Research Article doi: 10.12973/eu-jer.10.2.957

European Journal of Educational Research

Volume 10, Issue 2, 957 - 973.

ISSN: 2165-8714

https://www.eu-jer.com/

A Rasch Model Analysis of the Psychometric Properties of the Student-Teacher Relationship Scale among Middle School Students

Amal Alhadabi[*]
Kent State University, USA

Said Aldhafri
Sultan Qaboos University, Oman

Received: April 4, 2020 ▪ Revised: July 8, 2020 ▪ Accepted: December 16, 2020

Abstract: The current study investigated Student-Teacher Relationship Measure (STRM) psychometric properties using Rasch analysis in a sample of middle school female students (N = 995). Rasch Principal Components Analysis revealed psychometric support of two subscales (i.e., Academic and Social Relations). Summary statistics showed good psychometric properties. The category structure and individual statistics (i.e., items and person infit and outfit) were not ideal. Category structure showed that the distances between adjacent thresholds were lower than optimal criteria. Even though findings indicated that items mean square statistics (MNSQ) were optimal, standardized fit statistics (i.e., ZSTD) reflected many misfit persons and items in each subscale. After eliminating the misfit persons and items, the two subscales met the Rasch optimal criteria. The updated short 22-item scale had good psychometric properties, high item and person separation, and good item and person reliability for the two subscales and can be used as a reliable and valid scale.

Keywords: *Student-teacher relationship, psychometric properties, Rasch analysis, adolescents, middle and high school.*

To cite this article: Alhadabi, A., & Aldhafri, S. (2021). A Rasch model analysis of the psychometric properties of the student-teacher relationship scale among middle school students. *European Journal of Educational Research*, *10*(2), 957-973. https://doi.org/10.12973/eu-jer.10.2.957

Introduction

Students' perceptions about the quality of their relationships with teachers influence learning outcomes and academic engagement. A meta-analysis study classified the student-teacher relationship (STR) as the eleventh most vital factor associating with learning outcomes (Hattie, 2009). Literature has articulated that positive relations foster constructive students' growth, as projected by higher academic performance, stronger motivation, greater engagement, and better social-emotional adjustment (Greogory et al., 2014; Hughes, 2012; Ridwan et al., 2014). At the subject level (i.e., STEM subjects), the STR associates with student's ability to deal with difficult learning tasks in math and science (Mikk et al., 2016). In particular, the relationship with science teachers has a critical role among female students, owing to females having a weaker science identity than boys (Hill et al., 2018). Thus, positive relations with science teachers might empower females' science identity and enrollment in STEM majors, particularly in the Middle East.

An in-depth examination of previous studies investigating STR has stressed many points. First, a burgeoning research line has revealed a diversity of theoretical frameworks investigating STR (e.g., Attachment Theory, Sociocultural Theory, Ecological Theory, Social Cognitive Theory, and Developmental Systems Theory; Bandura, 1986; Bowlby, 1969; Bronfenbrenner, 1994; Pianta, 2001; Vygotsky, 1978). Second, many approaches have been adopted in measuring this construct (i.e., assessing teachers' views, students' perspectives, observations, and case studies). Related to student perspectives, which is the main interest in the current study, many scales assess this construct. Examples of these scales are the student-teacher relationship measure (STRM) (Aldhafri & Alhadabi, 2019), the emotional quality scale of the relatedness questionnaire (Lynch & Cicchetti, 1997), the quality of teacher-student relationship scale (Davis, 2001); the network of relationships inventory (Meehan et al., 2003); and inventory of teacher-student relationships (Murray & Zvoch, 2011).

Correspondingly, psychometric studies have illustrated manifold factorial-solutions. Given that, literature has supported various factorial solutions, including two-factor (i.e., academic and social relations; Aldhafri & Alhadabi, 2019), three-factor (e.g., closeness, dependency, and conflict; Pianta, 2001), and four-factor (i.e., recognition,

[*] **Corresponding author:**
Amal Alhadabi, College of Education, Health, and Human Services- School of Foundations, Leadership, and Administration, Kent State University, USA. ✉ aalhadab@kent.edu

commitment, familiarity, and respect; Cranley-Gallagher & Mayer, 2006). The majority of the above-mentioned well-established scales examined kindergarten and elementary school students (Lynch & Cicchetti, 1997; Murray & Zvoch, 2011). Student-Teacher Relationship Measure (STRM) is the only scale, to the author's knowledge, that measures these relations from the perspective of middle and high school students in the Arabic context (i.e., 7th-12th grade students; Aldhafri & Alhadabi, 2019).

Aldhafri and Alhadabi (2019), in their recent study, extensively examined the psychometric properties of STRM. The findings showed that STRM had a two-factor structure (i.e., academic relations and social relations) with good reliability coefficients. Evidences of first-order and second-order factorial structures were obtained. Only one level of measurement invariance (i.e., configural invariance) was supported. However, other levels of measurement invariance were not established, suggesting the need for further refinement and modification. For the most part, the STRM's psychometric properties were assessed using Classical Test Theory (CTT). Statistical literature has acknowledged many limitations of CTT (Crocker & Algina, 2008). These limitations lead to well-known drawbacks regarding the robustness of scales' psychometric properties and the accuracy of inferences drawn from the CTT-developed scales (Bond & Fox, 2015). Some of these limitations are: (1) CTT statistics are dependent on scale length and sample characteristics (Allen & Yen, 1979), (2) The central unit of testing is the full scale by providing one total-item scale correlation coefficient between each item and the total score (Crocker & Algina, 2008), and (3) No examination of items and persons misfit is conducted (Bond & Fox, 2015).

In a nutshell, prior research has revealed many concerns regarding the STRM's psychometric soundness, highlighting the need for assessing the scale psychometric properties using a more precise methodological approach. Item Response Theory (IRT), as a more recent and advanced methodological framework, can achieve this task by providing more informative parameters (e.g., item-difficulty, person-ability, and item-discrimination parameters) that are not sample-dependent and provide accurate estimates of misfit at the items and persons levels (De Ayala, 2009). The optimal IRT model selection depends on the items, response scale, and the number of estimated parameters in the IRT models. Rasch Analysis is a one-parameter logistic IRT model, mainly Rating Scale Model (RSM), is appropriate to analyze the low-stake scales like STRM because the items have the same response scale structure (i.e., five-point Likert scale) and category weights (Andrich, 1978; Linacre, 2017), aiming to provide a more sound and shorter version of the STRM.

Estimating accurate measures assessing the quality of relations between middle school students and their science teachers is a prime concern in Oman by establishing a sounded and abbreviated STRM scale. Omani Ministry of Education aims to prepare students for the fourth industrial revolution (Al Harthy, 2019; Al-Rubaie, 2019). Omani eighth-grade students, along with students in other developing countries in the Middle East, score (M = 431) in TIMSS science tests placed Oman at the middle to bottom ranks (Mullis et al., 2020). This placement is inconsistent with fourth industrial revolution requirements, raising a red flag and genuine concern. Hence, establishing a sounded scale measuring the external factors' role in science achievement, and in this study, the student-teacher relationship, was justified. Particularly, examining the perspective of female adolescents in rural middle schools was necessitated. Therefore, this paper aimed to investigate the psychometric properties of the Student-Teacher Relationship Measure (STRM) using Rasch Analysis.

Literature Review

Student-Teacher Relation (STR)

The STR is defined as the social and academic relations between a teacher and students, considering the teacher's personal and instructional characteristics (Aldhafri & Alhadabi, 2019). In other words, the teacher's personal features (e.g., warmth, caring, promoting trust, and establishing a safe learning environment); as well as instructional characteristics (e.g., considering students' learning styles, applying appropriate classroom management styles, and motivating students) contribute to the formation of positive students' outcomes (i.e., cognitive, behavioral, and social outcomes). A series of recent studies has indicated that constructive STR reinforce students' learning and social adjustment, resulting in fruitful academic and social growth (Greogory et al., 2014; Lee, 2012; Northup, 2011).

Related to academic growth, literature has revealed that positive STR correlated with higher grades (Lee, 2012), intellectual engagement (Gregory et al., 2014), learning motivation (Aldhafri & Alhadabi, 2019; Ridwan et al., 2014). A similar research line has found significant differences in the reading scores based on the quality of STRM (i.e., close or conflict), associating with externalizing behavior problems (i.e., overactive, impulsive, or aggressive behaviors; Baker et al., 2008). For instance, students who had warmth STR scored better in reading achievement than the misbehaved students who experienced substantial conflict in the relations with their teachers. Another meta-analysis study (n = 99 studies from 1990 to 2011) reviewed the associations between STR and two academic outcomes (i.e., engagement and achievement; Roorda et al., 2011). In this meta-analysis study, 63 studies examined the STR among lower grades' students (i.e., preschool, kindergarten, and elementary schools), and 31 studies assessed it among higher grades (i.e., secondary schools). Findings revealed that engagement and achievement positively correlated with constructive STR. However, the associations' strength varied across the academic outcome (i.e., engagement and achievement) and school (i.e., primary and secondary grades). Meaning, STR had a weaker association with achievement relative to engagement.

Alongside this, the effects of positive STR on engagement and achievement were more potent in the secondary schools relative to primary schools.

In contrast, STR's pessimistic influence on the academic outcomes has been explored in prior studies (Baker et al., 2008; Brewster & Bowen, 2004; DiLalla et al., 2004). Baker et al. (2008) revealed that positive work habits among students who had conflicted STR were more inadequate relative to those who had warmth STR, particularly those who had internalizing behavior problems (e.g., depression, anxiety, and social withdrawal). Negative STR associated with low grades and school dropout, especially among at-risk students (Brewster & Bowen, 2004). Another longitudinal study supported this finding by predicting the adolescents' academic achievement conditioning on STR quality at preschool students (i.e., at age five years old; DiLalla et al., 2004). It revealed that students who had negative STR (i.e., more dependent or conflicting STR) got significantly lower grades during adolescence.

On the other hand, effective relations have promoted healthy students' social/behavioral outcomes, including self-concept (McFarland et al., 2016), personal/school adjustment (Baker, 2006), and gaining social skills (Berry & O'Connor, 2010). Prior studies revealed that good STR correlated with a lower level of aggression (Hughes et al., 2008), less discipline problems (Sáez et al., 2012), better subjective wellbeing (Suldo et al., 2014), and more adaptive emotional functioning (Reddy et al., 2003). It also reduced the developmental vulnerabilities, external behavioral problems, and social-emotional issues (e.g., shyness, anxiety, school avoidance, and social withdrawal; Silver et al., 2005). Another longitudinal study examined the social skill growth among a relatively large sample (N = 1,364) that followed students from kindergarten to sixth grade (Berry & O'Connor, 2010). Findings demonstrated that students who had more positive STR showed more productive social skills growth than students with poorer STR.

Contrarily, adverse STR hinder students' social development. Previous studies showed that negative STR (i.e., high level of conflict) significantly associated with an increase in conduct problems (e.g., often fights with other children) and hyperactive behaviors (e.g., Restless and overactive) among middle and high school students (Longobardi et al., 2016). A more recent study investigated the role of the STR during three school transitions (i.e., from kindergarten to elementary school, from elementary to middle school, and from middle to high school; Longobardi et al., 2019). Results revealed a significant association between more conflicted STR and an increase in the externalizing behavioral problems. Meaning, students who experienced a rise in STR conflict showed an exaggeration of externalizing problems, particularly during the first year of the new transition. As such, these students were more likely to negotiate the system, avoid class, and drop out of school (Fredricks et al., 2004).

Students' Characteristics and Student-Teacher Relations

Students' characteristics (e.g., grade, age, and gender) identify how students perceive STR quality and moderate the associations between these STR and students' outcomes. That is, the perceptions of adolescents in middle schools about the supportive relations vary from that of younger students in elementary school and older students in high schools (Lee, 2012; Wentzel, 1997). For instance, middle school students described an excellent teacher as non-judgmental, supportive, and fair with all students (Seaton, 2007). Additionally, they reported other qualities of a good teacher as someone who likes them and listens to them (Kinney, 2007). Previous studies also have suggested that teenagers in middle schools rely on their teachers for emotional support in ways that vary across grades. Meaning, providing challenging learning activities combined with appropriate scaffolding may be proper for middle school adolescents (Greogory et al., 2014; Wentzel, 1997). Older students in high school, in contrast, rated good teachers as those who hold high expectations for their students, provide demanding learning tasks, offer encouragement, and show proper scaffolding when needed (Northup, 2011).

Simultaneously, age moderates the association between STR and other cognitive and noncognitive outcomes. A meta-analysis study (n = 65 studies from 1994 to 2016) revealed that age moderated the correlation between STR and academic emotions (i.e., positive emotions [e.g., enjoyment, pride] and negative emotions [e.g., shame, anxiety]; Lei et al., 2018). For instance, the negative association between STR and pessimistic academic emotions was the strongest among middle school students relative to other age groups. Meaning, middle school students experienced more negative academic emotions when STR was poor. Aldhafri and Alhadabi (2019) found significant STR differences across grades, supporting the association between STR and age. Meaning, students in younger grades (i.e., 7th grade) had better relationships with their science teachers than students in higher grades (i.e., 10th and 11th grade).

Regarding gender, the influence of positive relations on students' outcomes differs among male and female students. Lei et al. (2018) found that the strength of the negative association between teacher support and negative academic emotion was greater among females relative to males. Meaning, constructing positive relations with female students correlate with a lower level of pessimistic academic feelings, resulting in a better engagement and more proactive learning experience during difficult science lessons. Another meta-analysis study (n = 57 studies from 2000 to 2016) showed that gender moderated the association between positive student-teacher relations and externalizing behavior problems (Lei et al., 2016). In particular, the strength of this association was larger among females relative to males, supporting the selection of the current study sample.

IRT in Psychological Test development

IRT is an accurate methodological framework that has been widely used in developing standardized tests in education and health, but not in psychology (Rubio et al., 2007; Zanon et al., 2016). Despite the limited use of IRT during psychological scale construction, IRT can overcome the limitations of CTT due to three justifications. Unlike CTT that mainly examines the total scale, the primary IRT analysis unit is the single item. Meaning, IRT allows researchers to explore more details such as item difficulty, person performance in the aptitude test or ability in attitude scale, and item discrimination parameters (Bond & Fox, 2015). Therefore, terms like total-scores or summative information are not included in the statistical vocabulary in IRT.

Second, the scale statistics of IRT are invariant across items and persons. To clarify this point, it is essential to emphasize that scale reliability and validity in CTT depend on the samples and the scale items (Allen & Yen, 1979). In the first case (i.e., sample-dependency), researchers administer the test multiple times to the same sample to obtain test-retest reliability. However, indices of internal consistency reliability are not identical across numerous administrations despite using the same sample. Additionally, researchers tend to measure the degree of consistency by determining the relationships between these scores. In the second case (i.e., item-dependency), researchers use parallel forms of the same test to ensure that all possible items measure the construct consistently by estimating the parallel test reliability. Yet, scale scores vary to some extent across these parallel forms. All these reliability procedures are not needed in the IRT because the scale scores are invariant across samples and items.

Third, the relationship between item performance and trait/ability in IRT can be estimated by one, two, or three-parameter logistic function (i.e., 1-PL, 2-PL, and 3-PL, see Equation 1). In this equation, De Ayala (2009) identifies three primary parameters: (1) Item difficulty or also is known as item threshold (b_i; the easiness of endorsing an item that reflects the latent trait), (2) Item discrimination (a_i; the steepness of the item characteristic curve and how well it can differentiate among individuals located at different points of ability), and (3) Pseudo guessing parameter (c_i). In the Likert scale, b_i represents the point at which the individual with a certain level of measured trait has an equal probability (50:50) of endorsing an item across adjacent response categories (e.g., Agree vs. Strongly Agree). Only four threshold parameters are estimated (k-1, where k is the number categories) for a five-point scale.

$$P_1(\theta) = c_i + (1 - c_i)\frac{e^{-Da_i(\ - bi)}}{1 + e^{-Da_i(\ - bi)}} \qquad \rightarrow \text{Equation 1}$$

The additional undefined symbols in Equation 1 are: e is an exponential constant (e = 2.718), and D is a scaling factor whose value is 1.7 (Han, 2013). Correspondingly, IRT offers several models (e.g., Partial Credit Model, the Generalized Partial Credit Model, and the Rating Scale Model, Sequential Rasch Model, Graded Response Model, etc.; Bond & Fox, 2015; De Ayala, 2009; Linacre, 2002). The selection of the optimal IRT model depends on the items, response scale, and number of estimated parameters (i.e., 1-PL, 2-PL, and 3-PL). For example, under Rasch analysis (i.e., 1-PL), Rating Scale Model is used to analyze polytomous data, precisely the Likert scale, where all items have the same response structure and category weight (Bond & Fox, 2015). In contrast, when estimating the 2-PL model, Graded Response Model (GRM) is the appropriate model for analyzing polytomous items that use the Likert rating scale (Samejima, 2010). The Rasch analysis, the current study's scope, emphasized one principal assumption, unidimensionality (Linacre, 2002). Several techniques can be followed to ensure unidimensionality, which are: (1) Rasch Principal Components Analysis (PCA), as supported by small unexplained variance in the first contrast (i.e., < 2.00) in case of conducting 1-PL/Rasch analysis (Linacre, 2002), (2) Parallel analysis (Zanon et al., 2016), or (3) Confirmatory factor analysis (De Ayala, 2009). In case unidimensionality is met, one IRT model can be fit. Otherwise, the IRT model should be fitted for each dimension individually.

Method

Study Aim

As discussed earlier, the STR are assessed from the teachers' perspective and the students' perspective. Related to students' perspective, many scales were identified, as mentioned in the introduction. Student-Teacher Relationship Measure (STRM), which is a core interest in the current study, measures the relations from the perspective of 7th-12th students in the Arabic context (Aldhafri & Alhadabi, 2019). This scale was developed for various reasons. First, many STR scales measure the relation from a teachers' perspective (Ang, 2005; Pianta, 2001). Second, while most of the prior research in this area examined young children, little was known about the STR from the perspective of middle school and high school students (Saft & Pianta, 2001). Third, no standardized Arabic STR scale was available. Lastly, prior research found that the perception of STR among middle and high school students differs from that of elementary school students, necessitating developing a tailored scale for middle and high school students.

The psychometric properties of STRM (e.g., factorial structure, reliability, validity, and measurement invariance) were assessed using Classical Test Theory (CTT). That is, Aldhafri and Alhadabi (2019) supported a first-order two-factor solution (i.e., academic and social relations) with good Internal Consistency Reliability coefficients (i.e., Cronbach's α

=.94 and .92, respectively). As well, the second-order factorial structure was substantiated, where the academic and social relations loaded in a higher-order factor (i.e., STR). Construct validity was supported by two pieces of evidence: (1) Significant differences in STR across grades and (2) A positive association between STR and learning motivation (Aldhafri & Alhadabi, 2019). Only one level of measurement invariance (i.e., configural invariance) was supported. Metric, scaler, and strict invariance were not substantiated, implying no meaningful differences across grades can be obtained until these levels of invariance are supported.

As mentioned above, STRM's psychometric properties were estimated using CTT, which has many limitations. Spence et al. (2012) indicated that IRT results in developing scales with more solid psychometric properties by overcoming the limitations of CTT. That is, IRT provides three estimates (i.e., item difficulty, person ability, and item discrimination). Though, item discrimination estimates are not crucial in the present study since the STRM is considered a low stake scale. Meaning, the scale is not used as a diagnostic test in the educational context. Rasch Rating Scale Model (RSM) is appropriate to analyze the STRM because the items have the same response scale structure and category weights (Linacre, 2017). Therefore, the current research aimed to examine STRM using RSM, answering two main research questions:

(RQ1) What is the component structure of the STRM for middle and high school students?

(RQ2) What are the psychometric properties of the STRM for middle and school students?

Procedure and Sample

After obtaining the Ministry of Education Research Ethics committee's approval, the general director sent official invitation letters with a general overview of the study aims, duration of the scale, and procedures to preserve the responses' confidentiality. These letters also contained an attached invitation to attend a preparatory meeting to explain the study in-depth for voluntary science teachers. Later on, each teacher explained the study's purpose and procedures to the students. Parental consent forms were sent home for parents' approval of their children's participation in the study. The teachers emphasized voluntary participation and gave assurances that no risk was associated with completing the study scale.

A sample of female students (*N* = 995) was obtained from one large rural governorate in Oman. This governorate was selected because it was a large governorate that has more than 36 female middle schools and assimilates other Omani rural governorates, covering diverse geographical regions. The collected sample covered four grades, which were: 7th (*n* = 297, 29.8%), 8th (*n* = 250, 25.1), 9th (*n* = 234, 23.5%), and 10th grade (*n* = 214, 29.8%).

Measures

A survey was administered containing two sections, including (1) Demographic information and (2) STRM (Aldhafri & Alhadabi, 2019). The STRM is a 25-item scale that evaluates students' perceptions of the STR with science teachers, capturing two dimensions (i.e., academic relation [AR] and social relation [SR]). Items were rated on a 5-point Likert scale ranging from "Never applies" (Coded 1) to "Definitely Applies" (Coded 5). Examples of AR items include "My teacher makes me feel that I am able to solve difficult questions" and "My teacher encourages me to ask about things that I did not understand." Examples of SR items are "My teacher cares about my performance" and "My teacher listens to what I say". The AR and SR had good internal consistency reliability (Cronbach's α = .94 and .89, respectively; Aldhafri & Alhadabi, 2019).

Data Analysis

Data were screened and descriptive demographic statistics were examined using the Statistical Package for Social Sciences (SPSS) for Windows Version 24.0 in order to identify missing data, nonnormality, and ceiling or floor effects. Winsteps 4.01 for Windows was used for the Rasch Analysis. The dimensionality of each STRM sub-factor was examined using Rasch Principal Components Analysis (PCA). Small unexplained variance (i.e., < 2.00) in the first contrast supports unidimensionality (Linacre, 2017). Following Rasch PCA, two Rasch RSM were conducted for the two subscales (i.e., AR and SR). Several statistics were assessed to evaluate the psychometric properties of STRM, including the model global fit, Wright item-person map, summary statistics (i.e., persons/items separation and reliability), category function, and item/person fit statistics (i.e., MNSQ and ZSTD values; Bond & Fox, 2015; Linacre, 1999; Linacre, 2002; Linacre, 2017).

Related to the first statistic, a non-significant Chi-square justified the model global fit. Yet, relying solely on the global fit is not satisfactory, suggesting the examination of items/persons fit indices (Linacre, 2017). The second indicator, Wright item-person map, provides a visual representation of the items and persons across the logit vertical line. The optimal Wright map should graphically present a normal distribution of persons and items along the logit line (Linacre, 2002).

Third, two statistics describe the persons' fit (i.e., person separation and person reliability). Another two statistics gauge the items' fit (i.e., item separation and item reliability; Bond & Fox, 2015). The person separation index estimates the spread between persons on the measured construct and is obtained by dividing adjusted person standard deviation with average measurement error. In other words, the person separation index refers to the scale's efficiency in classifying respondents across the logit scale. Values less than two indicate that the scale did not distinguish between respondents who scored low and high in the measured construct and, in this case, STR (Linacre, 2017). Person reliability evaluates the reproducibility of persons' locations across the vertical line when another set of similar scale items are introduced to the same persons (Bond & Fox, 2015). It is analogs to internal consistency reliability and ranges from zero to one. Comparatively, the item separation index gauge the spread between item locations. Item reliability illustrates the replicability of item locations across the vertical line when the same items are administered to other persons with similar ability levels. Overall, persons indices (i.e., separation ≥ 2.00 and reliability ≥ .80) and items indices (i.e., item separation ≥ 3.00 and reliability ≥ .90) demonstrate an optimal fit (Bond & Fox, 2015).

For the fourth indicator, category function, the following criteria was used to identify the acceptable category structure: (1) A monotonic increase in the average category measures, reflecting un-disordered categories, (2) Large observed count in each category (i.e., ≥ 10), and (3) Well-spaced distance between any two Andrich thresholds ($1.4 \geq d > 5$; Bond & Fox, 2015). The literature also provides another liberal criterion, which states that the lowest distance should not be less than 1.00 (Linacre, 2002). Graphically, the satisfactory categories structure assimilates smooth rolling hills (Linacre, 2017).

Considering the last statistics, item/person fit, items and persons mean square (MNSQ) should range from .60 to 1.7 for the rating scale (Bond & Fox, 2015). Since STRM is a low-stake scale, a more liberal criterion was selected in the current study (i.e. .50 < MNSQ < 2.00; Linacre, 2017). Small values (MNSQ/ZSTD > .50) and larger values (MNSQ/ZSTD < 2.00) reflect overfit and underfit respectively (Bond & Fox, 2015). Additionally, the point-measure correlation coefficients were reviewed. Overall, negative and zero values indicate the items' scoring contradicts the orientation of the measured latent construct (i.e., negative coding) and signals any items miscoding (Boone & Staver, 2020). Positive values indicate that the items are functioning well (Linacre, 2017).

Results

This section articulates three segments of findings, including (1) Assessment of unidimensionality assumption and the results of Rasch PCA, (2) Rasch analysis of the academic relation subscale, and (3) Rasch analysis of the social relation subscale. In the second and third segments, the psychometric properties of the STRM were investigated by fitting separate RSM for the two subscales (i.e., AR and SR). Several statistics were examined to evaluate the psychometric properties, which include: (1) Summary statistics of the initial dimension/subscales, (2) Initial misfit items statistics (see Table 1), (3) Category structure (see Table 2), (4) Psychometric properties of multiple modification rounds (see Table 3), and (5) The summary statistics of the modified subscales.

Assessment of Unidimensionality Assumption and Findings of Rasch PCA

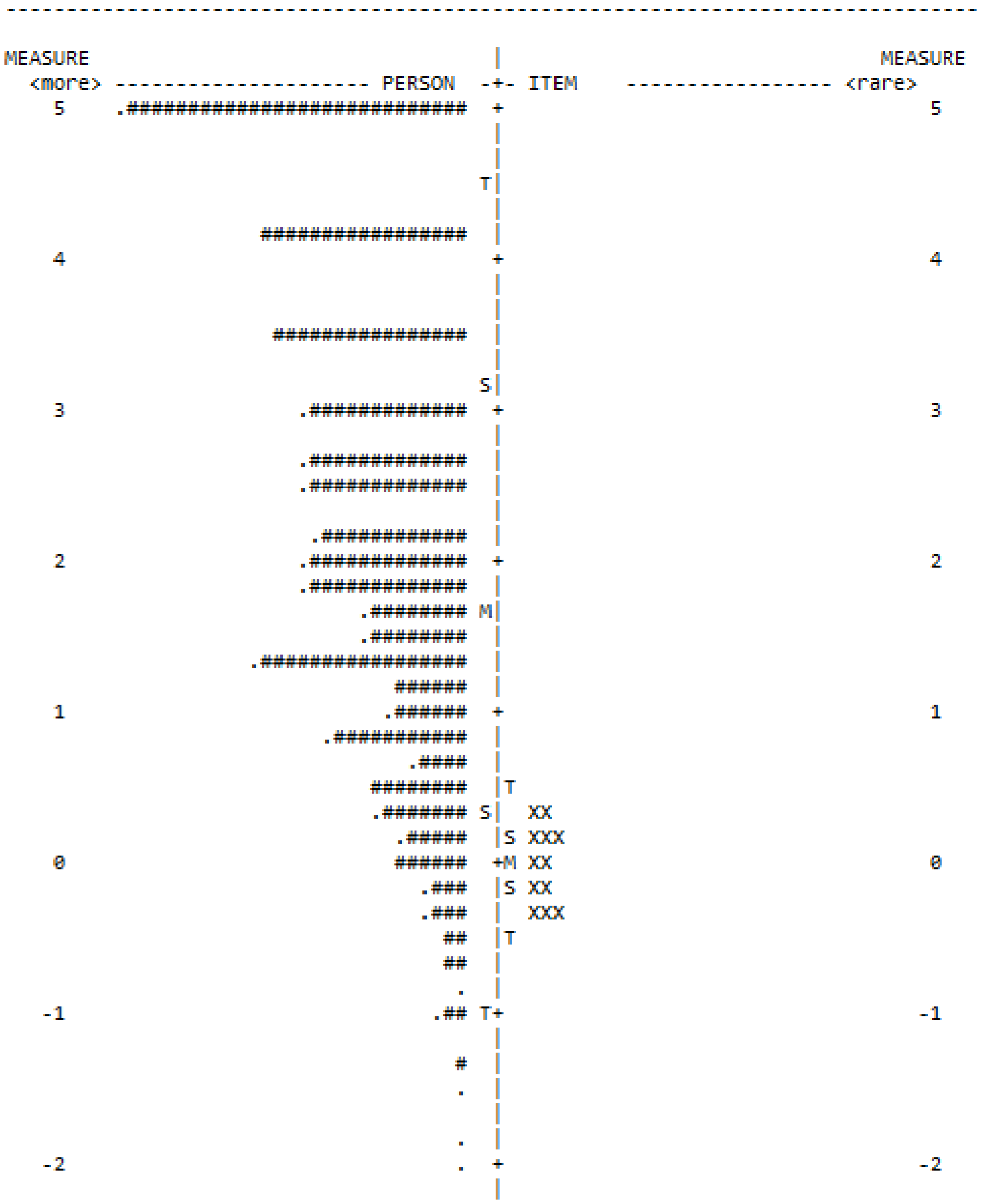
Preliminary examination was conducted by assessing the unidimensionality using PCA. Results revealed that the STRM had two components. The first component of the STRM had sixteen items (i.e., Items 15, 14, 22, 11, 17, 25, 20, 12, 16, 13, 24, 19, 10, 18, 23, 21), labeled as "Academic Relation [AR]". The second component had nine items (i.e., Items 6, 1, 4, 9, 8, 3, 2, 5, 7), was labeled as "Social Relation [SR]". The observed explained variance by the AR was 12.81%, shared by persons (7.70%) and by items (5.11%). Comparatively, SR explained 12.30% of the variance, shared by persons (7.58%) and by items (4.72%). Examination of the standardized unexplained variance in the first contrast reflected that dimensionality was met for AR and SR (i.e., 1.35 and 1.43, respectively).

Results of Rasch Analysis Academic Relation Subscale

Overall, the non-significant Chi-Square (χ^2 [30554] = 3034.28, p = .82) indicated a good global model fit. Good global model fit, however, did not capture any problematic persons and items misfit issues (Linacre, 2017). Given that, items and person summary statistics were reviewed. Items were located from -.28 to .35 on the logits scale, indicating that items were distributed normally around a mean of zero, as shown in the Wright Item-Person Map (see Figure 1). Even though some items were located at the same point in the logit scale, which may indicate redundancy, a content investigation of these items did not show any redundancy in the item content. The average person measure (1.35) indicated that persons had a mean greater than zero. Meaning, a larger number of persons highly endorsed AR with their science teachers, indicating that this subscale was easy to endorse from the perspective of the study sample.

Initial summary statistics showed that person separation and reliability were 2.66 and .88, respectively. Item separation and reliability were 4.55 and .95, respectively. These statistics indicated good person/items separation and reliability (Linacre, 2017). As well, all items had a good fit, as indicated by individual item mean square infit and outfit (MNSQ) estimates (see Table 1.A). However, ZSTD infit and outfit illustrated some contradictory findings, suggesting nine items were diagnosed as potentially misfit (i.e., Items 23, 15, 18, 25, 13, 16, 24, 19, and 21) because the ZSTD values were beyond the optimal range (i.e., ± 2.00; De Ayala, 2009). The point-measure correlation coefficients were positive, indicating no concerns about miscoding and the items measured the AR, as intended. After reviewing the Item

Characteristic Curve (ICC) of these nine misfit items, Items 23 and 21 were the most problematic (see Figure 2). Item 18 showed a smaller deviation relative to Item 23 and 21. Comparatively, person estimates reflected that a large number of persons were misfit. As such, eliminating most misfit items (23, 21,18) and persons (*n* = 439) were suggested.



```
-------------------------------------------------------------------------------

MEASURE                                |                              MEASURE
  <more> -------------------- PERSON  -+- ITEM    ----------------- <rare>
    5   .############################  +                                   5
                                       |
                                       |
                                      T|
                                       |
                    #################  |
    4                                  +                                   4
                                       |
                                       |
                     ################  |
                                       |
                                      S|
    3                  .#############  +                                   3
                                       |
                       .#############  |
                       .#############  |
                                       |
                        .############  |
    2                  .#############  +                                   2
                       .#############  |
                            .######## M|
                            .########  |
                   .#################  |
                               ######  |
    1                         .######  +                                   1
                         .###########  |
                                .####  |
                             ########  |T
                             .####### S|  XX
                               .#####  |S XXX
    0                          ######  +M XX                               0
                                 .###  |S XX
                                 .###  |  XXX
                                   ##  |T
                                   ##  |
                                    .  |
   -1                             .## T+                                  -1
                                       |
                                    #  |
                                    .  |
                                       |
                                    .  |
   -2                               .  +                                  -2
                                       |
```

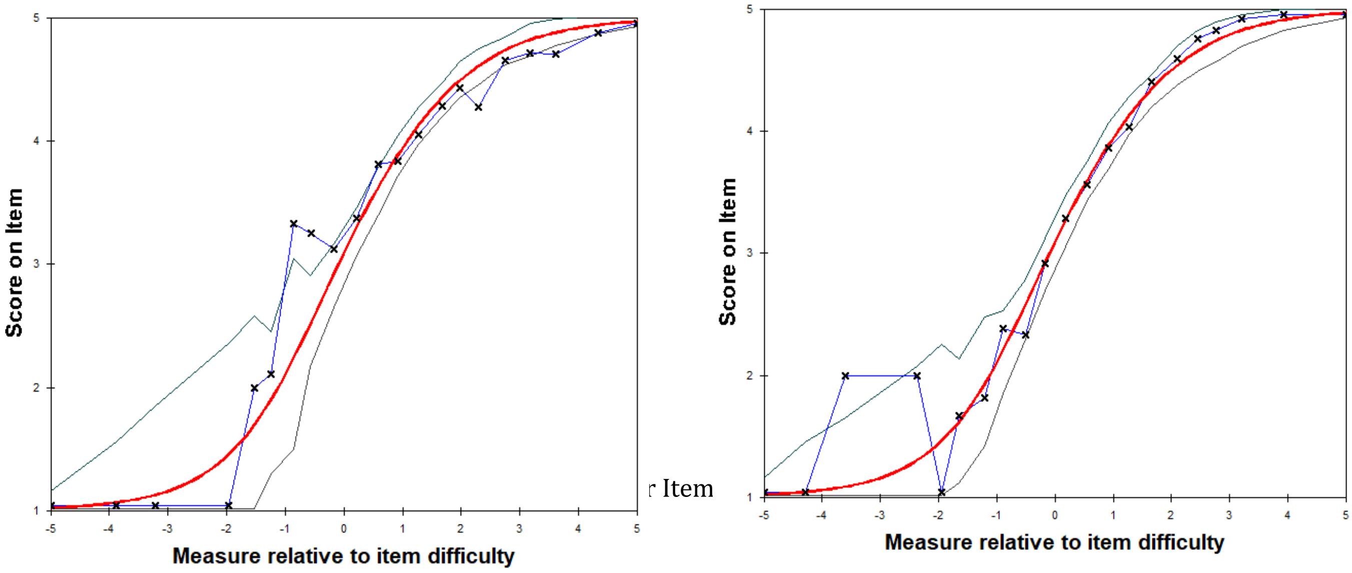
Figure 1. The Wright item-person map for the "Academic Relation" subscale.

Note: The left side of the vertical logit line shows the distribution of person and the right side represents the distribution of items. The top part across the logit line demonstrates the person who highly endorsed the AR on the right side and the most difficult items on the left side. The bottom part illustrates the persons who reported poor AR on the right side and easiest items on the left side. The letter *M* on the vertical line is the person and item mean. The letter *S* shows the location of one standard deviation from the mean, and the letter *T* refers to the location of two standard deviations from the mean. Each "#" represents four respondents, while each "." represents one to three respondents. "x" represents the individual items.

Table 1: Initial item statistics for misfit order (N = 995)

#	Items	Measure	SE	Infit		Outfit		Pt-Measur Co
				MNSQ	ZSTD	MNSQ	ZSTD	
	A. Academic Relation (16 Items)							
23	My teacher involves other students to answer the questions that are asked by their peers.	-.05	.04	1.30	**5.4**	1.49	**7.1**	.57
15	My teacher encourages me to ask about things that I did not understand.	-.23	.04	1.12	**2.3**	1.11	1.7	.62
22	My teacher allows students to think before answering questions.	-.28	.05	1.07	1.4	1.07	1.1	.62
18	My teacher gives some hints to provide the right answer.	.11	.04	1.22	**4.3**	1.27	**4.4**	.62
14	My teacher encourages good behavior in the class.	-.22	.04	1.08	1.5	1.09	1.4	.62
10	My teacher asks exciting questions related to the subject.	-.14	.04	1.07	1.3	1.10	1.6	.63
11	My teacher expects me to participate effectively in the classroom.	-.27	.04	.99	-.2	1.02	.3	.63
17	My teacher makes me feel proud when I achieve certain goals.	-.17	.04	1.01	.2	.98	-.3	.64
12	My teacher encourages positive interaction between students.	-.02	.04	.89	-2.2	.89	-2.0	.67
25	My teacher encourages students to find more than one way to solve problems.	.02	.04	.88	**-2.4**	.88	**-2.8**	.68
20	My teacher shows remarkable enthusiasm during class.	.15	.04	.98	-.3	.93	-1.2	.68
13	My teacher makes me feel that I'm able to solve difficult questions.	.22	.04	.87	**-2.8**	.93	-1.2	.69
16	My teacher encourages me to be the best I can.	-.01	.04	.82	**-3.9**	.81	**-3.9**	.69
24	My teacher develops my self-confidence to succeed in science.	.21	.04	.89	**-2.3**	.85	**-2.9**	.70
19	My teacher uses teaching methods that develop my ability to cooperate with others.	.34	.04	.90	**-2.2**	.87	**-2.5**	.71
21	My teacher believes in me and my potential.	.35	.04	.84	**-3.5**	.81	**-3.7**	.72
	B. Social Relation (9 Items)							
1	My teacher listens to what I say.	-.57	.05	1.27	**5.1**	1.19	**3.1**	.63
6	My teacher cares about me.	-.47	.04	1.25	**4.8**	1.18	**3.0**	.64
9	My teacher links the subject's topics with characters that matter to us.	.58	.04	1.14	**3.0**	1.19	**3.9**	.70
4	My teacher encourages me to ask questions.	.13	.04	.99	-.1	.99	-.3	.71
8	My teacher uses a variety of ways that captivate my attention.	.00	.04	.93	-1.5	.92	-1.6	.73
5	My teacher provides practical implications about the taught lessons.	-.21	.04	.81	**-4.4**	.74	.70	.74
3	My teacher strengthens my confidence in my ability and talents.	-.01	.04	.96	-.9	.88	-2.6	.74
2	My teacher excites me to learn science.	.14	.04	.90	**-2.4**	.88	**-2.6**	.75
7	My teacher uses methods that suit my interest.	.41	.04	.78	**-5.3**	.79	**-4.9**	.78

Note. Measure = Item Difficulty Estimate; MNSQ = Mean Square; ZSTD = Standardized Fit; Pt-Measur Cor= Point-measure correlation coefficient.



Item 23

Item 21

Figure 2. Academic relation: Item characteristic curves for Items 23 and 21.

Note: The solid red line is the model line that is estimated by Rasch analysis. The black “x” points identify the observations in an interval on the latent variable, which was, in this case, AR. The blue line shows the empirical item characteristic curve estimated directly from the data. The top and bottom gray lines represent the upper and lower 95% confidence interval boundaries, respectively. The optimal trend is identified by “x” values that are located between the two lines. Points outside these boundaries are problematic and imply that some of the variance in these observations was not fully modeled and explained by the Rasch analysis.

The examination of the Likert category structure for each item (see Table 2.A) showed that observed average values increased monotonically, and thresholds between categories were not disordered. However, the distance between the Andrich thresholds was too small (i.e., < 1.0). Graphically, category label 2 (i.e., Applies little) was lower than other categories; yet, the category probability curves appeared approximately as smooth rolling hills. Thus, no suggestion about collapsing categories had been made.

Three rounds of modification were conducted (see Table 3.A). The suggested modifications were implemented in separate iterations. A final round of analysis reflected that the final results were satisfactory at the persons and items level (see Table 4). Several indicators illustrate this improvement, as follows: (1) Good person and items separation (2.32 and 3.79, respectively; see Table 4), (2) Acceptable person and item reliability (.84, .94 respectively; see Table 4), (3) A monotonic increase in category measures without any disorder (see Table 2.B), and (4) An increase in the distance between the threshold across categories (see Table 2.B), aligning with Linacre (2002) liberal criteria.

Table 2: Category frequency and threshold values for the STRM dimensions

Category Label	Observed Count	%	Observed Average	Sample Expect	Infit MNSQ	Outfit MNSQ	Andrich Threshold	Category Measure
A.	**Category Structure for Initial Academic Relations (16 Items)**							
1	537	3	-.63	-.74	1.17	1.37	NONE	-2.56
2	1046	7	-.11	-.02	.88	.90	-1.13	-1.12
3	2673	17	.50	.54	.91	.90	-.68	-.09
4	4714	30	1.29	1.24	.90	.98	.30	1.08
5	6950	44	2.29	2.30	1.08	1.05	1.51	2.80
B.	**Category Structure for Modified Academic Relations (12 Items)**							
1	209	2	-1.36	-1.20	.78	.77	NONE	-2.93
2	497	6	-.03	-.01	.92	.86	-1.62	-1.26
3	1267	14	.73	.74	.97	.88	-.57	-.02
4	2445	27	1.63	1.58	1.03	1.07	.49	1.25
5	4558	51	2.58	2.60	1.09	1.05	1.71	2.99

Table 2: Continued

Category Label	Observed Count	%	Observed Average	Sample Expect	Infit MNSQ	Outfit MNSQ	Andrich Threshold	Category Measure
A. Category Structure for Initial Social Relation (9 Items)								
1	412	4	-2.10	-1.16	1.09	1.13	NONE	-2.87
2	812	8	-.42	-.34	.88	.88	-1.50	-1.31
3	1937	20	.44	.46	.91	.85	-.81	-.10
4	3017	31	1.41	1.36	.96	.99	.46	1.27
5	3713	38	2.41	2.44	1.11	1.07	1.85	3.10
B. Category Structure for Modified Social Relation (9 Items)								
1	196	3	-1.65	-1.61	.89	.90	NONE	-3.31
2	499	7	-.38	-.33	.87	.83	-2.06	-1.45
3	1157	17	.71	.71	.96	.85	-.64	.03
4	1842	28	1.75	1.70	1.09	1.08	.74	1.46
5	2975	45	2.75	2.77	1.05	1.04	1.96	3.24

Note. Category Label 1 = Never applies, Label 2 = Applies little, Label 3 = Applies sometimes, Label 4 = Applies often, and Label 5 = Definitely applies.

Table 3: Psychometric properties in three modification rounds of rasch RSM for the STRM

	Categorization	Average Measures	Person Separation	Person Reliability	Item Separation	Item Reliability
	A. Academic Relation					
1	12345 (original data)	ordered	2.66	.88	4.55	.95
2	12345 after eliminating misfit persons (*n* = 439)	ordered	2.22	.83	3.91	.94
3	12345 after eliminating misfit items (*n* = 3) & persons (*n* = 34)	ordered	2.32	.84	3.79	.94
	B. Social Relation					
1	12345 (original data)	ordered	2.27	.84	7.32	.98
2	12345 after removing misfit persons (*n* = 465)	ordered	2.33	.84	6.56	.98
3	12345 after removing items (9, 1)	ordered	1.82	.77	5.81	.97

Note: 12345 refers to the order of the five Likert categories. 1 = Never applies, 2 = Applies little, 3 = Applies sometimes, 4 = Applies often, and 5 = Definitely applies.

Table 4: Item and person summary statistics for the modified "Academic Relations" subscale (n = 12)

Statistic	Total Score	Measure	Model Error	Infit		Outfit	
				MNSQ	ZSTD	MNSQ	ZSTD
A. Person level (659 Persons)							
Mean	49.2	1.74	.48	.97	.0	.97	.0
SD	8.6	1.34	.20	.31	.8	.33	.8
Max	59.0	4.18	1.02	1.93	2.1	1.97	2.2
Min	13.0	- 4.10	.31	.44	-1.6	.45	-1.6
Real	RMSE: .53 True SD: 1.23 Person Separation: 2.32 Person Reliability: .84						
Model	RMSE: .52 True SD: 1.24 Person Separation: 2.40 Person Reliability: .85						
B. Item level (12 Items)							
Mean	3131.2	.00	.05	.99	-.2	.97	-.5
SD	73.4	.22	.00	.10	1.6	.09	1.3
Max	3218.0	.32	.06	1.18	2.7	1.12	1.6
Min	3022.0	-.26	.05	.86	-2.5	.85	-2.2
Real	RMSE: .06 True SD: .21 Item Separation: 3.79 Item Reliability: .94						
Model	RMSE: .06 True SD: .21 Item Separation: 3.88 Item Reliability: .94						

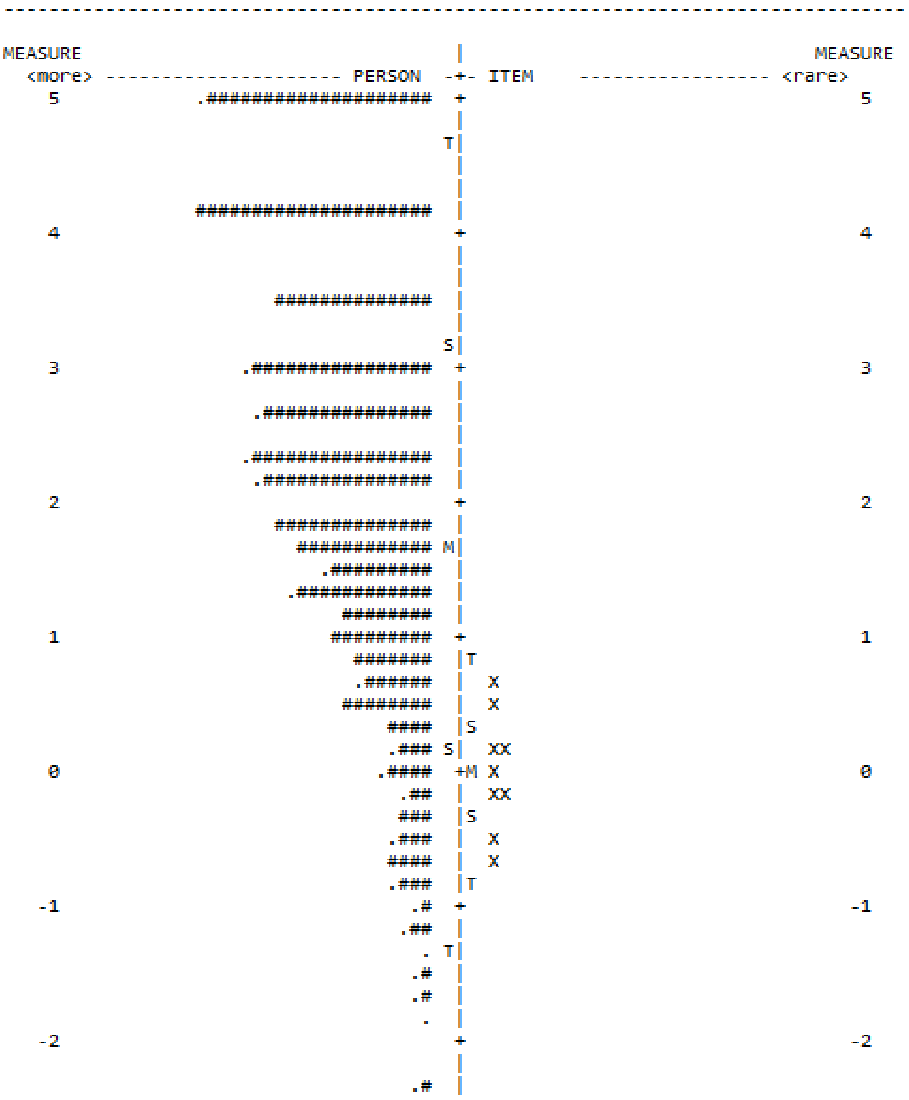
Note: Measure = Item Calibration Estimated; MNSQ = Mean Square; ZSTD = Standardized Fit; RMSE = Root Mean Square Error

Results of Rasch Analysis Social Relation Subscale

Overall, the non-significant Chi-Square (χ^2 [17382] = 17205.46, *p* = .83) indicated a good model fit. Furthermore, Person Separation (2.15) and Item Separation (7.23) were acceptable (Linacre, 2017). The person reliability (.82) and item reliability (.98) met the optimal standard (Bond & Fox, 2015). Additionally, the Wright map showed that items were located from -.57 to .58 logits on the scale, indicating that items distributed normally around the mean of zero (see Figure 3). The person average measure value (1.30) showed that persons had a mean greater than zero, implying that item difficulty was above average. However, many participants scored highly in the social relation, suggesting it was easy for students to endorse SR items.

Individual item fit statistics illustrated that all items had a good fit, as indicated by MNSQ infit and outfit estimates (see Table 1.B). Like AR dimension, ZSTD infit and outfit flagged six misfit items (i.e., Item 1, 6, 9, 5, 2, and 7). Three items showed underfit (i.e., Items 1, 6, and 9 where the MNSQ values were greater than 1.0, reflecting that the data are less predictable than the model expects). The other three items were considered overfitting (i.e., Items 5, 2, and 7 where the MNSQ values were below 1.0, reflecting that the data are more predictable than the model expects (Wright & Linacre, 1994). The correlation coefficients were positive, implying that all items functioned well.

In contrast, only Item 9 and 1 were flagged as the most problematic items according to Item Characteristic Curve (see Figure 4). The examination of the Likert category structure for each item revealed that the categories were monotonically ordered across all items. Only in Item 9, the distances between categories were not evenly spaced. Item 9 was collectively a candidate for elimination due to item misfit, as indicated ZSTD, problematic ICC, and uneven spaced distance between Likert categories. As well, Item 1 was nominated for deletion because of unacceptable ZSTD and problematic ICC. Comparatively, person estimates indicated a large number of misfit persons (*N* = 465). Thus, eliminating misfit persons was suggested first, followed by removing misfit items.

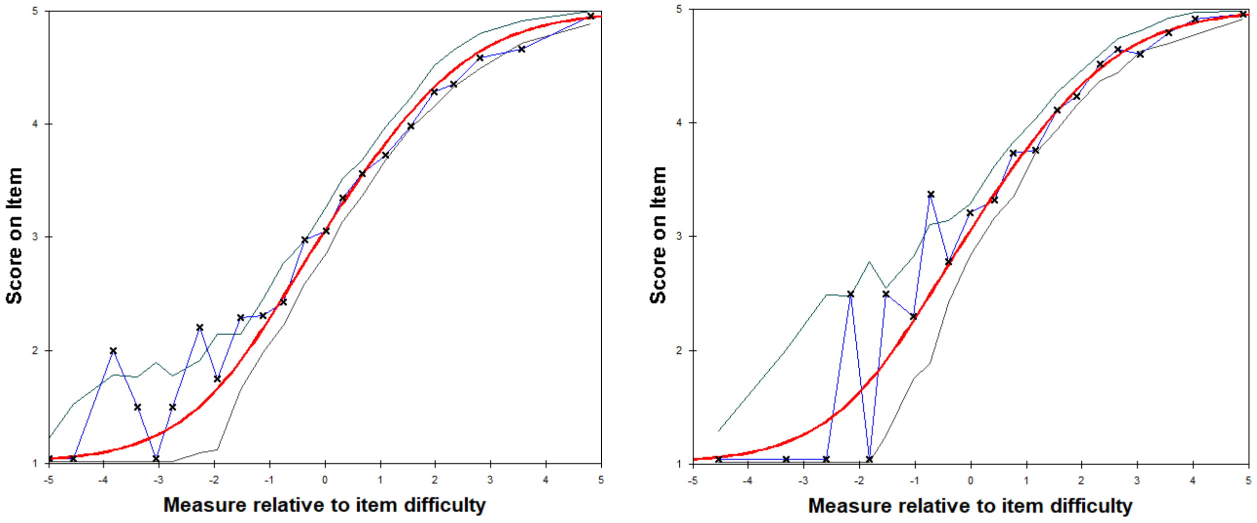


```
--------------------------------------------------------------------------------

MEASURE                                 |                               MEASURE
  <more> --------------------- PERSON  -+- ITEM    ----------------- <rare>
    5            .####################  +                                   5
                                        |
                                       T|
                                        |
                                        |
                 #####################  |
    4                                   +                                   4
                                        |
                                        |
                        ##############  |
                                        |
                                       S|
    3                .################  +                                   3
                      .###############  |
                                        |
                     .################  |
                      .###############  |
    2                                   +                                   2
                        ##############  |
                          ############ M|
                            .#########  |
                         .############  |
                              ########  |
    1                        #########  +                                   1
                               #######  |T
                               .######  |  X
                              ########  |  X
                                  ####  |S
                                  .### S|  XX
    0                            .####  +M X                                0
                                   .##  |  XX
                                   ###  |S
                                  .###  |  X
                                  ####  |  X
                                  .###  |T
   -1                               .#  +                                  -1
                                   .##  |
                                     . T|
                                    .#  |
                                    .#  |
                                     .  |
   -2                                   +                                  -2
                                        |
                                    .#  |
```

Figure 4. The Wright item-person map for the "Social Relation" subscale.

Note: The left side of the vertical logit line shows the distribution of person and the right side represents the distribution of items. The top part across the logit line demonstrates the person who highly endorsed the SR on the right side and the most difficult items on the left side. The bottom part illustrates the persons who reported mediocre SR on the right side and the easiest items on the left side. The letter *M* on the vertical line is the person and item mean. The letter *S* shows the location of one standard deviation from the mean, and the letter *T* refers to the location of two standard deviations from the mean. Each "#" represents four respondents, while each "." represents one to three respondents. "x" represents the individual items.



Item 9

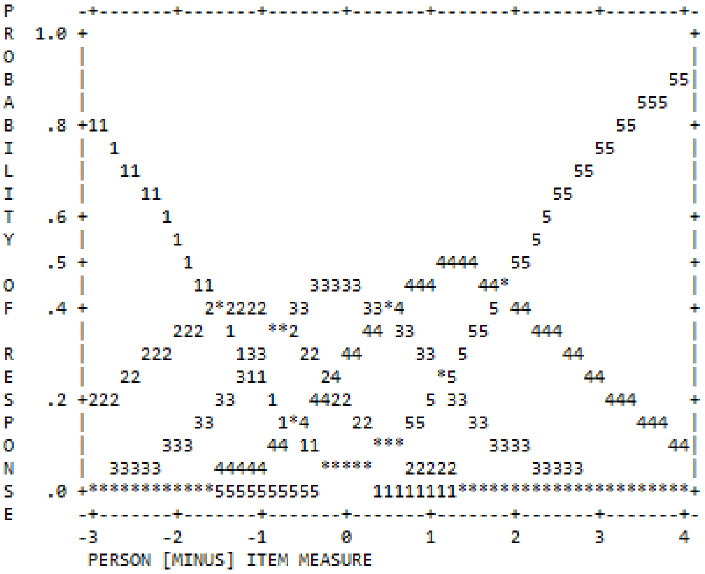
Item 1

Figure 5. Item characteristic curves for items 9 and 1 in the SR subscale.

Note: The solid red line is the model line that is estimated by Rasch analysis. The black "x" points identify the observations in an interval on the latent variable, which was, in this case, SR. The blue line shows the empirical item characteristic curve estimated directly from the data. The top and bottom gray lines represent the upper and lower 95% confidence interval boundaries, respectively. The optimal trend is identified by "x" values that are located between the two lines. Points outside these boundaries are problematic and imply that some of the variance in these observations was not fully modeled and explained by the Rasch analysis.

Category structure findings revealed that the categories were not disordered as indicated by a monotonic increase in the average category measures. The observed count in each category was large (i.e., > 10). The category probability curves graphically appeared as smooth rolling hills, suggesting no need for collapsing categories. Nevertheless, the distances between the Andrich thresholds of adjunct categories were small (*d* < 1.4; see Table 2.C), displaying a problematic issue. Linacre (2002) explained this issue as follows: "can indicate that a category represents too narrow segment of the latent variable or corresponds to a concept that is poorly defined in the minds of the respondents" (p. 98).

Like academic relations, the suggested modifications were conducted in separate iterations (See Table 3.B). Eliminating misfit persons (*N* = 465) showed a substantial enhancement in the item and person estimates. Also, there was a decrease in the items and person separation, though the values were within the optimal range, according to Bond and Fox (2015). There was an increase in three distances between the adjunct thresholds (i.e., (1) 1.42 between first and second thresholds, (2) 1.38 between the second and third thresholds, and (3)1.22 between the third and the fourth thresholds). These values met Linacre's liberal and conservative criteria (see Table 2.D). Graphically, the category probability curves illustrated smooth rolling hills (see Figure 5).



```
P      -+-------+-------+-------+-------+-------+-------+-------+-
R  1.0 +                                                         +
O      |                                                         |
B      |                                                       55|
A      |                                                    555  |
B   .8 +11                                                55     +
I      |  1                                             55       |
L      |   11                                         55         |
I      |     11                                     55           |
T   .6 +       1                                   5             +
Y      |        1                                 5              |
    .5 +         1                       4444   55               +
O      |          11         33333    444    44*                 |
F   .4 +           2*2222  33     33*4        5 44               +
       |        222  1   **2      44 33     55    444            |
R      |     222      133   22  44     33  5         44          |
E      |   22         311     24         *5            44        |
S   .2 +222         33   1   4422       5 33             444     +
P      |          33      1*4    22   55    33              444  |
O      |       333       44 11     ***        3333             44|
N      |  33333     44444     *****   22222       33333          |
S   .0 +************5555555555     11111111**********************+
E      -+-------+-------+-------+-------+-------+-------+-------+-
       -3      -2      -1       0       1       2       3       4
        PERSON [MINUS] ITEM MEASURE
```

Figure 5. Category probability curves for SR items.

Note. 1 = Never applies, 2 = Applies little, 3 = Applies sometimes, 4 = Applies often, and 5 = Definitely applies. The optimal category probability curves should look like smooth rolling hills.

The third round of modification (i.e., eliminating the two most misfit items; see Table 3.B) reflected non-significant Chi-Square (χ^2 [8906] = 8709.55, *p* = .93). The person separation got worse (1.82 < 2.00) as well as person reliability (.77). Similarly, the item separation (5.81) declined; fortunately, it was above the cutoff of three (Linacre, 2017). Item reliability (.97) was good. As well, there was an increase in the distance between the threshold and met the criteria of 1.4. The category probability curves showed a similar presentation of categories to the second round. Rounds 2 and 3 provided approximately similar results. Meaning, Round 2 (i.e., eliminating misfit person) resulted in better distance between the thresholds, good estimation of items/persons separations, and reliabilities, concurring with the optimal standard. Eliminating misfit items in Round 3 resulted in wider distances between thresholds; however, it resulted in unacceptable persons' separation and reliability. A trade-off between the two rounds, Round 2 results were considered as the most desirable for the SR subscale (See Table 5).

Table 5: Item and Person Summary Statistics for the Modified "Social Relations" Subscale (n = 9)

Statistic	Total Score	Measure	Model Error	Infit		Outfit	
				MNSQ	ZSTD	MNSQ	ZSTD
Person level (675 Persons)							
Mean	35.7	1.68	.53	.96	.1	.97	.1
SD	7.1	1.48	.19	.33	.7	.35	.7
Max	44.0	4.17	1.03	2.02	1.9	1.98	1.9
Min	10.0	-4.26	.39	.45	-1.4	.49	-1.3
Real	RMSE: .58 True SD: 1.36 Person Separation: 2.33 Person Reliability: .84						
Model	RMSE: .56 True SD: 1.37 Person Separation: 2.43 Person Reliability: .86						
Item level (9 Items)							
Mean	2989.8	.00	.06	1.00	-.02	.97	-.7
SD	125.9	.38	.00	.16	2.9	.15	2.3
Max	3188.0	.62	.06	1.25	3.8	1.20	2.5
Min	2775.0	-.63	.05	.74	-5.1	.76	-4.5
Real	RMSE: .06 True SD: .38 Item Separation: 6.56 Item Reliability: .98						
Model	RMSE: .06 True SD: .38 Item Separation: 6.81 Item Reliability: .98						

Note. Measure = Item Calibration Estimated; MNSQ = Mean Square; ZSTD = Standardized Fit; RMSE = Root Mean Square Error.

Discussion

Constructive relations between students and their teachers facilitate students' growth academically and socially. Literature has provided ample indicators about the promising impact of fruitful STR and the pessimistic effect of calamitous STR on the quality of students' outcomes (Baker et al., 2008; Brewster & Bowen, 2004; DiLalla et al., 2004; Greogory et al., 2014; Hughes, 2012; Longobardi et al., 2016; Ridwan et al., 2014). The STR quality is moderated by gender, suggesting that females in middle school perceive these relations differently from their male peers (Lei et al., 2018). This STR is also moderated by student age, implying that the adolescents and young students in elementary schools hold beliefs about the quality of these relations (Aldhafri & Alhadabi, 2019; Lee, 2012; Lei et al., 2018). STRM is one of the scales, which assesses STR in the Arabic context. The scale has good psychometric properties, as obtained by CTT, except that higher measurement invariance levels (e.g., metric, scalar, and strict invariance) were not met (Aldhafri & Alhadabi, 2019). Though, no previous attempt was conducted to analyze this scale using IRT, notably Rasch Analysis. Therefore, the current study evaluated the STRM psychometric properties by conducting Rasch PCA (i.e., estimating the factorial structure and testing dimensionality) and two RSM for STRM sub-scales among female students in rural middle schools (i.e., 7th-10th grades).

Rasch PCA revealed psychometric support of two components (i.e., academic relations and social relations). Summary statistics showed good psychometric properties. These findings concur with the criteria presented by experts in Rasch analysis (Bond & Fox, 2015; Linacre, 2017). However, category structure and individual statistics (i.e., items and person infit and outfit) were not ideal. In other words, the category structure showed that the distances between adjacent thresholds were below the liberal cutoff (1.0; Linacre, 2002) and the conservative criteria (i.e., 1.4; Linacre, 2017) for the two subscales. Whereas findings indicated that items mean square statistics were optimal, standardized fit statistics reflected many misfit persons and items in each subscale.

For the academic relations subscale, the distance between thresholds met the criteria of 1.00 after eliminating misfit items (Item 23, 21, and 18) and misfit persons (*n* = 473) in the third round of modifications. For the social relations subscale, the second round of modifications (i.e., eliminating misfit persons, *n* = 465) resulted in substantial improvement, meeting the optimal liberal criteria related to distance between thresholds. The final round (i.e., eliminating misfit items based on ZSTD values) resulted in unacceptable persons' separation and reliability estimates. Trade-off between the two rounds' findings suggested the acceptance of Round 2 results since the scale is not a diagnostic tool (i.e., low-stake scale). Findings showed that the final 22-item had good psychometric properties, high item/person separation, and good item/person reliability for the two subscales. Unlike the results found by Aldhafri and Alhadabi (2019), the Rasch Analysis among middle school students revealed a shorter scale (i.e., 22 items).

This shorter scale has many practical advantages, including its convenience and feasibility to be administered in the school context, considering teachers' large workload. Understanding the quality of girls' relations with their science teachers may benefit educational policymakers and administrators by gaining deeper understating and data-driven knowledge about the role of productive relations in shaping students' engagement in STEM subjects (Mandinach, 2012; Schildkamp, 2019). This understanding facilitates the implementation of interventions that may enhance the enrollment of rural females in STEM majors, echoing the several prior studies suggestions (Aldhafri & Alhadabi, 2019; Hill et al., 2018; Lei et al., 2018).

Conclusion

In short, learning and willing to excel in science is facilitated by holding proper and supporting student-teacher relationships, whether academically or socially. This relationships seems to be more influential among female students in rural middle school. Assessing the quality of these relations without adding more workload to the teachers is an important aspect. In other words, establishing a valid and short scale gauging the relationships between students and their science teacher is a priority, particularly in Arabic, for Middle East countries. This assessment will help in making data-driven policies and strategies to meet the fourth industrial revolution requirements. The evaluation of the scale using the Classical Test Theory is not enough. To this end, the current study assessed the psychometric properties of STRM using IRT, precisely Rasch Analysis. This study provided a more reliable and shorter scale with more adequate properties and estimates that are sample-independent. This statement was supported by (1) High persons and items' separation and reliability, (2) Acceptable individual statistics (i.e., items and person infit and outfit), (3) Good category structure, and (4) Optimal distance between thresholds.

Recommendations

The current study had several recommendations for psychometricians and practitioners. Related to the psychometric side, future studies should conduct Differential Item Functioning Analysis to ensure that STRM items are reliable to use in different demographics (e.g., public/private schools, males/females, lower and higher grades). The findings should also be interpreted with caution as this study just used the liberal criteria; however, future studies could use more conservative criteria while providing more support to the psychometric properties of the STRM. On the psychological/educational side, the current study recommends educators use STRM as a data-driven approach that orient the modification of their teaching practices to strengthen their relations with students. Frequent assessment of students' perceptions about their relationships with their teachers provides an indicator about the aspects that require additional improvement. Conducting targeted and well-design training programs focusing on building constructive STRs is advised. Furthermore, running experimental studies that examine these courses' effects on students involvement and learning using STRM is highly suggested.

Limitations

This study had multiple limitations. That is, the study sample was limited to female students in rural middle schools. The scale assessed their relationship with science teachers only. No examination of Differential Item Functioning was conducted across different groups (e.g., gender, grades, and subjects). The current study used the liberal criteria when assessing items functionality.

References

Aldhafri, S., & Alhadabi, A. (2019). The psychometric properties of the student–teacher relationship measure for Omani grade 7–11 students. *Frontiers in Psychology, 10,* 2283. https://doi.org/10.3389/fpsyg.2019.02283

Al Harthy, H. (2019, January 21-23). *The orientations of the ministry of education in the Sultanate of Oman to keep up with the fourth industrial revolution* [Paper presentation]. The Fourth Industrial Revolution and its Impact on Education Conference, Ministry of Education, Sohar, Sultanate of Oman.

Allen, M., & Yen, W. (1979). *Introduction to measurement theory*. Brooks/Cole Publishing.

Al-Rubaie, S. (2019, January 21-23). *The orientations of education in Oman under the fourth industrial revolution* [Paper presentation]. The Fourth Industrial Revolution and its Impact on Education Conference, Ministry of Education, Sohar, Sultanate of Oman.

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43*(4), 561-73. https://doi.org/10.1007/BF02293814

Ang, R. (2005). Development and validation of the teacher-students relationship inventory using exploratory and confirmatory factor analysis. *The Journal of Experimental Education, 47*(1), 55-73.

Baker, J. (2006). Contributions of teacher–child relationships to positive school adjustment during elementary school. *Journal of School Psychology, 44*(3), 211-229.

Baker, J., 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.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice-Hall.

Berry, D., & O'Connor, E. (2010). Behavioral risk, teacher-child relationships, and social skill development across middle childhood: A child-by-environment analysis of change. *Journal of Applied Developmental Psychology, 31*(1), 1–14.

Bond, T., & Fox, C. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd ed.). Routledge.

Boone, W., & Staver, J. (2020). *Advances in Rasch analyses in the human sciences.* Springer.

Bowlby, J. (1969). *Attachment and Loss*. Basic Books.

Brewster, A., & Bowen, G. (2004). Teacher support and the school engagement of Latino middle and high school students at risk of school failure. *Child and Adolescent Social Work Journal, 21*(1), 47–67.

Bronfenbrenner, U. (1994). Ecological models of human development. *Readings on the Development of Children, 2*(1), 37-43.

Cranley-Gallagher, K., & Mayer, K. (2006). Teacher-child relationships at the forefront of effective practice. *Young Children, 61*(6), 44-49.

Crocker, L., & Algina, J. (2008). *Introduction to classical & modern test theory* (2nd ed.). Cengage Learning.

Davis, H. (2001). The quality and impact of relationships between elementary school students and teachers. *Contemporary Educational Psychology, 26*(4), 431–453.

De Ayala, R. (2009). *The theory and practice of item response theory*. The Guilford Press.

DiLalla, L., Marcus, T., & Wright-Phillips, M. (2004). Longitudinal effects of preschool behavioral styles on early adolescent school performance. *Journal of School Psychology, 42*(5), 385–401.

Fredricks, J., Blumenfeld, P., & Paris, A. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74*(1), 59–109.

Greogory, A., Allen, J., Mikami, A., Hafen, C., & Pianta, R. (2014). Effects of a professional development program on behavioral engagement of students in middle and high school. *Psychology in the Schools, 51*(2), 143-163.

Han, X. (2013, May 6). *Item Response Models Used within WinGen*. WinGen. https://www.umass.edu/remp/software/simcata/wingen/modelsF.html

Hattie, J. (2009). *A synthesis of over 800 meta-analyses relating to achievement*. Routledge.

Hill, P., Spiegel, A., McQuillan, J., & Diamond, J. (2018). Discovery orientation, cognitive schemas, and disparities in science identity in early adolescence. *Sociological Perspectives, 61*(1), 99-125.

Hughes, J. (2012). Teacher-student relationships and school adjustment: Progress and remaining challenges. *Attachment & Human Development, 14*(3), 319–327. https://doi.org/10.1080/14616734.2012.672288

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

Kinney, P. (2007). A voice from the middle. *Principal leadership, 8*(2), 35-36.

Lee, J. (2012). The effects of the teacher-student relationship and academic press on student engagement and academic performance. *International Journal of Educational Research, 53*, 330-340. https://doi.org/10.1016/j.ijer.2012.04.006

Lei, H., Cui, Y., & Chiu, M. (2016). Affective teacher-student relationships and students' externalizing behavior problems: A meta-analysis. *Frontiers in Psychology, 7*, 1311-1324. https://doi.org/10.3389/fpsyg.2016.01311

Lei, H., Cui, Y., & Chiu, M. (2018). The relationship between teacher support and students' academic emotions: A meta-analysis. *Frontiers in Psychology, 8*, 2288-2300 https://doi.org/10.3389/fpsyg.2017.02288

Linacre, J. (1999). Understanding Rasch measurement: Estimation methods for Rasch measures. *Journal of Outcome Measurement, 3*(4), 381-405.

Linacre, J. (2002). Understanding Rasch measurement: Optimizing rating scale category effectiveness. *Journal of Applied Measurement, 3*(1), 85-106.

Linacre, J. (2017). *Winsteps® (Version 3.93.0)* [Computer Software]. http://www.winsteps.com

Longobardi, C., Prino, L., Marengo, D., & Settanni, M. (2016). Student-teacher relationships as a protective factor for school adjustment during the transition from middle to high school. *Frontiers in Psychology, 7*(1988), 1-9. https://doi.org/10.3389/fpsyg.2016.01988

Longobardi, C., Settanni, M., Prino, L., Fabris, M., & Marengo, D. (2019). Students' psychological adjustment in normative school transitions from kindergarten to high school: Investigating the role of teacher-student relationship quality. *Frontiers in Psychology, 10*, 1238. https://doi.org/10.3389/fpsyg.2019.01238

Lynch, M., & Cicchetti, D. (1997). Children's relationships with adults and peers: An examination of elementary and junior high school students. *Journal of School Psychology, 35*(1), 81–99.

Mandinach, E. (2012). A Perfect Time for Data Use: Using Data-Driven Decision Making to Inform Practice. Educational Psychologist, 47(2), 71–85. https://doi.org/10.1080/00461520.2012.667064

McFarland, L., Murray, E., & Phillipson, S. (2016). Student-teacher relationships and student self-concept: Relations with teacher and student gender. *Australian Journal of Education*, *60*(1), 5–25. https://doi.org/10.1177/0004944115626426

Meehan, B., Hughes, J., & Cavell T. (2003). Teacher–child relationships as compensatory resources for aggressive children. *Child Development, 74*(4), 1145–1157.

Mikk, J., Krips, H., Saalik, U., & Kalk, K. (2016). Relationships between student perception of teacher-student relations and PISA results in mathematics and science. International *Journal of Science & Math Education, 14*(8), 1437-1454. https://doi.org/10.1007/s10763-015-9669-7

Mullis, I., Martin, M., Foy, P., Kelly, D., & Fishbein, B. (2020). *TIMSS 2019 international results in mathematics and science*. TIMSS 2019. https://timssandpirls.bc.edu/timss2019/international-results/

Murray, C., & Zvoch, K. (2011). Teacher-student relationships among behaviorally at-risk African American youth from low-income backgrounds: Student perceptions, teacher perceptions, and socioemotional adjustment correlates. *Journal of Emotional and Behavioral Disorders, 19*(1), 41-54.

Northup, J. (2011). *Teacher and student relationships and student outcomes* (Publication No. 3456052). [Doctoral dissertation, University of Colorado Denver]. ProQuest Dissertations and Theses database.

Pianta, R. (2001). *Student–Teacher Relationship Scale (STRM): Professional manual*. Psychological Assessment Resources, Inc.

Reddy, R., Rhodes, J., & Mulhall, P. (2003). The influence of teacher support on student adjustment in the middle school years: A latent growth curve study. *Development and Psychopathology, 15*(1), 119-138.

Ridwan, M., Marie-Christine, O., & Roel, B. (2014). Teacher–student interpersonal relationships do change and affect academic motivation: A multilevel growth curve modelling. *British Journal of Educational Psychology, 84*(3), 459–482.

Roorda, D., Koomen, H., Spilt, J., & Oort, F. (2011). The influence of affective teacher–student relationships on students' school engagement and achievement: A meta-analytic approach. *Review of Educational Research, 81*(4), 493-529.

Rubio, V., Hernández, J., Aguado, D., & Hontangas, P. (2007). Psychometric properties of an emotional adjustment measure: An application of the graded response model. *European Journal of Psychological Assessment, 23*(1), 39–46.

Sáez, L., Folsom, J., Al Otaiba, S., & Schatschneider, C. (2012). Relations among student attention behaviors, teacher practices, and beginning word reading skill. *Journal of Learning Disabilities, 45*(5), 418– 432.

Saft, E., & Pianta, R. (2001). Teachers' perceptions of their relationships with students: Effects of child age, gender, and ethnicity of teachers and children. *School Psychology Quarterly, 16*(2), 125-141.

Samejima, F. (2010). The general graded response model. In M. Nering & R. Ostini (Eds.), *Handbook of polytomous item response theory models* (pp. 77–107). Routledge/Taylor & Francis Group.

Schildkamp, K. (2019). Data-based decision-making for school improvement: Research insights and gaps. *Educational Research, 61*(3), 257–273. https://doi.org/10.1080/00131881.2019.1625716

Seaton, E. (2007). If teachers are good to you: caring for rural girls in the classroom. *Journal of Research in Rural Education, 22*(6), 1-16.

Silver, R., Measelle, J., Essex, M., & Armstrong, J. (2005). Trajectories of externalizing behavior problems in the classroom: Contributions of child characteristics, family characteristics, and the teacher–child relationship during the school transition. *Journal of School Psychology, 43*(1), 39–60.

Spence, R., Owens, M., & Goodyer, I. (2012). Item Response Theory and validity of the NEO-FFI in adolescents. *Personality and Individual Differences, 53*(6), 801-807.

Suldo, S., McMahan, M., Chappel, A., & Bateman, L. (2014). Evaluation of the teacher–student relationship inventory in American high school students. *Journal of Psycheducational Assessment, 32*(1), 3–14.

Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Wentzel, K. (1997). Student motivation in middle school: The role of perceived pedagogical caring. *Journal of Educational Psychology, 89*(3), 411-419.

Wright, B., & Linacre, J. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions, 8*(3), 370-371.

Zanon, C., Hutz, C., Yoo, H., & Hambleton, R. (2016). An application of item response theory to psychological test development. *Psychology: Research and Review/ Psicologia: Reflexão e Crítica, 29*(1), 1-10. https://doi.org/10.1186/s41155-016-0040-x