Published in final edited form as:

Sch Psychol Q. 2015 March ; 30(1): 23–36. doi:10.1037/spq0000063.

Teacher Network of Relationships Inventory: Measurement Invariance of Academically At-risk Students Across Ages 6 to 15

Jiun-Yu Wu¹ and Jan N. Hughes²

Jiun-Yu Wu: jiunyu.rms@gmail.com; Jan N. Hughes: jhughes@tamu.edu

¹Institute of Education, National Chiao Tung University, 1001 Ta-Hsueh Rd, Hsinchu, 30010, Taiwan

²Department of Educational Psychology, 4225 Texas A&M University, College Station, TX 77843-4225, United States

Abstract

We tested the longitudinal measurement invariance of the Teacher Network of Relationships Inventory (TNRI), a teacher-report measure of teacher-student relationship quality (TSRQ), on a sample of 784 academically at-risk students across ages 6 to 15 years by comparing the model for each subsequent year with that of the previous year(s). The TNRI was constructed with 22 items to form three correlated factors: Warmth, Conflict, and Intimacy. Cronbach's alphas (α) ranged from .87 to .96 across 9 years. Both metric and scalar measurement invariance held for 9 years, indicating that scores on the TNRI have similar meaning across these ages. The TNRI also demonstrated measurement invariance across gender and race/ethnicity. Findings support that the TNRI is an appropriate measure for investigating substantive issues related to developmental changes in TSRQ from early childhood through adolescence, including gender and ethnic/racial differences in TSRQ across these ages. Based on RM-ANOVAs, each scale decreased across the 9 years, although the growth patterns for scales differed somewhat: Conflict had a linearly decreasing pattern, Warmth declined most notably as students make the transition to adolescence, whereas Intimacy scores dropped off noticeably at the transition from early to late childhood. Research limitation and implication for practice are discussed.

Keywords

Teacher-Student Relationship Quality; Teacher Network of Relationships Inventory; longitudinal measurement invariance; multigroup comparison; child development

An extensive body of research documents the importance of a supportive relationship with one's teacher for a student's school success (for narrative review see Hamre & Pianta, 2006; for meta-analytic review see Roorda, Koomen, Spilt, & Oort, 2011). Students at risk for academic failure also benefit from a supportive teacher-student relationship, and in fact are more affected by relational supports at school compared with the normative sample (Hamre & Pianta, 2005 and Hughes et al, 2012). Therefore, research on teacher-student relationship

among the at-risk students is a worthwhile pursuit. In particular, teacher reports of support (i.e., closeness or warmth) and conflict in the teacher-student relationship predict students' social, behavioral, and academic outcomes, above prior levels of outcomes, from preschool through the elementary grades (Hughes, Luo, Kwok, & Loyd, 2008; Ladd, Birch, & Buhs, 1999; Pianta, Steinberg, & Rollins, 1995). After the elementary grades, teacher-student relationship quality (TSRQ) is commonly assessed with student questionnaires, which are also associated with positive academic and behavioral adjustment (Wentzel, 1999). In developing teacher report measures of TSRQ, researchers have drawn from attachment theory (Pianta, 2001), social support theory (Hughes & Kwok, 2007), and self system models of motivation (Furrer & Skinner, 2003). Despite the diversity of theoretical frameworks, teacher report measures of TSRQ consistently identify a supportive dimension (i.e., close, warm, accepting) and a conflict dimension (Hughes & Kwok, 2007; Murray, Murray, & Waas, 2008; Pianta, 2001), with some measures identifying a third dimension of dependency (Pianta & Stuhlman, 2004) or intimacy (Hughes & Kwok, 2007). Both teacher support scores and teacher conflict scores are associated with children's social, emotional, and academic adjustment (Hamre & Pianta, 2001; Meehan, Hughes, & Cavell, 2003).

Recently, researchers have investigated developmental changes in teacher-reported TSRQ across the elementary grades and their implications for student performance. These studies have found that average levels of overall relationship quality tend to decrease across the elementary grades (O'Connor & McCartney, 2007; O'Connor, 2010). When closeness and conflict in relationships were examined separately from kindergarten to grade 6, a decrease in closeness across this period and an increase in conflict in the early grades followed by a decrease in grade 5 were found (Jerome, Hamre, & Pianta, 2009). Surmising that average growth trajectories may obscure subtypes of children who follow atypical trajectories, Spilt, Hughes, Wu, and Kwok (2012) identified subgroups of trajectory classes for warmth and conflict for boys and girls across grades 1 to 5 on a sample of students at risk for school failure. The typical child had high but somewhat declining warmth and low, stable conflict across these years. However, not all children followed these normative trends. Children whose trajectories were characterized by variably high conflict showed lower growth in reading and math across the elementary grades than did students with stable and low conflict. Because these researchers did not test the assumption that scores on their measures have the same meaning across ages (i.e., may not demonstrate longitudinal measurement invariance), conclusions regarding developmental changes in relationship constructs may not be justified.

Measurement invariance of tests of TSRQ

Definition of measurement invariance

In its most general sense, measurement invariance refers to the equivalence of measurement across groups. Groups can be based on any number of factors, such as ethnicity, gender, or age. Testing measurement invariance involves a series of increasingly stringent criteria (Millsap, 2011). Configural invariance is the first criterion, and examines if the number of factors and the pattern of zero factor loadings are the same among groups. Next, metric invariance (also referred to as weak invariance) refers to a finding that factor loadings are

equivalent across groups (i.e., that the relationship between the items and the underlying latent construct is the same across groups). Metric invariance is sufficient for examining the issue of factorial validity of a measure across groups. The next level of invariance is scalar invariance, which refers to a finding that item intercepts are also equivalent across groups. A finding of scalar invariance means that the construct is measured on a similar scale across groups. When both metric and scalar invariance are found, the measure is considered to demonstrate strong invariance. In order to reach valid conclusions regarding observed differences between group means or group differences in the patterns of correlations of the measure with external criteria, a measure must demonstrate strong invariance (Meredith, 1993).

Longitudinal measurement invariance

Longitudinal measurement invariance means that a measure assesses similar constructs in the same way at different ages or assessment waves. Substantive interpretations of observed score differences between different ages are not valid unless the measure demonstrates invariance across ages (Meredith, 1993). Measurement invariance across age is often assumed but rarely tested in studies assessing age-related changes in TSRQ. An exception is a study by Koomen, Verschueren, van Schooten, Jak, and Pianta (2012), which investigated measurement invariance of a Dutch-adaptation of the Student Teacher Relationship Scale (STRS; Pianta, 2001) across ages 3 to 12 and across gender. The STRS, a teacher-report measure derived from literature on parent-child attachment, was originally intended for use with children ages 3 to 8 and yields three scores: Closeness, Conflict, and Dependency. Although the three-factor structure of the STRS was supported between younger and older age levels and gender, neither metric invariance (i.e., equivalence of factor loadings) nor scalar invariance (i.e., equivalence of item intercepts) across age levels was supported. Gender non-invariance problems were found to be minor, indicating scores on the STRS have similar if not identical meanings across gender. The authors concluded that the STRS measures different constructs at different age levels; consequently, scores do not mean the same thing for younger and older students. The authors also concluded that the STRS, which is based on attachment theory, may be more appropriate for children under the age of 8, or even under the age of 6, than for older children.

The primary purpose of this study was to test longitudinal measurement invariance of the Teacher Network of Relationships Inventory (TNRI; Hughes et al., 2008) from age 6 to 15. The TNRI was developed from the Network of Relationships Inventory (NRI; Furman & Buhrmester, 1985), a measure designed to assess children's and adolescents' perceptions of relationship quality across different types of relationships or across different ages or groups. The NRI is based on social support theory (Weiss, 1974) and has been used extensively in research to assess how youths' perceptions of overall support and negative interactions in a given relationships (e.g., best friend or parent) are associated with other individual outcomes (e.g., loneliness or self-concept) or relationship outcomes (e.g., continuation or termination of relationships) (Furman & Buhrmester, 2009). The TNRI was adapted from the child-report NRI and assesses teachers' reports of the provision of various types of social support to students and yields three scales: Warmth, Intimacy, and Conflict. The TNRI has demonstrated good internal consistency and construct and predictive validity in grades 1–5,

with Warmth and Support scores being positively associated with peer acceptance, behavioral engagement, and reading and math achievement and negatively associated with aggression (Meehan, Hughes & Cavell, 2003; Hughes & Kwok, 2007; Hughes et al., 2008; Liew, Chen & Hughes, 2010; Hughes, 2011; Hughes et al., 2012). For example, in a 3-year longitudinal study, TNRI Support scores (which includes Warmth and Intimacy scales) predicted changes in elementary students' academic achievement, an effect that was mediated by the effect of Support scores on children's behavioral engagement in the classroom (Hughes et al., 2008). The TNRI Warmth scale has demonstrated good convergent and discriminant validity with both child and peer reports of teacher warmth and child conduct problems (Li, Hughes, Kwok & Hsu, 2012).

The hypothesis of longitudinal measurement invariance of the TNRI across elementary and middle school is reasonable because the TNRI assesses the provision of sources of social support (e.g., affection, intimacy, enhancement of worth, low conflict) that have been found to be relevant to individuals from early childhood to adulthood (Furman, 1996; Furman & Buhrmester, 2009). The availability of a measure of TSRQ that is invariant across elementary and middle school would facilitate efforts to understand the development of TSRQ across childhood and adolescence, age differences in TSRQ, and differences in correlates of TSRQ across developmental levels. Given the saliency of gender and race/ethnicity both to teacher-student interactions and to educational outcomes, it is also important to investigate invariance of the TNRI across gender and race/ethnicity. The limited research on the role of gender and race/ethnicity in the measurement of TSRQ is summarized next.

Gender and Measurement of TSRQ

Consistently, studies report that teachers report greater closeness and lower conflict in their relationships with girls than with boys (Spilt et al., 2012; Wu, Hughes, Kwok, 2010). However, without a finding of metric and scalar invariance across gender, a conclusion that gender differences in observed scores on a measure of TSRQ reflect differences on the same construct is not warranted. The current study tested measurement invariance of the TNRI across gender in a sample of youth from age 6 to 15.

Race/Ethnicity and Measurement of TSRQ

Researchers have found that teachers report lower levels of support and higher levels of conflict in their relationships with African American students than with Hispanic or White students, who experience similar levels of teacher support and conflict (Hamre & Pianta, 2001; Meehan, Hughes & Cavell, 2003; Hughes & Kwok, 2007). These race and ethnic differences are minimized, however, when teachers and students are matched on race or ethnicity (Saft and Pianta, 2001). To the authors' knowledge, only one study has investigated ethnic or racial invariance of measures of TSRQ. Using CFA procedures, Webb and Neuharth-Pritchett (2011) examined the factorial validity of the STRS and its measurement equivalence across African American and White children in a sample of pre-kindergarten children (age 5). The hypothesized three-factor model of the STRS did not obtain an adequate fit to the whole sample, and the factor structure was not equivalent across

racial groups. Relatively large racial discrepancies were found in patterns of factor loadings for items on the Closeness and Dependency scales. The authors concluded that studies reporting ethnic differences on observed scores of the STRS may not reflect differences in the latent constructs. In the current study, we tested measurement invariance of the TNRI across Hispanic, White, and African American students.

Study Purpose

The primary purpose of the current study was to investigate measurement invariance of the TNRI across ages 6 to 15 (generally corresponding to grades 1 to 9) as well as across gender and ethnicity. We pursue these aims in an ethnically and linguistically diverse, academically at-risk sample of students (see participants). The TNRI was administered to students' teachers annually for 9 years, beginning when the children were in first grade. A finding of strong measurement invariance for the TNRI across age, gender, and race would support its use in substantive investigations of developmental patterns in TSRQ and of gender and race/ethnicity differences in students' relationships with teachers. Upon a finding of longitudinal, ethnic, and gender measurement invariance of the TNRI, the secondary purpose of the study was to describe developmental changes in each construct measured by the TNRI (i.e., Warmth, Intimacy, and Conflict) from age 6 to 15.

Method

Participants

Participants were recruited from three school districts in Texas (1 urban and 2 small cities) across two sequential cohorts in first grade during the fall of 2001 and 2002. Children were eligible to participate in the longitudinal study if they scored below the median score on a state-approved district-administered measure of literacy, spoke either English or Spanish, were not receiving special education services, and had not previously been retained in first grade. The study was conducted under the approval of the Institutional Review Board of the second author's university. School records identified 1,374 children as eligible to participate. Because teachers distributed consent forms to parents via children's weekly folders, the exact number of parents who received the consent forms could not be determined. Small gifts to children and the opportunity to win a larger prize in a random drawing were instrumental in obtaining 1,200 returned consent forms, of which 784 parents (65%) provided consent.

Analyses of a broad array of archival variables, including performance on the district-administered test of literacy, age, gender, ethnicity, eligibility for free or reduced-price lunch, bilingual class placement, cohort, and school context variables (i.e., ethnic composition and percentage of economically disadvantaged students), did not indicate any differences between children with and without consent. The resulting sample of 784 participants (52.6% male) closely resemble the population from which they were drawn on demographic and literacy variables relevant to students' educational performance. The ethnic composition of the study sample was 37% Hispanic (of whom 39% were Spanish language dominant), 34% White, 23% African American, and 6% other; 62% of the children qualified for free or reduced cost lunch. In Year 1, the mean scores on the Woodcock

Johnson III Tests of Achievement (WJ-III; Woodcock McGrew, & Mather, 2001) Broad Reading and Broad Math were 98.25 ($SD = 17.26$) and 101.70 ($SD = 13.80$), respectively.

In Year 1, the 784 students were in 207 classrooms in 36 schools. The sample was followed for 9 years. The effective sample differed by year, based on student attrition and teacher completion of the TNRI. By Year 9, 488 students (55.1% male) were still active in the study. The largest factor in subject attrition was parental non-response to a request for continuation in the study after the first 5 years, at which time parental written consent for continued participation was received for 569 participants. Attrition analysis of Year 9 sample found that attritted participants scored lower on math achievement¹ at Year 1 ($M_{WJ-III\ Math} = -2.09$, $t(754) = 2.08$, $p = .038$), but did not differ on a wide range of other variables (e.g., gender, ethnicity, bilingual education status, economically disadvantaged status, retention status, literacy scores, reading achievement, school engagement, parent education level and parent literacy status)². Across 9 years, on average 243.3 (MIN-MAX, 148–335) teachers from 72 (36–108) schools completed the TRNI and provided demographic information each year. The number of teachers and schools involved each year increased over time due to the increasing dispersion of students outside the original three participating school districts. Most teachers were female (54.3%–87%) and White (58.8%–73.9%), while 1.4%–11.7% of teachers were Hispanic and 1.9%–8.7% were African American. From Year 1 to Year 9, there were 2.5%–6.4% teachers with less than one year of experience, 11%–19.9% with 1–3 years, 11%–15.5% with 4–6 years, 4.5%–11.2% with 7–9 years, 3.5%–8.2% with 10–12 years and 20.4%–37.7% with more than 12 years. As expected for an educationally at-risk sample, some students were retained in grade. By Year 9, 154 (31.5%) students had been retained once and 10 (2.0%) students had been retained twice.

Teacher Network of Relationships Inventory (TNRI)

In the spring of each year, from Year 1 to Year 9, teachers were mailed questionnaires for each study child in their classroom that included the TNRI. Teachers received \$25 for each questionnaire returned. In the elementary grades, students' primary classroom teacher completed the questionnaire. Beginning at the transition to middle school, questionnaires were completed by the student's language arts teacher (93%) or a teacher named by the language arts teacher as having greater knowledge of the student (7%).

As described above, the 22-item TNRI was developed from the child version of the Network of Relationships Inventory (Furman & Buhrmester, 1985). Specifically, NRI items were rephrased so that teachers reported on a 5-point Likert scale (*not at all true* to *very true*) their provision of six types of social provisions (reliable alliance, dependable bond, enhancement of worth, affection, intimacy, and nurturance) as well as negativity and conflict in the relationship. Some items were reworded so that the focus of the item was on the child

¹We used the WJ-III Broad Reading W Scores (Letter-Word Identification, Reading Fluency, Passage Comprehension subtests), the WJ-III Broad Math W Scores (Calculations, Math Fluency, and Math Calculation Skills subtests) and the comparable Spanish version for individuals ages 6 to 15, The Extensive research documents the reliability and construct validity of the WJ-III (Woodcock et al., 2001).

²Detailed tables of attrition analysis of Year 9 sample and normality check of 9-year items can be requested from corresponding author.

rather than on the teacher's behavior in order to reduce the threat of the question and minimize a teacher's tendency to respond in a socially desirable manner. For example, an item on the child version asked "How much does your teacher treat you like you're admired and respected?" The comparable item on the teacher version is "This child gives me many opportunities to praise him/her." Other items were changed minimally (e.g., from "How satisfied are you with your relationship with your teacher?" to "I am satisfied with my relationship with this child"). Earlier exploratory and confirmatory factor analysis (Hughes, Gleason & Zhang, 2005) with the same longitudinal sample in Year 3 identified three factors: Warmth (13 items; $\alpha = .96$), Intimacy (3 items; $\alpha = .86$) and Conflict (6 items; $\alpha = .91$). Based on their high correlation, the Warmth and Intimacy items are sometimes combined to form a Support score (Hughes & Kwok, 2006). As stated previously, the TNRI has demonstrated good convergent and discriminant validity.

Data Analysis Overview

In order to examine the longitudinal stability of the measurement structure of the TNRI across 9 years, a series of measurement invariance (MI) analyses were conducted to sequentially test the configural, metric and scalar invariance assumptions within CFA framework. The normality of TNRI item responses was checked across 9 years. All the skewness and kurtosis values were within the commonly used criteria (e.g. skewness ranging within ± 3 and kurtosis within ± 8 , Kline, 2010)², and the normality assumption held for our responses across 9 years. We followed the practices of prior researchers assessing teacher-student relationships in treating items as continuous responses on a 5-point scale (e.g. Hughes, 2012; Koomen, Verschueren, van Schooten, Jak, & Pianta, 2012; Pianta, 2001; Solheim, Berg-Nielsen, & Wichstrøm, 2012 to name a few), thus permitting comparisons between our results and those of other researchers and allowing us to investigate the complex factorial structures with invariance assumptions (Rhemtulla, Brosseau-Liard, & Savalei, 2012; Stark, Chernyshenko, & Drasgow, 2006). The nested nature of our dataset was taken into consideration by using TYPE = COMPLEX routine in Mplus 6.12 with the rescaled model-fit test statistic and robust standard error estimator (L. K. Muthén & Muthén, 2010; Wu & Kwok, 2012). Participants with scores on the TNRI for at least one assessment wave were included in the analysis, and the missingness of the dataset was handled with full information maximum likelihood (FIML) parameter estimator (Enders, 2010).

We tested the three-factor CFA model from previous research with the TNRI (Hughes, Gleason & Zhang, 2005) with the 9 annual datasets. Besides model fit chi-square test statistics, the commonly used criteria of model fit indices suggested by Hu and Bentler (1999) were used to evaluate the model goodness of fit. Modification indices were used to provide statistical evidence for the unknown underlying relationship among items. An examination of modification indices found that 4 pairs of correlated items had consistently large modification indices across 9 years (i.e., modification index ≥ 40). Item 14 and item 7 both involved praising child; item 20 and item 19 both involved affection/nurturance; item 17 and item 9 both involve general satisfaction with relationship; item 21 and item 6 both involved need to discipline. Similarity in wording or phrasing on these pairs may suggest that correlated residuals are necessary (Brown, 2006). Therefore, we revised the original three-factor CFA with these four pairs of correlated items. This revised three-factor CFA

demonstrated an adequate model goodness of fit across the 9 datasets; the average of model-fit indices were as follow:

$$\overline{CFI}=.928, \overline{TLI}=.918, \overline{RMSEA}=.080 \text{ and } \overline{SRMR}=.047.$$

This revised CFA structure was then used as the factor structure in testing the longitudinal measurement invariance of the TNRI measures across 9 years. First, we examined the configural invariance of measurement structures; second, we tested the metric invariance of factor loadings; last, we tested the scalar invariance of intercepts (Millsap, 2011). For instance, in Table 1, we first confirmed the configural invariance assumption with Model 1.1 and proceeded to test metric invariance. The longitudinal MI analysis was conducted by comparing the model for each subsequent year with that of the previous year(s). This approach has two major advantages. First, we can know if the parameters are the same across 9 years. Second, we can detect when items become non-invariant. To do so, metric invariance between Year 1 and Year 2 was tested ($\Lambda_1 = \Lambda_2$) with the factor loadings of the rest of years being freely estimated in Model 2.1. In Model 2.2, we fixed Year 1 and Year 2 factor loadings to be the same and tested if the loadings are invariant through Year 3 ($\Lambda_1 = \Lambda_2 = \Lambda_3$) with the factor loadings of the following years being freely estimated. The same procedure was applied through Model 2.8, where we fixed all the factor loadings of CFA models for 9 years to be the same. After we confirmed the metric invariance from Model 2.1 to Model 2.8, the same yearly comparison was applied to test scalar invariance from Model 3.1 to Model 3.8.

To evaluate the longitudinal measurement invariance, the Satorra-Bentler scaled differential chi-square test (χ^2) was conducted to take the data nesting structure into consideration (Muthén & Muthén, 2010). Given the sensitivity of the χ^2 to the slight difference between two comparative models for large sample size dataset, four other criteria of invariance analysis were used, including changes in CFIs (CFI; Cheung & Rensvold, 2002), and TLIs (TLI; Little, 1997), Root Deterioration per Restriction Index (RDR; Browne & Du Toit, 1992), and the Expected Cross-Validation Index difference (EVCi; Browne & Cudeck, 1993; Oort, 2009). When CFI $\geq$ 0.92 (Cheung & Rensvold, 2002), TLI $\geq$ 0.95 (Little, 1997), RDR $\leq$.08, and/or the 90% confidence interval of EVCi includes zero (Dudgeon, 2004), it would suggest that the two comparative models were not substantially different from each other. To the author's knowledge, there is not a single criterion that can claim measurement invariance; therefore a finding that the majority of criteria are within the suggested thresholds is interpreted as evidence of measurement invariance (Koomen et al., 2012; Vandenberg & Lance, 2000).

We conducted multi-group comparison analyses to test if the invariance assumptions hold for gender and ethnicity. Three ethnic groups were included in this analysis, African American (AA), Hispanic (HIS) and White (W). In the case of non-invariance, modification indices were used to locate the non-invariant items across different groups.

After we confirmed the invariance assumptions across years, genders, and ethnicities of the TNRI, we examined the developmental pattern of TNRI factors across 9 years. The shift and

the growth trend of factor means of the TNRI scales were tested using RM-ANOVA with Proc GLM in SAS 9.3 (SAS Institute, 2012). RM-ANOVA allowed us to investigate if the TNRI construct means would maintain or shift across 9 years. If significant mean shift was found, we tested the linear, quadratic, and cubic average growth trend of the TNRI scales to see if the decreasing or increasing trends would change linearly or if the trends would level off using the polynomial contrast option in Proc GLM procedure.

Results

Invariance of Model Structure across Years

The model information indicated the revised CFA model with Warmth, Conflict and Intimacy factors and four pairs of corrected items adequately fit the longitudinal data, $\chi^2(1818) = 6220.03$ with $p < .001$, CFI = .926, TLI = .915, RMSEA = .068 and SRMR = .048. Therefore, the configural factor invariance with the same pattern of freely estimated parameters held for 9 year data.

Metric and Scalar Longitudinal Measurement Invariance

The results of the metric and scalar longitudinal measurement invariance tests are reported in Table 1. Though the model fit test statistics (χ^2) indicated that the hypothesized models fit did not exactly fit to the data, all model fit indices were within the commonly used criteria (Hu & Bentler, 1999) and demonstrated adequate goodness of fit of the hypothesized models to the data.

Metric Invariance—For yearly MI analyses from Model 2.1 to Model 2.8, except for Model 2.7, the differential chi-square test statistics were smaller than the critical values, which indicated the hypothesized model equivalence was not rejected. The χ^2 between Model 2.6 and Model 2.7 were significant but the other four model comparison criteria indicated no substantial difference between the factor loadings of Year 8 construct and those of the seven previous years, $\chi^2(19) = 34.69$ with $p = .02$, CFI=0, TLI = 0, $EVCI = .002$ with 90% CI=[-.002, .003], and RDR=.045. Therefore, metric invariance was established and we proceeded to examine the scalar/intercept invariance hypothesis.

Scalar Invariance—When we conducted the analyses of scalar invariance for yearly measurements from Model 3.1 to Model 3.8, the pattern of the model comparison information was similar. For all the yearly MI analyses, the majority of criteria (i.e., CFI 0.02, TLI 0.05, and RDR 0.08) suggested negligible differences between the constrained and baseline models (Koomen et al., 2012; Vandenberg & Lance, 2000). The MI result supported both metric and scalar invariance and suggested that the constructs of TNRI scales were longitudinally equivalent in the sense of strong invariance.

Measurement Invariance across Gender and Ethnicity

Table 2 presents the multigroup comparison result on gender and ethnicity. In Model G.1, we had freely estimated factor loadings and item intercepts for boys and girls. In Model G.2, we fixed factor loadings to be the same for boys and girls but allowed item intercepts to be freely estimated for the two groups. In Model G.3, we further fixed the item intercepts to be

the same for boys and girls. For ethnic group comparison of AA, HIS, and W, we followed the same procedure as that for the gender group comparison and yielded Model E.1, Model E.2, and Model E.3. All models exhibited adequate goodness-of-fit to the data.

Comparing Model G.2 to Model G.1 and Model E.2 to E.1, the χ^2 test results were not statistically significant, revealing that neither gender nor ethnicity moderated the factor loadings in the CFA models for both gender and ethnicity, $\Delta\chi^2_{Metric\ Inv., Gender}(19)=26.09$ with $p=.13$, and $\Delta\chi^2_{Metric\ Inv., Ethnicity}(38)=53.21$ with $p=.05$. The χ^2 tests of scalar invariance hypothesis was not significant for gender but was marginally significant for ethnicity. Nevertheless, except for *EVCI*, the other three model comparison criteria provided evidence for intercept equivalence for ethnic groups, *CFI* = .002, *TLI* = -.002, *EVCI* = .022 with 90% *CI*[.014, .032], and *RDR* = .049. Therefore, we concluded that the TNRI measure is largely invariant across gender and ethnic groups. Coefficient alphas were also adequate across the 9 years, as shown in Table 3, ranging from .869 to .962.

Developmental Changes in Mean Level of TNRI scales

The yearly factor means of TNRI constructs were reported in standardized scores with factor means at Year1 as the baseline, which is equivalent to Cohen's *d* (Cohen, 1992), in Table 3. The growth trends of three scales were depicted in Figure 1. The Warmth factor means were consistent across Year 1 to Year 4 but those of Year 5 to Year 9 were significantly smaller than the Year 1 mean (partial $\eta^2 = 10.1\%$, $F_{(8,6.676)}=12.052$, $p<.001$)³. The cubic growth trend fit the data better with the factor means decreasing after Year 4 and leveling off at Year 8 ($R^2 = 3.6\%$, $F_{(3,4652)}=57.27$, $p<.001$). Though Conflict factor means were consistent for the first three years, they were statistically different from Year 4 to Year 9 (partial $\eta^2=3.5\%$, $F_{(8,6.477)}=12.052$, $p<.001$) and showed a linearly decreasing pattern ($R^2 = 1.2\%$, $F_{(1,4654)}=56.31$, $p<.001$). Like the Warmth scores, the Intimacy factor means exhibited a cubic growth pattern ($R^2 = 13.7\%$, $F_{(3,4652)}=246.54$, $p<.001$). Despite the consistent factor means for the first two years, the Intimacy factor means decreased from Year 3 and leveled off at Year 8 (partial $\eta^2=22.5\%$, $F_{(8,6.768)}=31.08$, $p<.001$).

Discussion

Measurement Invariance

The primary purpose of the current study was to investigate the invariance of the measurement model of the TNRI across grades 1 to 9 (ages 6 – 15) as well as across gender and ethnicity. In each test of invariance, we followed the sequence of steps recommended by Millsap (2011). First, we tested whether the same three-factor structure of the TNRI was a satisfactory representation of dimensions of teacher-student relationship quality across groups (i.e., configural invariance). Next, we applied multi-group invariance analysis across groups to test equivalence of factor loadings (metric invariance) and then equivalence of item intercepts (scalar invariance), using a variety of model comparison statistics. Statistical

³Due to the significant Sphericity test result of 9-year factor scores, we reported Huynh-Feldt corrected *F* values for RM-ANOVA tests of three scales.

support for strong invariance of the TNRI across grades/ages, gender, and a racial/ethnic group was found.

These results mean that observed scores on the TNRI scales have the same or very nearly the same meaning across these groups, which provide evidence to our hypothesis of longitudinal measurement invariance of the TNRI across elementary and middle school based on social support theory (Furman, 1996). Therefore, the TNRI is an appropriate research instrument for answering substantive questions about developmental changes from age 6 to 15 and about gender or racial/ethnic differences in TSRQ among students within this broad age span.

Developmental Shifts in of TNRI Scales

This study is the first to investigate changes in mean levels of TSRQ from the early elementary grades to the end of middle school. Intimacy showed the steepest decline (partial $\eta^2=22.5\%$), followed by Warmth (partial $\eta^2=10.1\%$) and Conflict (partial $\eta^2=3.5\%$). As shown in Figure 1, Intimacy and Warmth had a cubic average growth pattern and Conflict had a linearly decreasing pattern. Findings suggest that the normative decline in positive aspects of the teacher-student relationship reported across the elementary grades (Jerome et al., 2009; O'Connor, 2010) continues in the early years of middle school before leveling off in Year 8, and that the normative decline in conflict observed at the end of elementary school (Jerome et al., 2009) continues through middle school. Prior research has found that middle school students' perceptions of teacher support is a strong predictor of their sense of school belonging and engagement in school (Barber & Olsen, 2004; Niehaus, Rudasill, & Rakes, 2012; Roeser, Eccles, & Sameroff, 1998). Furthermore, low school engagement in middle school is a strong predictor of low academic achievement in high school and dropping out of school (Casillas, Robbins, Allen, Kuo, & Hanson., 2012; Johnson, McGue, & Iacono, 2006). Thus the downward trend in perceived teacher support in middle school underscores the need for school psychologists and other school practitioners to implement school-wide interventions focused on building positive, supportive relationships during middle school. Findings from a recent randomized clinical trial suggests that providing middle school teachers with the opportunity to reflect on individualized feedback on classroom teaching interactions with a supportive consultant results in more positive teacher-student interactions and achievement (Allen, Pianta, Gregory, Mikami, & Lun, 2011).

Study Limitations Implications for Future

Although the study models acceptably fit our longitudinal dataset, findings need to be considered in light of certain limitations. The study sample was selected based on scoring below the median on a district-administered measure of literacy at the beginning of first grade; however, their scores at baseline on the nationally standardized Woodcock Johnson III Broad Reading and Broad Math were similar to those in the standardization sample in terms of the sample mean and standard deviation. Thus study results may not generalize to normally and higher achieving students, because this study did not establish measurement invariance with regard to more diverse sample. Analyses of whole school climate and student behavioral norms that involve a broader population and also use the TNRI must be

approached with caution. Additionally, teacher support was assessed only from the perspective of the teacher. Student report of teacher social support might lead to different conclusions regarding developmental trends. Future research is needed on measurement properties of the TNRI with more diverse samples, including its correspondence with student report of similar relationship constructs.

Conclusion

The TNRI appears to meet the need for a measure of TSRQ with strong invariance across age, gender, and race/ethnicity. Thus, the TNRI is the only measure of TSRQ in the published literature that meets measurement invariance standards for investigations of developmental changes in teacher support from early childhood through early adolescence. Among questions that could be pursued with the TNRI are “Do boys and girls or racial/ethnic groups differ in their growth trajectories for TSRQ across elementary and middle school?” and “What are the implications of Warmth and Conflict trajectories on adolescents’ commitment to school and high school completion?” The TNRI is also appropriate for investigations of age-related changes in the associations between TSRQ and student social and academic outcomes as well as research on the mechanisms that account for these associations at different developmental periods. For example, in the early elementary grades a supportive relationship with one’s teacher may provide a secure basis for children, thereby promoting behavioral self-regulation (Liew, Chen & Hughes, 2010). In middle school, teacher social support may buffer students from the normative decline in identification with and bonding to school (Catalano, Haggerty, Osterle, Fleming, & Hawkins, 2004; Liljeberg, Eklund, Fritz, & Klinteberg, 2011).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by Grant R01 HD39367 to Jan Hughes from the National Institute of Child Health and Human Development, USA, and Grant 101-2628-H-009-003-MY3 to Jiun-Yu Wu from the National Science Council, Taiwan. The authors are grateful to the editor and anonymous referees for their thoughtful comments and suggestions in revising this article, and Drs. Helma Koomen, Suzanne Jak and Bellinda King-Kallimanis for their help in providing the statistical program.

References

- Allen JP, Pianta RC, Gregory A, Mikami AY, Lun J. An interaction-based approach to enhancing secondary school instruction and student achievement. *Science*. 2011; 333(6045):1034–1037. [PubMed: 21852503]
- Barber BK, Olsen JA. Assessing the transitions to middle and high school. *Journal of Adolescent Research*. 2004; 19(1):3–30.10.1177/0743558403258113
- Brown, TA. *Confirmatory Factor Analysis for applied research*. New York, NY: Guilford Press; 2006.
- Browne, MW.; Cudeck, R. Alternative ways of assessing model fit. In: Bollen, KA.; Long, JS., editors. *Testing structural equation models*. Newbury Park, CA: SAGE; 1993. p. 136-162.
- Browne MW, Du Toit SHC. Automated fitting of nonstandard models. *Multivariate Behavioral Research*. 1992; 27:269–300.10.1207/s15327906mbr2702_13

- Casillas A, Robbins S, Allen J, Kuo Y, Hanson MA, Schmeiser C. Predicting early academic failure in high school from prior academic achievement, psychosocial characteristics, and behavior. *Journal of Educational Psychology*. 2012; 104(2):407–420.10.1037/a0027180
- Catalano RF, Haggerty KP, Oesterle S, Fleming CB, Hawkins JD. The importance of bonding to school for healthy development: Findings from the social development research group. *The Journal of School Health*. 2004; 74:252–261.10.1111/j.1746-1561.2004.tb08281.x [PubMed: 15493702]
- Cheung GW, Rensvold RB. Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*. 2002; 9:233–255.10.1207/S15328007SEM0902_5
- Cohen J. A power primer. *Psychological bulletin*. 1992; 112(1):155–159. [PubMed: 19565683]
- Dudgeon P. A note on extending Steiger's (1998) multiple sample RMSEA adjustment to other noncentrality parameter-based statistics. *Structural Equation Modeling: A Multidisciplinary Journal*. 2004; 11(3):305–319.10.1207/s15328007sem1103_1
- Enders, CK. *Applied missing data analysis*. New York, NY: The Guilford Press; 2010.
- Furman, W. The measurement of children and adolescents' perceptions of friendships: Conceptual and methodological issues. In: Bukowski, WM.; Newcomb, AF.; Hartup, WW., editors. *The Company They Keep: Friendships in Childhood and Adolescence*. Cambridge, MA: Cambridge University Press; 1996. p. 41-65.
- Furman W, Buhrmester D. Children's perceptions of the personal relationships in their social networks. *Developmental Psychology*. 1985; 21:1016–1024.10.1037//0012-1649.21.6.1016
- Furman W, Buhrmester D. The Network of relationships inventory: Behavioral systems version. *International Journal of Behavioral Development*. 2009; 33:470–478.10.1177/0165025409342634 [PubMed: 20186262]
- Furrer C, Skinner EA. Sense of relatedness as a factor in children's academic engagement and performance. *Journal of Educational Psychology*. 2003; 95(1):148–162.
- Hamre BK, Pianta RC. Early teacher-child relationships and the trajectory of children's school outcomes through eighth grade. *Child Development*. 2001; 72:625–638.10.1111/1467-8624.00301 [PubMed: 11333089]
- Hamre BK, Pianta RC. Can instructional and emotional support in the first-grade classroom make a difference for children at risk of school failure? *Child Development*. 2005; 76:949–967.10.1111/j.1467-8624.2005.00889.x [PubMed: 16149994]
- Hamre, BK.; Pianta, RC. Student-teacher relationships. In: Bear, GC.; Minke, KM., editors. *Children's needs III: Development, prevention, and intervention*. Washington, DC: National Association of School Psychologists; 2006. p. 59-71.
- Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*. 1999; 6(1):1–55.10.1080/10705519909540118
- Hughes JN. Longitudinal effects of teacher and student perceptions of teacher-student relationship qualities on academic adjustment. *Elementary School Journal*. 2011; 112:38–60.10.1086/660686 [PubMed: 21984843]
- Hughes JN. Teacher-student relationships and school adjustment: Progress and remaining challenges. *Attachment and Human Development*. 2012; 14(3):319–327. [PubMed: 22537527]
- Hughes JN, Gleason K, Zhang D. Relationship influences on teachers' perceptions of academic competence in academically at-risk minority and majority first-grade students. *Journal of School Psychology*. 2005; 43:303–320.10.1016/j.jsp.2005.07.001 [PubMed: 20401185]
- Hughes JN, Kwok O. Classroom engagement mediates the effect of teacher-student support on elementary students' peer acceptance: A prospective analysis. *Journal of School Psychology*. 2006; 43(6):465–480.10.1016/j.jsp.2005.10.001 [PubMed: 20431706]
- Hughes JN, Kwok O. The influence of student-teacher and parent-teacher relationships on lower achieving readers' engagement and achievement in the primary grades. *Journal of Educational Psychology*. 2007; 99:39–51. [PubMed: 18084625]
- Hughes JN, Luo W, Kwok O, Loyd LK. Teacher-student support, effortful engagement, and achievement: A 3-year longitudinal study. *Journal of Educational Psychology*. 2008; 100:1–14.10.1037/0022-0663.100.1.1 [PubMed: 19578558]

- Hughes JN, Wu J, Kwok O, Villarreal V, Youngman AY. Indirect effects of child reports of teacher-student relationship on achievement. *Journal of Educational Psychology*. 2012; 104:350–365.10.1037/a0026339 [PubMed: 23226873]
- Jerome EM, Hamre BK, Pianta RC. Teacher-child relationships from kindergarten to sixth grade: Early childhood predictors of teacher-perceived conflict and closeness. *Social Development*. 2009; 18:915–945.10.1111/j.1467-9507.2008.00508.x [PubMed: 24991084]
- Johnson W, McGue M, Iacono WG. Genetic and environmental influences on academic achievement trajectories during adolescence. *Developmental Psychology*. 2006; 42:514–532.10.1037/0012-1649.42.3.514+ [PubMed: 16756442]
- Kline, RB. Principles and practice of structural equation modeling. 3. New York, NY: The Guilford Press; 2010.
- Koomen HMY, Verschueren K, van Schooten E, Jak S, Pianta RC. 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*. 2012; 50:215–234.10.1016/j.jsp.2011.09.001 [PubMed: 22386121]
- Ladd GW, Birch SH, Buhs ES. Children's social and scholastic lives in kindergarten: Related spheres of influence? *Child Development*. 1999; 70:1373–1400.10.1111/1467-8624.00101 [PubMed: 10621962]
- Li Y, Hughes JN, Kwok O, Hsu H. Evidence of convergent and discriminant validity of child, teacher, and peer reports of teacher-student support. *Psychological Assessment*. 2012; 24:54–65.10.1037/a0024481 [PubMed: 21767024]
- Liew J, Chen Q, Hughes JN. Child effortful control, teacher-student relationships, and achievement in academically at-risk children: Additive and interactive effects. *Early Childhood Research Quarterly*. 2010; 25:51–64.10.1016/j.ecresq.2009.07.005 [PubMed: 20161421]
- Liljeberg JF, Eklund JM, Fritz MV, Klinteberg B. Poor school bonding and delinquency over time: Bidirectional effects and sex differences. *Journal of Adolescence*. 2011; 34:1–9.10.1016/j.adolescence.2010.03.008 [PubMed: 20417551]
- Little TD. Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research*. 1997; 32(1):53–76.10.1207/s15327906mbr3201_3
- Meehan BT, Hughes JN, Cavell TA. Teacher-student relationships as compensatory resources for aggressive children. *Child Development*. 2003; 74:1145–1157.10.1111/1467-8624.00598 [PubMed: 12938710]
- Meredith W. Measurement invariance, factor analysis and factorial invariance. *Psychometrika*. 1993; 58(4):525–543.10.1007/BF02294825
- Millsap, RE. Statistical approaches to measurement invariance. New York, NY: Routledge; 2011.
- Murray C, Murray KM, Waas GA. Child and teacher reports of teacher-student relationships: Concordance of perspectives and associations with school adjustment in urban kindergarten classrooms. *Journal of Applied Developmental Psychology*. 2008; 29(1):49–61.10.1016/j.appdev.2007.10.006
- Muthén, LK.; Muthén, BO. Mplus user's guide. 6. Los Angeles, CA: Muthén & Muthén; 2010.
- Niehaus K, Rudasill KM, Rakes CR. A longitudinal study of school connectedness and academic outcomes across sixth grade. *Journal of School Psychology*. 2012; 50(4):443–460.10.1016/j.jsp.2012.03.002 [PubMed: 22710015]
- O'Connor E. Teacher-child relationships as dynamic systems. *Journal of School Psychology*. 2010; 48:187–218. [PubMed: 20380947]
- O'Connor E, McCartney K. Examining teacher-child relationships and achievement as part of an ecological model of development. *American Educational Research Journal*. 2007; 44:340–369.10.3102/0002831207302172
- Oort FJ. Three-mode models for multitrait-multimethod data. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*. 2009; 5(3):78–87.10.1027/1614-2241.5.3.78
- Pianta, RC. Student-Teacher Relationship Scale: Professional manual. Lutz, FL: Psychological Assessment Resources; 2001.

- Pianta RC, Steinberg MS, Rollins KB. The first two years of school: Teacher-child relationships and deflections in children's classroom adjustment. *Development and Psychopathology*. 1995; 7:295–312.10.1017/S0954579400006519
- Pianta RC, Stuhlman MW. Teacher-child relationships and children's success in the first years of school. *School Psychology Review*. 2004; 33(3):444–458.
- Roeser RW, Eccles JS, Sameroff AJ. Academic and emotional functioning in early adolescence: Longitudinal relations, patterns, and prediction by experience in middle school. *Development and Psychopathology*. 1998; 10(2):321–352. [PubMed: 9635227]
- Rhemtulla M, Brosseau-Liard P, Savalei V. How many categories is enough to treat data as continuous? A comparison of robust continuous and categorical SEM estimation methods under a range of non-ideal situations. *Psychological Methods*. 2012 Advance online publication. 10.1037/a0029315
- Roorda DL, Koomen HMY, Spilt JL, Oort FJ. The influence of affective teacher–student relationships on students' school engagement and achievement: A meta-analytic approach. *Review of Educational Research*. 2011; 81:493–529.10.3102/0034654311421793
- Saft EW, Pianta RC. Teachers' perceptions of their relationships with student: Effects of child age, gender, and ethnicity of teachers and children. *School Psychology Quarterly*. 2001; 16:125–141.
- SAS Institute. SAS 9.3 [Computer software]. Cary, NC: SAS Institute, Inc; 2012. Spilt JL, Hughes JN, Wu J, Kwok O. Dynamics of teacher-student relationships: Stability and change across elementary school and the influence on children's academic success. *Child Development*. 2012; 83:1180–1195.10.1111/j.1467-8624.2012.01761.x [PubMed: 22497209]
- Stark S, Chernyshenko OS, Drasgow F. Detecting differential item functioning with confirmatory factor analysis and item response theory: Toward a unified strategy. *Journal of Applied Psychology*. 2006; 91(6):1292–1306.10.1037/0021-9010.91.6.1292. [PubMed: 17100485]
- Solheim E, Berg-Nielsen TS, Wichstrøm L. The three dimensions of the student–teacher relationship scale: CFA validation in a preschool sample. *Journal of Psychoeducational Assessment*. 2011 0734282911423356. 10.1177/0734282911423356
- Vandenberg RJ, Lance CE. A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*. 2000; 3(1):4–70.10.1177/109442810031002
- Webb ML, Neuharth-Pritchett S. Examining factorial validity and measurement invariance of the Student–Teacher Relationship Scale. *Early Childhood Research Quarterly*. 2011; 26(2):205–215.10.1016/j.ecresq.2010.09.004
- Weiss, RS. The provisions of social relationships. In: Rubin, Z., editor. *Doing unto others*(pp.). Englewood Cliffs, NJ: Prentice-Hall; 1974.
- Wentzel KR. Social-motivational processes and interpersonal relationship: Implications for understanding motivation at school. *Journal of Educational Psychology*. 1999; 91:76–97.10.1037//0022-0663.91.1.76
- Woodcock, RW.; McGrew, KS.; Mather, N. Woodcock-Johnson III Tests of Achievement. Itasca, IL: Riverside Publishing; 2001.
- Wu J, Hughes JN, Kwok O. Teacher-student relationship quality type in elementary grades: Effects on trajectories for achievement and engagement. *Journal of School Psychology*. 2010; 48(5):357–387.10.1016/j.jsp.2010.06.004 [PubMed: 20728688]
- Wu J, Kwok O. Using structural equation modeling to analyze complex survey data: A comparison between design-based single-level and model-based multi-level approaches. *Structural Equation Modeling-A Multidisciplinary Journal*. 2012; 19(1):16–35.10.1080/10705511.2012.634703

Figure 1.

The development change in mean Level of TNRI scales. From Year 1 to Year 9, Warmth and Intimacy show cubic growth pattern, while Conflict demonstrates linearly decreasing trend. The solid regression line best fitted the means of factor scores shown as in circles.

Table 1
Model fit test statistics and fit indices and their changes for the measurement models of Year 1 through Year 9

Model fit test statistics and fit indices							Change of model fit test statistics & fit indices					
	χ^2	df	CFI	TLI	RMSEA	SRMR	χ^2	df	CFI	TLI	EVCI	RDR
Step 1: Configural Invariance												
1.1	6220.03	1818	.926	.915	.068	.048	—	—	—	—	—	—
Step 2: Metric Invariance												
2.1 $\Lambda_1 = \Lambda_2$	6243.64	1837	.925	.916	.068	.049	23.61	19	.001	-.001	-.0002	.018
2.2 $\Lambda_1 = \dots = \Lambda_3$	6281.10	1856	.925	.916	.068	.049	19.65	19	0	0	-.0001	.035
2.3 $\Lambda_1 = \dots = \Lambda_4$	6319.09	1875	.925	.917	.068	.051	18.60	19	0	-.001	0	.036
2.4 $\Lambda_1 = \dots = \Lambda_5$	6357.22	1894	.925	.917	.067	.051	10.42	19	0	0	0	.036
2.5 $\Lambda_1 = \dots = \Lambda_6$	6357.22	1913	.925	.918	.067	.052	16.08	19	0	-.001	-.004	.002
2.6 $\Lambda_1 = \dots = \Lambda_7$	6396.41	1932	.924	.919	.067	.053	21.19	19	.001	-.001	.0002	.037
2.7 $\Lambda_1 = \dots = \Lambda_8$	6445.90	1951	.924	.919	.067	.055	34.69	19	0	0	.002	.045
2.8 $\Lambda_1 = \dots = \Lambda_9$	6475.03	1970	.924	.920	.066	.056	29.13	19	0	-.001	-.001	.026
Step 3: Scalar Invariance												
3.1 $\tau_1 = \tau_2$	6512.04	1989	.923	.920	.066	.056	—	—	—	—	—	—
3.2 $\tau_1 = \dots = \tau_3$	6593.60	2008	.922	.920	.066	.056	87.78	19	.001	0	.006	.065
3.3 $\tau_1 = \dots = \tau_4$	6678.76	2027	.921	.919	.067	.056	92.60	19	.001	.001	.007	.067
3.4 $\tau_1 = \dots = \tau_5$	6762.61	2046	.920	.919	.067	.057	88.08	19	.001	0	.007	.066
3.5 $\tau_1 = \dots = \tau_6$	6842.98	2065	.919	.919	.067	.057	86.28	19	.001	0	.006	.064
3.6 $\tau_1 = \dots = \tau_7$	6929.38	2084	.917	.917	.068	.058	88.57	19	.002	.002	.009	.073
3.7 $\tau_1 = \dots = \tau_8$	7029.21	2103	.917	.918	.067	.059	103.26	19	0	-.001	.010	.078
3.8 $\tau_1 = \dots = \tau_9$	7125.00	2122	.915	.917	.068	.060	212.39	19	.002	.001	.008	.072

Note. Δ : factor loading matrix; τ : item intercept vector; the subscript indicates the year of the measures collected.

Bold font indicates the difference between two comparative models is statistically negligible.

Table 2

Model fit test statistics and fit indices and their changes for the measurement models for gender and ethnicity

	Model fit test statistics and fit indices						Change of model fit test statistics & fit indices					
	χ^2	df	CFI	TLI	RMSEA	SRMR	χ^2	df	CFI	TLI	EVCI	RDR
Gender												
G.1 Λ_Q	Λ_Q	τ_Q	4102.04	404	.931	.921	.063	.047	—	—	—	—
G.2 $\Lambda_Q = \Lambda_Q \& \tau_Q$	τ_Q	4138.13	423	.930	.924	.061	.048	26.09	19	.001	-.0003	.020
G.3 $\Lambda_Q = \Lambda_Q \& \tau_Q = \tau_Q$	τ_Q	4163.60	442	.930	.927	.060	.048	2.28	19	0	-.0003	.012
Ethnicity												
E.1 Λ_{AA}	Λ_{HS}	$\Lambda_W \& \tau_{HS}$	4110.97	606	.934	.924	.062	.045	—	—	—	—
E.2 $\Lambda_{AA} = \Lambda_{HS} = \Lambda_W \& \tau_{AA}$	τ_{HS}	4206.83	644	.932	.927	.061	.048	53.21	38	.002	-.0003	.032
E.3 $\Lambda_{AA} = \Lambda_{HS} = \Lambda_W \& \tau_{AA} = \tau_{HS} = \tau_W$	τ_W	4382.12	682	.930	.929	.060	.049	163.77	38	.002	-.002	.049

Note. Λ represents the factor loading matrix; τ is the item intercept vector; the subscript indicates the group of the measures collected; AA = African American; HS = Hispanic; W=Whites.

G.1/E.1: freely estimated Λ and τ for gender and ethnicity group; G.2/E.2: fixed Λ to be the same and freely estimated τ for gender and ethnicity group; G.3/E.3: fixed Λ and τ to be the same for gender and ethnicity group;

Bold font value indicates the difference between two comparative models is statistically negligible.

Table 3

The factor means, correlations and Cronbach's alpha of TNRI scales across 9 years.

Factor correlations						Factor means (κ)						Reliability (α)														
Conf with Warf			Intf with Warf			Intf with Conf			Warf mean			Conf mean			Intf mean			Warf			Conf			Intf		
Est.	SE		Est.	SE		Est.	SE		Est.	SE		Est.	SE		Est.	SE		Est.	SE		Est.	SE		Est.	SE	
Year1	-.446	**	.038	.376	**	.035	.032	**	-.112	.000	—	.000	.000	—	.000	.000	—	.000	.000	—	.000	.000	—	.000	.000	—
Year2	-.470	**	.042	.384	**	.038	.034	*	-.077	.034	-.070	.048	-.035	.045	-.062	.058	.951	.916	.880							
Year3	-.378	**	.040	.407	**	.038	.030	*	-.069	.030	-.058	.050	-.077	.045	-.220	.058	.956	.916	.863							
Year4	-.385	**	.042	.393	**	.039	.030	*	-.077	.030	-.093	.052	-.114	.044	-.318	.060	.950	.909	.869							
Year5	-.379	**	.039	.412	**	.039	.034		-.022	.034	-.145	.051	-.121	.044	-.402	.060	.944	.926	.901							
Year6	-.434	**	.047	.323	**	.042	.031		-.042	.031	-.262	.057	-.157	.046	-.725	.060	.960	.937	.900							
Year7	-.459	**	.055	.324	**	.042	.031		.014	.031	-.393	.059	-.158	.048	-.857	.059	.960	.937	.904							
Year8	-.391	**	.045	.420	**	.046	.029		-.016	.029	-.532	.062	-.201	.044	-.935	.059	.962	.923	.913							
Year9	-.293	**	.037	.263	**	.036	.021		-.016	.021	-.438	.058	-.253	.043	-.916	.057	.949	.919	.886							

Note. Est. = parameter estimate; SE = Standard error; Conf = Conflict factor; Warf = Warmth factor; Intf = Intimacy factor. Factor means (κ) reported were computed from standardized factor scores with Year1 factor mean as the baseline, which can be interpreted as Cohen's d (Cohen, 1992).

* p .05,
 ** p .01.