of their professional development. A part of the ECE teacher sample participated in an in-service training organized by universities. According to legislation (*Act on ECEC, 2018*), the municipalities are obligated to provide ECE staff in-service training to update their professional development on a yearly basis with full salary benefits.

The current study addresses the following research questions regarding the participants' perceptions of their overall relationships with a group of children assessed with the aggregated STRS-SF (Whitaker et al., 2015): (1) Can we verify the hypothesized two-factor model with negatively correlated factors among the student teachers, teachers in training, and teachers? (2) Can we show at least partial strong measurement invariance, suggesting that the constructs of Closeness and Conflict have the same underlying meaning for the three groups? (3) Are there differences in the factor means and factor variances between the three groups in how they perceive their overall relationships with children?

Method

Participants and Procedure

In the present study, we collected self-reports online of first-year ECE student teachers studying in a bachelor's degree program in early childhood teacher education (180 ECTS credits; for more details about the European Credit Transfer and Accumulation System, see the European Commission's website, <https://ec.europa.eu>) at a university. At the beginning of a study module, 103 (females = 93) out of 115 student teachers gave their consent for using the self-report data in the present study. The students had completed around 30 ECTS points, including one week of practical training in daycare centers with 3- to 5-year-old children. Almost all students were taking the modules of the curriculum in the recommended order.

In addition to the student teachers, we recruited two samples of qualified ECE teachers working in daycare centers in two southwestern provinces of Finland. The eligibility criteria for the ECE teachers were at least a bachelor's degree in either education from a university or in social sciences from an applied university. Invitations to either participate in the training or respond to the online questionnaire were sent via email to the administration of 8 and 44 municipalities, respectively. The administration was responsible for forwarding the invitations to the daycare centers, and the heads, in turn, forwarded them to the ECE teachers working in the center. Altogether, 40 ECE teachers from 31 daycare centers participated in the in-service training organized by two universities. The aim of the training was to enhance teachers' professional learning and development by practicing new evidence-based ways of assessing and supporting children's development. On the first training day, 39 (females = 38) ECE teachers responded to the online questionnaire and gave their consent for the study. In addition, another sample of 116 ECE teachers (females = 113) who did not attend the training responded to the online questionnaire.

Teacher, Child, and Classroom Characteristics

The online questionnaire (Webropol) contained items about demographic characteristics, relationships with children, and cultural competence (not reported here). Group comparisons revealed $F(2,253) = 105.79, p < .001$; that is, the student teachers were younger ($M = 25.13, SD = 6.32$, and range 19–49) than the two qualified ECE teacher groups, Tukey's test $p_s < .001$ (for descriptive statistics, see Table 1). As shown in Table 1, the teachers in training, on average, had less working experience and younger children in their classrooms, compared to the teachers. In all other respects, the two groups of qualified teachers did not differ: the educational background, child–teacher ratio, and number of immigrant and special needs children in their classrooms were similar. This was expected, as the ECEC legislation provides that multi-professional teams generally consisting of three members with different educational backgrounds are responsible for a group of 12 children or 24 children if the children are older than three years. Moreover, most children with special needs attend mainstream early education settings, which means that the children–ECE staff ratio can be even smaller in such classrooms than mainstream ones.

Assessment of Teacher–Child Relationship

The participants rated the overall quality of their relationships, Closeness and Conflict, with a whole group of children. We used a Finnish translation of the modified version (Whitaker et al., 2015) of the original STRS-SF (Pianta, 2001b). The content of the 15 items was the same in the modified version, except that the word “child” and the singular verbs had been changed to “children” and plural verbs. We also used the instructions modified by Whitaker et al. (2015): “Please, assess how well each of the statements below currently applies to your relationship with

Table 1. Demographic Characteristics of Early Childhood Education Teachers and Children.

| Variable | Teachers in Training | | | Teachers | | | t | df | p |
|-------------------------------------|----------------------|-------|-----------|----------|-------|------------|-------------------|--------|------|
| | M | SD | Range | M | SD | Range | | | |
| Teachers' age, years | 42.58 | 11.18 | 26–60 | 42.90 | 11.17 | 23–63 | — ² | — | — |
| Years of experience | 10.79 | 6.89 | 1–30 | 15.40 | 11.47 | 1–40 | –3.05 | 113.47 | .003 |
| Children's age, ¹ years | 3.92 | 1.48 | 1.38–7.00 | 4.69 | 1.42 | 1.67–7.00 | –2.51 | 141 | .013 |
| Immigrant children ¹ | 3.12 | 3.37 | 0–12 | 3.59 | 4.75 | 0–22 | –.48 | 140 | .633 |
| Special needs children ¹ | 2.19 | 2.29 | 0–8 | 2.38 | 2.23 | 0–9 | –.41 | 141 | .686 |
| Child–teacher ratio ¹ | 5.71 | 1.49 | 3.00–7.67 | 5.74 | 1.84 | 2.33–12.00 | –.07 | 140 | .945 |
| | n | % | | n | % | | χ^2 | df | p |
| BA in social sciences | 22 | 56 | | 53 | 46 | | 1.34 ³ | 2 | .511 |
| BA in education | 14 | 36 | | 52 | 45 | | | | |
| MA in education | 3 | 8 | | 11 | 9 | | | | |

¹Only 26 or 27 out of 39 teachers in training responded these questions.

²The test statistics are reported in the text.

³One cell has expected count less than 5.

children in your classroom. All relationships are individual, but in responding, please think about your relationships with the children in general. Use the scale below to choose the appropriate response for each item.” For the student teachers with no stable classroom, we used “... children attending ECE settings.” instead of “... children in your classroom.” The participants took about 10–15 minutes to complete the questionnaire.

As in the original version, closeness (7 items) represents the positive, warm, and affectionate aspect of relationships (e.g., “I share an affectionate, warm relationship with the children”; “The children openly share their feelings and experiences with me”), whereas conflict (8 items) evaluates the negative, conflictual, and hostile side of relationships (e.g., “The children and I always seem to be struggling with each other”; “The children are uncomfortable with physical affection or touch from me”). The participants rated each statement using a 5-point Likert scale (1 = *definitely does not apply* and 5 = *definitely applies*). In the original version, 3 stands for “neutral, not sure,” but in the Finnish version, we used “in-between.” The internal consistency of the 15-item closeness and conflict scales were .64 and .68 (Cronbach's alpha) and .74 and .80 (ordinal alpha, see Zumbo et al., 2007), respectively.

Three ECE professionals, native Finnish speakers who speak English as their second language, translated the English version into Finnish, which was then back-translated into English by a professional translator native in English and proficient in Finnish. There were minor language discrepancies between the original and the back-translated English version and only a few cultural differences. The original item 10 states: “The children remain angry or are resistant after being disciplined,” but the back-translation reads: “The children are angry or do not care if they have been told off.” This reflects a more passive and typical way of being resistant in the Finnish culture. Moreover, any connotation to the physical discipline of children highly contradicts the ethical principles of ECEC and Finnish legislation. In addition to these minor lexical differences, there is a difference in syntax throughout all the back-translated items: the word order of the sentences reflects the order that is appropriate in Finnish. The developer of the STRS and STRS-SF (Pianta, 2001a, 2001b) approved the back-translated version.

Results

Analytic Strategy

We used Mplus 8 (Muthén & Muthén, 2017) to explore the factor structure and measurement invariance of the sample data. In the first step, we ran CFA to test the postulated two-factor model in the whole sample ($N = 258$) and the three subsamples, that is, student teachers, teachers in training, and teachers. In the second step, we tested the measurement invariance of the constructs across the three groups, using multigroup CFA. Finally, we compared whether the factor means and variances differed between the three groups. We fitted the hypothetical models to the covariance matrix by use of the maximum likelihood method and chose the following indicators to assess the goodness of fit of a specified model: χ^2 likelihood ratio test, the comparative fit index (CFI), the Tucker–Lewis index (TLI), the standardized root mean square residual (SRMR), and the root mean square error of approximation (RMSEA). In accordance with Hu and Bentler (1995), the model is considered reasonably good with the following values: p -value $> .05$, CFI and TLI $\geq .90$, and SRMR and RMSEA $\leq .06$. We inspected the modification indices to identify the sources of misfits. We also inspected $\Delta\chi^2$ and Δ CFI to test and evaluate the differences between nested models (Cheung & Rensvold, 2002; van de Schoot et al., 2012).

Very few missing items ($n = 3$) out of a total of 3870 items were found in the data. The missing data were treated as missing at random and handled in the full information maximum likelihood estimation (Enders, 2010).

Factor Structure

We tested whether the hypothesized two-factor measurement model (with no constraints) fitted the data of the whole sample based on the 15 items (for descriptive statistics, see Table 2). As shown in the upper part of Table 3, the fit indices were less than adequate. The standardized factor loadings

Table 2. Closeness and Conflict Items: Descriptive Statistics for Student Teachers, Teachers in Training, and Teachers.

| Item | Student Teachers | | Teachers in Training | | Teachers | |
|-------------|------------------------|-------|------------------------|-------|------------------------|-------|
| | <i>M</i> (<i>SD</i>) | Range | <i>M</i> (<i>SD</i>) | Range | <i>M</i> (<i>SD</i>) | Range |
| Closeness1 | 4.71 (.45) | 4–5 | 4.75 (.48) | 3–5 | 4.78 (.42) | 4–5 |
| Conflict2 | 1.80 (.79) | 1–5 | 2.26 (.90) | 1–4 | 2.09 (.79) | 1–4 |
| Closeness3 | 4.32 (.66) | 2–5 | 4.77 (.42) | 4–5 | 4.79 (.40) | 4–5 |
| Conflict4 | 1.40 (.67) | 1–4 | 1.44 (.50) | 1–2 | 1.56 (.88) | 1–5 |
| Closeness5 | 4.29 (.60) | 3–5 | 4.44 (.55) | 3–5 | 4.50 (.55) | 3–5 |
| Closeness6 | 4.63 (.57) | 3–5 | 4.69 (.51) | 3–5 | 4.73 (.50) | 3–5 |
| Closeness7 | 4.48 (.62) | 3–5 | 4.28 (.68) | 3–5 | 4.53 (.65) | 2–5 |
| Conflict8 | 1.83 (.76) | 1–5 | 2.47 (.96) | 1–5 | 1.93 (.89) | 2–5 |
| Closeness9 | 4.33 (.60) | 3–5 | 4.21 (.65) | 3–5 | 4.20 (.67) | 1–5 |
| Conflict10 | 2.24 (.82) | 1–5 | 2.56 (1.03) | 1–5 | 2.15 (.97) | 1–5 |
| Conflict11 | 2.72 (.94) | 1–5 | 3.33 (1.07) | 1–5 | 3.51 (1.01) | 1–5 |
| Conflict12 | 2.86 (.94) | 1–5 | 2.80 (1.09) | 1–5 | 2.68 (1.03) | 1–5 |
| Conflict13 | 1.92 (.73) | 1–4 | 2.15 (1.03) | 1–5 | 2.06 (1.14) | 1–5 |
| Conflict14 | 1.39 (.53) | 1–3 | 1.51 (.75) | 1–5 | 1.30 (.53) | 1–3 |
| Closeness15 | 4.48 (.54) | 3–5 | 4.28 (.60) | 3–5 | 4.46 (.58) | 3–5 |

Note. The Finnish version and the back-translated English version are available upon request from the authors.

Table 3. Goodness of Fit Statistics of Closeness and Conflict Constructs: CFA for the Whole Sample and Subsamples.

| Sample | df | χ^2 | RMSEA | SRMR | TLI | CFI | p |
|-------------------------------|----|----------|-------|------|-------|-------|-------|
| Whole sample (N = 258) | | | | | | | |
| Original (15 items) | 89 | 178.14 | .062 | .059 | .811 | .840 | <.001 |
| Modified (12 items) | 50 | 66.02 | .035 | .044 | .956 | .967 | .064 |
| Student teachers (n = 103) | 50 | 61.92 | .048 | .068 | .905 | .928 | .120 |
| Teachers in training (n = 39) | 50 | 46.98 | <.001 | .091 | 1.038 | 1.000 | .595 |
| Teachers (n = 116) | 50 | 55.85 | .032 | .059 | .964 | .973 | .265 |

Note. CFA = confirmatory factor analysis; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual; TLI = Tucker–Lewis index; CFI = comparative fit index.

appeared quite low for items 4, 9, and 11 (.21, .34, and .21, respectively). Moreover, there were correlated measurement errors between three pairs of items (7 and 15 in the Closeness factor; 12 and 13 as well as 13 and 14 in the Conflict factor). In order to establish a well-fitting model for the whole sample, we reran a series of alternative models by removing one at a time the three items with the lowest factor loadings and allowing for the three error covariances. The fit of the final specified model with the remaining 12 items was acceptable for the whole sample. Next, we tested the fit of the final model for the three groups separately. As shown in the lower part of Table 3, the fit indices indicated that the two-factor measurement model was acceptable for each group. The standardized correlations between the two factors for the whole sample, student teachers, teachers in training, and teachers were $-.54$, $-.58$, $-.73$, and $-.47$, respectively.

Measurement Invariance

Measurement invariance is consecutively tested through three increasingly restrictive levels, namely, configural, metric, and scalar invariance (e.g., Byrne, 2006; Meredith, 1993). Configural invariance tests whether the factor structure and patterns of factor loadings of a model are similar across different groups. Failure to demonstrate configural invariance would suggest that the underlying constructs are different from one group to another (for example, see Webb & Neuharth-Pritchett, 2011). At the next level, when testing for metric invariance, the magnitude of the factor loadings is constrained equally across different groups. Failure at this level reveals that one or more items are stronger indicators of the construct for one group than another. Passing the configural as well as the metric invariance tests demonstrates weak measurement invariance (for example, see Tsigilis & Gregoriadis, 2008). The next level tests for scalar invariance by imposing additional item intercepts equally across groups (for example, see Milatz et al., 2014). Success in passing scalar invariance represents strong measurement invariance, which is a requirement for comparing factor means and factor variances between different groups (Little, 2013; Meredith & Teresi, 2006). However, as strong measurement invariance is difficult to achieve, researchers have accepted partial scalar invariance by letting the non-invariant items vary in the model (Greiff & Scherer, 2018; Steenkamp & Baumgartner, 1998; for examples, see Aboagye et al., 2019; Koomen et al., 2012; Milatz et al., 2014).

Table 4 presents the results of the procedure for testing measurement invariance across the student teachers, teachers in training, and teachers, using multigroup CFA. At the basic level, the results showed a good fit for the configural invariance model (M_1), which suggests that the data were fairly well described by the two constructs of Closeness and Conflict. At the next level, weak measurement invariance was tested by constraining the factor loadings of the three groups to be equal and then comparing the metric invariance model (M_2) with the previous configural model (M_1), in which only the number of factors was held invariant. As shown by the fit statistics and the

Table 4. Measurement Invariance of Closeness and Conflict Constructs: Models, Factor Loadings, Intercepts, Means, and Variances.

| Model (M) | χ^2 | df | CFI | TLI | RMSEA | $\Delta\chi^2$ -test (df) | p | Δ CFI |
|---|----------|-----|------|------|-------|------------------------------|-------|--------------|
| M ₁ : Configural | 164.74 | 150 | .970 | .960 | .034 | | | |
| M ₂ : Metric | 196.07 | 170 | .946 | .937 | .042 | 31.33 (20) | .051 | .024 |
| M ₃ : Scalar | 268.74 | 190 | .837 | .831 | .069 | 72.66 (20) | <.001 | .109 |
| M ₄ : Partial scalar | 218.01 | 186 | .934 | .930 | .045 | 21.94 (16) | .145 | .012 |
| M ₅ : Partial scalar equal means in Closeness and Conflict | 238.63 | 190 | .900 | .895 | .055 | 20.61 (4) | .010 | .034 |
| M ₆ : Partial scalar equal means in Closeness | 225.77 | 188 | .922 | .918 | .048 | 7.76 (2) | .021 | .012 |
| M ₇ : Partial scalar equal means in Conflict | 227.42 | 188 | .919 | .914 | .049 | 9.41 (2) | .009 | .015 |
| M ₈ : Partial scalar equal variances in Closeness and Conflict | 224.59 | 190 | .929 | .926 | .046 | 6.58 (4) | .160 | .005 |

| | Closeness | | | Conflict | | |
|------------------|------------------|----------------------|-----------|------------------|----------------------|-----------|
| | Student Teachers | Teachers in Training | Teachers | Student Teachers | Teachers in Training | Teachers |
| Factor loadings | .21–.65 | .31–.84 | .27–.71 | .27–.55 | .37–.73 | .31–.65 |
| Intercepts | 4.35–4.74 | 4.36–4.77 | 4.36–4.74 | 1.58–2.98 | 1.58–2.98 | 1.58–2.98 |
| Factor means | 4.70 | 4.74 | 4.78 | 1.82 | 2.21 | 1.80 |
| Factor variances | .03 | .03 | .02 | .17 | .37 | .28 |

Note. CFI = comparative fit index; RMSEA = root mean square error of approximation; TLI = Tucker–Lewis index.

nonsignificant χ^2 difference ($\Delta\chi^2$), the assumption of an invariant pattern of factor loadings was supported. The metric invariance indicated that the factor loadings were equally strong indicators of the Closeness and Conflict constructs in all three groups. To test scalar invariance, the item intercepts were next constrained as equal across the groups (M₃). The significant $\Delta\chi^2$ revealed that one or more items functioned differently across the three groups. On inspection, the teachers scored higher than the two other groups on Item 2 of Conflict (“The children and I always seem to be struggling with each other”). In Closeness, the teachers in training scored highest on Item 3 (“If upset, the children will seek comfort from me”), and the student teachers scored highest on Item 15 (“The children openly share their feelings and experiences with me”). In order to achieve partial scalar invariance, the intercepts of the non-invariant items were set free to take on any value for the corresponding groups. The fit statistics and the nonsignificant $\Delta\chi^2$ provided adequate support for the partial scalar invariance model (M₄) across the three groups.

Comparing Group Differences on Closeness and Conflict

After testing the measurement invariance of the observed variables, the final step was to explore the equality of the means and variances of the latent factors, Closeness and Conflict, across student teachers, teachers in training, and teachers. We re-estimated the M₄ by imposing equality constraints on the factor means across the three groups. As indicated in Table 4, when compared to the partial scalar invariance model M₄, the significant $\Delta\chi^2$ suggested mean differences in Closeness and Conflict (M₅), both for Closeness (M₆) and Conflict (M₇). Even though all the groups reported

high levels of closeness (means above 4.70), the teachers perceived their overall relationships with children as closer than the student teachers did ($\Delta M = .080, p = .012$), but the teachers in training did not differ either from the student teachers ($\Delta M = .040, p = .328$) or the teachers ($\Delta M = .040, p = .305$). Moreover, the teachers in training perceived their overall relationships with children as more conflictual than did the teachers ($\Delta M = .412, p = .004$) and the student teachers ($\Delta M = .395, p = .003$). The two latter groups did not differ on the level of conflict ($\Delta M = .017, p = .835$). When imposing equality constraints on the factor variances, the three groups of teachers showed an equal amount of variability in Conflict and Closeness.

Discussion

The original STRS with 28 items is a widely used and validated measure of teacher–child relationships across different languages and cultures in the ECEC context. The purpose of the present study was to examine the factorial validity and measurement invariance of a shortened version of STRS, further modified by Whitaker et al. (2015), to assess and to compare Finnish teachers' overall perceptions of their relationships with children, from the undergraduate level onward, across various stages of their careers. In general, ECE student teachers and qualified ECE teachers perceived high levels of overall closeness and low levels of conflictual relationships with children. This pattern of findings might reflect culture-specific beliefs and practices regarding relational processes in the Finnish ECEC context. Our study is the first to confirm, using multigroup CFA, a vigorous and advanced method, that the only existing self-report estimating teachers' overall relationships with a group of children rather than individual children, is a reliable and valid self-report measure that can be applied in teacher education and comparative research, at least in a Nordic cultural context.

The first purpose was to examine the factorial validity of the modified STRS-SF. We confirmed the two-factor structure Closeness and Conflict with a reduced number of items (12 out of 15). Previous CFA studies conducted in the ECEC setting could fit the two-factor structure without any item reduction (Aboagye et al., 2019; Tsigilis & Gregoriadis, 2008; for school-aged children, see Drugli & Hjemdal, 2013; Patrício et al., 2015). In other European studies, the three-factor structure of the STRS could only be confirmed in Italian, Dutch, German, and Austrian samples of teachers across different educational levels, after more or less significant item reductions or modifications (Fraire et al., 2013; Koomen et al., 2012; Milatz et al., 2014). Moreover, it has been typical to improve the model fit by adding parameter specifications that may represent nonrandom measurement errors due to item content overlap (Byrne, 2006). Similarly, in model fitting, we also identified and added error covariances between three pairs of items (see also Tsigilis & Gregoriadis, 2008; for school-aged children, see Patrício et al., 2015).

Studies on factorial validity have reported low or modest negative associations between Closeness and Conflict constructs for teachers of kindergarten-aged children in Ghana ($-.23$) and Greece ($-.38$), respectively (Aboagye et al., 2019; Tsigilis & Gregoriadis, 2008; for school-aged children in Norway and in Portugal, see Drugli & Hjemdal, 2013; Patrício et al., 2015). The negative association suggests that the more the teachers perceived their overall relationship with a group of children as positive and warm, the less they experienced negative feelings and conflict, and vice versa. In our study, the correlations between the latent factors were higher, ranging from moderate ($-.47$) for teachers to high ($-.73$) for teachers in training, compared to the moderate correlation ($-.37$) between summary scores reported by Whitaker et al. (2015) in their large population-based study. The difference in magnitude may be partly due to the fact that latent factors, in contrast to the summary scores, are free of random or systematic measurement errors (for a more detailed discussion, see Tsigilis & Gregoriadis, 2008). It is also possible that the strong negative association between the two relational constructs in our study reflects culture-specific values regarding expectations for, and approaches to, closeness and conflict in Finnish society.

The second purpose was to test whether the meaning of the Closeness and Conflict constructs is invariant across Finnish teachers from the undergraduate level onward, at different stages of their careers. We found evidence for a partial strong measurement invariance across student teachers, teachers in training, and teachers, but the evidence did not meet the criteria for strong measurement invariance, which is difficult to achieve as it requires equivalence of both factor loadings and item intercepts, that is, scalar invariance. Further inspection revealed that Items 2, 3, and 15 were non-invariant among the three groups, which do not prevent an exploration of equality in factor means and factor variances (Little, 2013). Knowing the psychometric properties of the STRS-SF ensures that researchers can meaningfully compare group differences in self-reported closeness and conflict and relate them to other factors such as teacher and classroom characteristics or teacher training.

Regarding the third purpose of the study, the two groups of qualified ECE teachers did not differ in perceived closeness with the children in the classrooms, which might reflect shared cultural values about high-quality professional competence influencing existing practice. Nor did the groups differ in educational background, child–teacher ratio, or composition of classrooms (number of immigrant and special needs children), as was expected in the Finnish ECEC context. It is interesting, however, that the teachers at the onset of the in-service training perceived their overall relationships with children as being more conflictual compared to those who did not participate in the training. This perceptual difference might reflect the teachers' actual experience of pedagogically demanding classrooms and explain their motivation for participating in the training. This interpretation is consistent with other data showing that the teachers in training, compared to the teachers, reported less working experience and had younger children who needed more attention and support in classrooms than the teachers. The ECE student teachers, on the other hand, perceived their overall relationships with children as less close than did the qualified ECE teachers and less conflictual than did the qualified ECE teachers in training. Given student teachers' limited actual experience in the ECEC context, they might have been less competent in accurately appraising their relational strengths at the beginning of their studies, or less aware of potentially problematic relational processes with diverse children. Some evidence on pre-service and in-service teacher training suggests that reflection on relational processes and support in the classroom context can alter the quality of relationships and, hence, influence on professional development (Sabol & Pianta, 2012). It might be expected that teachers change their perceptions of close relationships and actual behavior with diverse children when they gradually become aware of how their reflections about emotional experiences with the children relate to their actual interactions in the classroom.

Limitations and Future Directions

There are potential sources of bias that can limit the reliability and validity of self-report data, as well as the generalizability of the research findings to the population. In addition, the relatively small sample size, especially for the teachers in training, may have affected the accuracy and stability of the estimated parameters and standard errors (Li, 2016). On average, the participants reported high levels of closeness and low levels of conflict, which might partly be due to social desirability bias. Anonymity might minimize social desirability among the teachers but increase the non-response rate compared to the other two groups. By way of comparison, the response rates of the student teachers and the teachers in training were high, 89.6 and 97.5, respectively. Low response rates, in turn, may lead to sample selection bias if respondents and nonrespondents differ systematically. Due to anonymity, we could not compare the teachers with those who chose not to respond, but, as shown in Table 1, they had more experience and worked with older children compared to the teachers in training. The inclusion of observation data of teacher–children interactions in the classroom would be beneficial in future studies to explore to what extent the

aggregated version of the STRS-SF can be regarded as a generalized construct reflecting the teachers' and student teachers' accurate relational expectations based on real experiences with children. The aggregated version could be useful for researchers and teacher educators in the ECEC pre- and in-service teacher training context to improve professional learning, provide feedback to individuals and groups, such as multi-professional teams, regarding the quality of their relationships with children, and estimate the effect of training on the development of pedagogical competence. It could also be applied in comparative research to explore the similarities and differences in ECE professionals' pedagogical competence within or across daycare centers and municipalities at the national level or between different countries or cultures. According to Ross and Bruce (2007), see also Avalos (2011), self-assessment in combination with other growth fostering tools can not only support the recognition of existing good practice, which can strengthen teachers' competence and self-efficacy beliefs, but also provide information for improvement and facilitate motivation to change. Overall, the Finnish study extended the application of STRS-SF to a Nordic cultural context and added to the existing knowledge about the factorial validity and measurement invariance of this widely used instrument.

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) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research was supported by the Ministry of Education and Culture of the Finnish Government through Grant no. OKM/66/523/2017.

ORCID iD

Wenwen Yang  <https://orcid.org/0000-0003-2629-752X>

References

- Aboagye, M. O., Qin, J., Pekárková, S., Antwi, C. O., Jababu, Y., Asare, K., Affum-Osei, E., & Akinyi, N. (2019). Factorial validity of the student-teacher relationship scale-short form, latent means comparison of teacher-student relationship quality and association with child problem and prosocial behaviours. *Psychological Studies, 64*(2), 221-234. doi:10.1007/s12646-019-00488-0
- Act on Early Childhood Education and Care (2018). <https://www.finlex.fi/en/laki/kaannokset/2018/en20180540.pdf>.
- Avalos, B. (2011). Teacher professional development in teaching and teacher education over ten years. *Teaching and Teacher Education, 27*(1), 10-20. doi:10.1016/j.tate.2010.08.007
- Baker, J. A., Grant, S., & Morlock, L. (2008). The teacher-student relationship as a developmental context for children with internalizing or externalizing behavior problems. *School Psychology Quarterly, 23*(1), 3-15. doi:10.1037/1045-3830.23.1.3
- Birch, S. H., & Ladd, G. W. (1997). The teacher-child relationship and children's early school adjustment. *Journal of School Psychology, 35*(1), 61-79. doi:10.1016/S0022-4405(96)00029-5
- Bowlby, J. (1969). *Attachment and loss Vol. 1: Attachment*. Basic Books.
- Bronfenbrenner, U. (1986). Ecology of the family as a context for human development: Research perspectives. *Developmental Psychology, 22*(6), 723-742.
- Byrne, B. M. (2006). *Structural equation modeling with EQS: Basic concepts, applications, and programming* (2nd ed.). Lawrence Erlbaum Associates Publishers.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 9*(2), 233-255. doi:10.1207/S15328007SEM0902_5

- Downer, J. T., López, M. L., Grimm, K. J., Hamagami, A., Pianta, R. C., & Howes, C. (2012). Observations of teacher-child interactions in classrooms serving Latinos and dual language learners: Applicability of the Classroom Assessment Scoring System in diverse settings. *Early Childhood Research Quarterly*, 27(1), 21-32. doi:[10.1016/j.ecresq.2011.07.005](https://doi.org/10.1016/j.ecresq.2011.07.005)
- Drugli, M. B., & Hjemdal, O. (2013). Factor structure of the student-teacher relationship scale for Norwegian school-age children explored with confirmatory factor analysis. *Scandinavian Journal of Educational Research*, 57(5), 457-466. doi:[10.1080/00313831.2012.656697](https://doi.org/10.1080/00313831.2012.656697)
- Enders, C. K. (2010). *Applied missing data analysis*. Guilford Press.
- Finnish National Agency for Education (2018). *National core curriculum for early childhood education and care 2018*.
- Fraire, M., Longobardi, C., Prino, L. E., Sclavo, E., & Settanni, M. (2013). Examining the student-teacher relationship scale in the Italian context: A factorial validity study. *Electronic Journal of Research in Educational Psychology*, 11(3), 851-882. doi:[10.14204/ejrep.31.13068](https://doi.org/10.14204/ejrep.31.13068)
- Gregoriadis, A., & Tsigilis, N. (2008). Applicability of the student-teacher relationship scale (STRS) in the Greek educational setting. *Journal of Psychoeducational Assessment*, 26(2), 108-120. doi:[10.1177/0734282907306894](https://doi.org/10.1177/0734282907306894)
- Greiff, S., & Scherer, R. (2018). Still comparing apples with oranges? Some thoughts on the principles and practices of measurement invariance testing. *European Journal of Psychological Assessment*, 34(3), 141-144. doi:[10.1027/1015-5759/a000487](https://doi.org/10.1027/1015-5759/a000487)
- 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(2), 625-638. doi:[10.1111/1467-8624.00301](https://doi.org/10.1111/1467-8624.00301)
- Hamre, B. K., Pianta, R. C., Downer, J. T., & Mashburn, A. J. (2008). Teachers' perceptions of conflict with young students: Looking beyond problem behaviors. *Social Development*, 17(1), 115-136. doi:[10.1111/j.1467-9507.2007.00418.x](https://doi.org/10.1111/j.1467-9507.2007.00418.x)
- Howes, C. (2000). Social-emotional classroom climate in child care, child-teacher relationships and children's second grade peer relations. *Social Development*, 9(2), 191-204. doi:[10.1111/1467-9507.00119](https://doi.org/10.1111/1467-9507.00119)
- Hu, L. T., & Bentler, P. M. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 76-99). Sage Publications.
- Jerome, E. M., Hamre, B. K., & Pianta, R. C. (2009). Teacher-child relationships from kindergarten to sixth grade: Early childhood predictors of teacher-perceived conflict and closeness. *Social Development*, 18(4), 915-945. doi:[10.1111/j.1467-9507.2008.00508.x](https://doi.org/10.1111/j.1467-9507.2008.00508.x)
- Kesner, J. E. (2000). Teacher characteristics and the quality of child-teacher relationships. *Journal of School Psychology*, 38(2), 133-149. doi:[10.1016/S0022-4405\(99\)00043-6](https://doi.org/10.1016/S0022-4405(99)00043-6)
- Koomen, H. M. Y., Verschueren, K., van Schooten, E., Jak, S., & Pianta, R. C. (2012). Validating the student-teacher relationship scale: Testing factor structure and measurement invariance across child gender and age in a Dutch sample. *Journal of School Psychology*, 50(2), 215-234. doi:[10.1016/j.jsp.2011.09.001](https://doi.org/10.1016/j.jsp.2011.09.001)
- Li, C.-H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods*, 48(3), 936-949. doi:[10.3758/s13428-015-0619-7](https://doi.org/10.3758/s13428-015-0619-7)
- Little, T. D. (2013). *Longitudinal structural equation modeling*. Guilford Press.
- Meehan, B. T., Hughes, J. N., & Cavell, T. A. (2003). Teacher-student relationships as compensatory resources for aggressive children. *Child Development*, 74(4), 1145-1157. doi:[10.1111/1467-8624.00598](https://doi.org/10.1111/1467-8624.00598)
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58, 525-543. doi:[10.1007/BF02294825](https://doi.org/10.1007/BF02294825)
- Meredith, W., & Teresi, J. (2006). An essay on measurement and factorial invariance. *Medical Care*, 44(11), S69-S77. doi:[10.1097/01.mlr.0000245438.73837.89](https://doi.org/10.1097/01.mlr.0000245438.73837.89)
- Milatz, A., Glüer, M., Harwardt-Heinecke, E., Kappler, G., & Ahnert, L. (2014). The student-teacher relationship scale revisited: Testing factorial structure, measurement invariance and validity criteria in German-speaking samples. *Early Childhood Research Quarterly*, 29(3), 357-368. doi:[10.1016/j.ecresq.2014.04.003](https://doi.org/10.1016/j.ecresq.2014.04.003)
- Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide* (8th ed.).

- Patricio, J. N., Barata, M. C., Calheiros, M. M., & Graça, J. (2015). A Portuguese version of the student-teacher relationship scale—short form. *The Spanish Journal of Psychology*, *18*(30), 1-12. doi:[10.1017/sjp.2015.29](https://doi.org/10.1017/sjp.2015.29)
- Pianta, R. C. (1999). *Enhancing relationships between children and teachers*. American Psychological Association. doi:[10.1037/10314-000](https://doi.org/10.1037/10314-000)
- Pianta, R. C. (2001a). *Student-teacher relationship scale: Professional manual*. Psychological Assessment Resources.
- Pianta, R. C. (2001b). *Student-teacher relationship scale—Short form*. Psychological Assessment Resources.
- Pianta, R. C., & Steinberg, M. (1992). Teacher-child relationships and the process of adjusting to school. *New Directions for Child and Adolescent Development*, *57*, 61-80. doi:[10.1002/cd.23219925706](https://doi.org/10.1002/cd.23219925706)
- Pianta, R. C., & Stuhlman, M. W. (2004). Teacher-child relationships and children's success in the first years of school. *School Psychology Review*, *33*(3), 444-458. doi:[10.1080/02796015.2004.12086261](https://doi.org/10.1080/02796015.2004.12086261)
- Ross, J. A., & Bruce, C. D. (2007). Teacher self-assessment: A mechanism for facilitating professional growth. *Teaching and Teacher Education*, *23*(2), 146-159. doi:[10.1016/j.tate.2006.04.035](https://doi.org/10.1016/j.tate.2006.04.035)
- Sabol, T. J., & Pianta, R. C. (2012). Recent trends in research on teacher-child relationships. *Attachment & Human Development*, *14*(3), 213-231. doi:[10.1080/14616734.2012.672262](https://doi.org/10.1080/14616734.2012.672262)
- Solheim, E., Berg-Nielsen, T. S., & Wichstrøm, L. (2012). The three dimensions of the student-teacher relationship scale: CFA validation in a preschool sample. *Journal of Psychoeducational Assessment*, *30*(3), 250-263. doi:[10.1177/0734282911423356](https://doi.org/10.1177/0734282911423356)
- Sroufe, L. A. (1988). The role of infant-caregiver attachment in development. In J. Belsky, & T. M. Nezworski (Eds.), *Clinical implications of attachment* (pp. 18-38). Lawrence Erlbaum Associates.
- Steenkamp, J. B. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, *25*(1), 78-107. doi:[10.1086/209528](https://doi.org/10.1086/209528)
- Stipek, D., & Byler, P. (2004). The early childhood classroom observation measure. *Early Childhood Research Quarterly*, *19*(3), 375-397. doi:[10.1016/j.ecresq.2004.07.007](https://doi.org/10.1016/j.ecresq.2004.07.007)
- Tsigilis, N., & Gregoriadis, A. (2008). Measuring teacher-child relationships in the Greek kindergarten setting: A validity study of the student-teacher relationship scale-short form. *Early Education and Development*, *19*(5), 816-835. doi:[10.1080/10409280801975826](https://doi.org/10.1080/10409280801975826)
- van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, *9*(4), 486-492. doi:[10.1080/17405629.2012.686740](https://doi.org/10.1080/17405629.2012.686740)
- Webb, M.-y. L., & Neuharth-Pritchett, S. (2011). Examining factorial validity and measurement invariance of the student-teacher relationship scale. *Early Childhood Research Quarterly*, *26*(2), 205-215. doi:[10.1016/j.ecresq.2010.09.004](https://doi.org/10.1016/j.ecresq.2010.09.004)
- Whitaker, R. C., Dearth-Wesley, T., & Gooze, R. A. (2015). Workplace stress and the quality of teacher-children relationships in Head Start. *Early Childhood Research Quarterly*, *30*(Part A), 57-69. doi:[10.1016/j.ecresq.2014.08.008](https://doi.org/10.1016/j.ecresq.2014.08.008)
- Zumbo, B. D., Gadermann, A. M., & Zeisser, C. (2007). Ordinal versions of coefficients alpha and theta for likert rating scales. *Journal of Modern Applied Statistical Methods*, *6*(1), 21-29. doi:[10.22237/jmasm/1177992180](https://doi.org/10.22237/jmasm/1177992180)