Development and Validation of the Student Well-Being Scale for Elementary School Using Rasch Model Analysis

Wahyu Widodo
Faculty of Psychology
Universitas Negeri Malang
Faculty of Education
Universitas Tribhuwana Tunggadewi
Indonesia
drwahyuwidodo@outlook.com
Corresponding author

I Nyoman Sudana Degeng
Faculty of Psychology
Universitas Negeri Malang
Indonesia
nyoman.sudana.d.fip@um.ac.id

Marthen Pali
Faculty of Psychology
Universitas Negeri Malang
Indonesia
marthenpali@yahoo.com

Imanuel Hitipeuw
Faculty of Psychology
Universitas Negeri Malang
Indonesia
imanuel.hitipeuw.fip@um.ac.id

Abstract

Efforts to realize student well-being are of particular concern in Indonesia. Unfortunately, a scale measuring student well-being for elementary school students is still scarce. This study aimed to develop and validate a scale for the well-being of elementary school students in Indonesia. The scale was developed based on the Student Well-being Model and analyzed using the Rasch Model. This study involved 414 respondents who were Grade 5 or Grade 6 elementary school students in Malang city. Based on the analysis of the Rasch model, there are 19 items (out of 28 items) whose values meet the criteria for outfit MNSQ, outfit ZSTD, and point measure correlation. This scale has a dimensionality value of 42.9%, item separation of 6.93, and item reliability of 0.98, and the results of the DIF analysis showed no bias toward gender (female vs. male) and grade (5^{th} vs. 6^{th}). Overall, it can be concluded that the Student Well-being in Elementary Schools scale has good psychometric properties, so it is suitable for measuring student well-being in elementary schools.

Keywords: Scale development, student well-being, elementary school, Rasch model analysis

Received 27 January 2024/Accepted 5 June 2024

Introduction

The construct of student well-being is a complex psychological concept with diverse understandings. However, several studies generally formulate the construct of student well-being based on two

Greek philosophical views: hedonic and eudaimonic (Holzer et al., 2024). From the hedonic view, well-being is defined as a person's actions that are more directed toward achieving satisfaction/pleasure and tend to avoid pain. Meanwhile, in the eudaimonic view, well-being is defined as a person's condition that refers more to actualizing human potential to face challenges and problems to achieve happiness in life (Khatri et al., 2024). Fraillon, (2004) added that by limiting the broadness of student well-being meaning, the construct of student well-being formulation can be based on four principles: the school context, dimensions, interrelationships between dimensions, and the ability of measurements to describe the level of student well-being.

Previous researchers have formulated many constructs of student well-being, but their comparisons have not been studied much. The literature review conducted in this study revealed that the construct of student well-being compiled by Fraillon, (2004), Karyani et al. (2015), Wibowo et al. (2021), and Konu & Rimpela (2002) particularly contains an explanation of its constituent dimension composition. Meanwhile, the construct of the Student Well-Being Model (SWBM) formulated by Soutter et al., (2014) includes a conceptual framework in the form of dynamics of the interrelationships among the dimensions composing the construct of student well-being. These findings indicate that SWBM is more suitable for use as a conceptual basis in research revealing the dynamics of well-being experienced by students during school. As in the SWBM framework, student well-being can be understood as a thriving feeling experienced by students due to their ability to manage emotions and cognitions to be positively involved in school activities.

Student well-being can be interpreted as an agency process, not just a result or an achievement (Busseri et al., 2009), so that its dynamics are influenced by various factors, including personal, behavioral and environmental factors. Personal factors that influence student well-being include academic self-efficacy (Serinci et al., 2023), school identification (Dunstone et al., 2024) and self-regulation (Rodríguez et al., 2022). Moreover, behavioral factors that influence students' well-being include physical activity (Isoard-Gautheur et al., 2023), sleeping behavior (Armand et al., 2021), and eating behavior (Morshed et al., 2022). Therefore, student well-being is also influenced by environmental factors, such as school climate (Nguyen et al., 2021), classroom climate (Wang et al., 2020), and teacher factors (Bilz et al., 2022).

Over time, student well-being has become the main indicator of the quality of education in various countries (Courtney et al., 2023; Govorova et al., 2020; Morinaj & Held, 2023). This finding is supported by research results showing that student well-being plays an important role in both student self-development at school and academic success at the next level (Cárdenas et al., 2022; Hossain et al., 2023; Kellock, 2020). In the context of education in Indonesia, student well-being has become the government's primary concern since 2020, as stated in the Peraturan Direktur Jenderal Guru Dan Tenaga Kependidikan Kementerian Pendidikan Dan Kebudayaan Nomor 3813/B. According to B1/Hk/2020 (Kementerian Pendidikan dan Kebudayaan, 2020), every school principal must improve their ability to actualize student well-being.

In turn, efforts to actualize student well-being require psychological measuring tools at every level of education (Svane et al., 2019). Without an effective measuring instrument, it is problematic to measure the level of student well-being experienced by students. Elementary schools play an important role in building life skills, communication skills, well-being, literacy and numeracy skills, motivation and engagement, students' sense of identity and ownership, and the achievement of learning outcomes (FitzPatrick et al., 2014; Martanti & Fatkhuronji, 2023). In addition, the educational process experienced by students in elementary schools has had an impact on the cognitive development and academic abilities of students at the next level (Griet Vanwynsberghe & Fraine, 2017; Pustjens et al., 2007).

Unfortunately, the tools for measuring student well-being that have been developed in Indonesia (Durrotunnisa et al., 2018; Faizah et al., 2020; Hidayah et al., 2016; Kurniastuti & Azwar, 2014) still contain some weaknesses in the conceptual framework and data analysis techniques used. The conceptual framework used to develop the scale does not follow the principles of formulating student well-being constructs. Nevertheless, it is based on subjective well-being (Diener, 1984) and psychological well-being (Ryff & Keyes, 1995). This indicates that the scale can only measure the level of student well-being but cannot reveal the dynamics of student well-being at school.

In addition, the previous scale is still being developed using classical measurement theory, which is known to have several weaknesses. Yu (2020) explained that classical measurement theory cannot

accurately measure a respondent's ability and item difficulty level. This is because there is a circular dependence, namely, the ability of the respondent to depend on the quality of an item and vice versa. Over time, the weaknesses of classical measurement theory have amended with the emergence of item response theory, one of which is the Rasch model analysis technique. Rasch model analysis has several advantages, including being able to provide linear measurements, handle missing data, provide an accurate estimation process, find errors (misfits) or outliers, and provide objective measurement instruments (Tesio et al., 2024). Yu (2020) also confirmed that Rasch model analysis can simultaneously predict the item difficulty level and the respondent's ability because in Rasch model analysis, the respondent's ability is not affected by the item. Conversely, the quality of the item is not affected by the respondent's ability.

Overall, the description above states that efforts to realize student well-being in Indonesia need to be supported by the availability of measuring tools in the form of a scale that can overcome previous weaknesses by using a conceptual framework and more data analysis techniques. Therefore, this study uses Rasch model analysis to develop and validate a scale of student well-being in elementary schools based on the conceptual framework of the student well-being model.

Method

Participants

This study involved 414 students in grades 5 and 6 of elementary school (M_{age} = 12 with SD = 0.76, age group = 10-14 years, male = 192, female = 222). The respondents came from 5 schools in 5 subdistricts in Malang city and were selected based on the cluster sampling technique (Cohen et al., 2018).

Procedure

The development of this scale was carried out in 3 stages based on the procedure compiled by Boateng et al. (2018). The first is the item development stage. The item mentioned in this study is an item of statement that is designed to measure a desired psychological construct. Item development is carried out deductively by first conducting a literature review to determine the theoretical

grounding of a construct to be measured. Item development begins with an operational definition of each construct dimension. Based on the operational definition, the indicators and items are then prepared. Ten educational psychologists reviewed the quality of the indicators and items. The second stage is scale development. At this stage, the scale was reviewed by ten students from the targeted population of grades 5 and 6 every five to ensure that the editorial items were easily understood and interpreted correctly by the students. Furthermore, data collection was carried out by distributing scales to respondents using digital forms. The third stage is the evaluation of the scale. At this stage, the instrument was analyzed based on Rasch modeling analysis.

Instrument

The measuring tool developed in this research is based on the conceptual model of the Student Well-Being Model (SWBM), which consists of 7 domains, namely, having, being, relating, feeling, thinking, functioning, and striving (Soutter et al., 2014). Having is a dimension related to skills and access to learning resources that students obtain either through effort or through giving. Being is the student's intrapersonal dimension, which focuses on the trajectory of self-development, including who I am, what I am like now, and what I want to be like in the future. Relating is a student's interpersonal dimension that influences experiences, emotions, thoughts and action choices. Feeling is the student's ability to express and manage emotional states experienced during school. Thinking is a dimension of cognition that influences students' decisions to be actively involved in developing thinking skills in an effort to achieve meaningful learning. Functioning is a dimension of student behavior that involves extensive active involvement in carrying out student functions as members of a learning community. Striving is the dimension of student motivation and behavior in their efforts to achieve the desired goals.

Data Analysis

The measuring instruments used in this research were analyzed via Rasch modeling. Rasch modeling is a model that is capable of achieving objective measurements. This means that the instrument is able to measure respondents' abilities without relying on items. Likewise, the instrument is able to determine item difficulty levels without depending on the respondent (Tesio et al., 2024).

The data in this research were produced in polytomous form, namely, responses that have more than 2 categories expressed in the form of a Likert scale assessment, so that this research is a Rasch analysis with rating scale modeling. The scale developed in this research uses four Likert scale assessments, namely, very inappropriate (sts), not suitable (ts), suitable (s), and very suitable (ss) (Boone, Jr & Boone, 2012).

The data in this study were analyzed using WINSTEPS software version 5.2.4. WINSTEPS estimates parameters with PROX and Joint Maximum Likelihood Estimation (JMLE) with Newton–Raphson Iteration (Linacre, 2024). This estimation method has proven robust against missing data and allows easy incorporation into one analysis of data generated by variants of the Rasch model (Linacre, 1999). In addition, WINSTEPS is a Rasch analysis software that is popularly used by academics because it provides various analysis outputs so that it has a wealth of information that is able to explain the quality of an instrument based on Rasch modeling in a comprehensive manner.

The raw data from this research were analyzed in three stages (Widodo & Chotimah, 2023). The first stage is the data cleaning stage. At this stage, the quality of the respondents was checked through person-fit analysis (Dabaghi et al., 2020) and the Guttman scalogram (Linacre, 2024). The second stage included analysis at the item level. At this stage, item fit analysis is carried out (Rangka et al., 2023), and DIF items are identified (Aryadoust et al., 2024). In the third stage, analysis was performed at the instrument level. At this stage, unidimensionality (Aryadoust et al., 2021), local independence (Aryadoust et al., 2021), rating (Bond & Fox, 2015), person measurement (Rangka et al., 2023), person separation (Davis & Boone, 2021), person reliability (Davis & Boone, 2021), item measurement (Rangka et al., 2023), item separation (Davis & Boone, 2021), item reliability (Davis & Boone, 2021), Cronbach's alpha (Akhtar & Sumintono, 2023), Wright map analysis (Kim & Cho, 2022), and examination of the test information function (TIF) (Putra & Retnawati, 2020) are checked.

Results

Item Development Stage

The item development process begins by compiling 28 items based on 14 indicators consisting of seven dimensions of the Student Well-being Model (SWBM). In this stage, the quality of item development is reviewed based on the suitability of the framework of a conceptual construct, operational definitions of each dimension of a construct, indicators of each dimension, and editorial items of each indicator. The initial examination of the quality of item development involved ten educational psychologists; two were professors in psychology, and eight were scientists (doctoral degrees) in educational psychology. As a result, all the indicators developed represent the seven dimensions contained in the SWBM. However, considering the age of the respondents who were still children, there were editorial improvements in 5 items, as presented in Table 1 below.

Table 1
List of Item Wording Improvements

Item	Item Wording Before Improvement	Item Wording After Improvement
A1	saya mudah mendapatkan fasilitas belajar di sekolah	Saya mudah mendapatkan fasilitas belajar di sekolah seperti buku pelajaran, kuota internet, jaringan wifi, perlengkapan olahraga, peralatan musik, dan lainnya yang sejenis
A3	Saya mampu memahami isi bacaan dengan baik	Saya mampu memahami isi pelajaran dengan baik
D4	Saya mau memaafkan kesalahan teman	saya bersedia memaafkan kesalahan teman
G1	Saya menambah jam belajar di rumah	saya secara teratur dan disiplin belajar di rumah
G4	Saya mengikuti pelatihan yang diselenggarakan sekolah secara rutin	saya mengikuti berbagai pelatihan yang diselenggarakan sekolah secara rutin

Scale Development Stage

In the scale development stage, there was an instrument quality check involving ten students from the target population who were in grades 5 and 6 of each of the five elementary schools. As a result, the ten students stated that the editorial of each item on the instrument was easy to understand and could be interpreted well. Furthermore, data collection was carried out by distributing instruments to respondents using digital forms. As a result, the instrument was completed by 414 students—202

students in Grade 5 and 212 students in Grade 6. The respondents comprised 222 female and 192 male students aged 10-14 years.

The obtained data were then analyzed to determine the respondents' quality in providing answers using the person fit test and scalogram. The person fit test was examined using the criteria for the outfit mean square (MNSQ), outfit z-standard (ZSTD), and point measure correlation (Pt-Measure Corr) on the quality of the responses given by 414 respondents. Based on the MNSQ outfit value, 116 respondents identified misfits, with 56 respondents having an MNSQ outfit value > 1.5 and 60 respondents having an MNSQ outfit value < 0.5. Meanwhile, based on the outfit ZSTD value, 88 respondents were identified as misfits, with 39 respondents having an outfit ZSTD value > 2.0 and 49 respondents having an outfit ZSTD value < -2. Furthermore, the examination conducted based on the Pt-Measure Corr value resulted in the identification of 138 respondents as misfits, with 137 respondents having a Pt-Measure Corr value < 4 and 1 respondent having a Pt-Measure Corr value > 0.85. Finally, the quality of the respondents was checked with a scalogram. The results show that 41 respondents indicated answering in an unnatural pattern. Some of them are presented in Table 2 below.

Table 2
Examples of unreasonable respondents' answer patterns

Respondent Number	Respondent's answer*
413	44444444444444444444444444
156	22222222222222222222222222
126	44442443442333334233224422l4

Note:
* written according to the item number from the easiest to the most difficult sequentially from left to right (08, 13, 11, 17, 16, 24, 02, 10, 09, 04, 22, 07, 18, 14, 05, 25, 03, 12, 15, 01, 23, 21, 20, 19, 06, 28, 27, 26)

Table 2 shows an example of the pattern of respondents' answers to numbers 413 and 156, which looks unnatural because they give an even response to all items. This indicates that respondents 413 and 156 did not answer the item seriously. Likewise, the pattern of respondent's answer number 126 looks unnatural because it gives an inappropriate response. Respondent number 126 answered 1 (strongly disagree) on item number 27, whereas for the most difficult item to agree on (item 26), this respondent answered the opposite (very appropriate). Therefore, from the person fit test, 225 of

the 414 respondents were identified as misfits and eliminated. Therefore, 189 respondents were included in the item quality analysis.

At the scale development stage, after analyzing the quality of the respondents, the item quality analysis was then carried out using item fit and differential item function (DIF) analysis. The item quality checks are based on item fit with three criteria, namely, outfit mean square (MNSQ), outfit z-standard (ZSTD), and point-measure correlation (Pt-Measure Corr). The result based on the outfit MNSQ value revealed that the item with code A1 had an outfit MNSQ value > 1.5, so the item was inappropriate and eliminated. Item A1 has an editorial titled "*Saya mudah mendapatkan fasilitas belajar di sekolah seperti buku pelajaran, kuota internet, jaringan wifi, perlengkapan olahraga, peralatan musik, dan lainnya yang sejenis*". Furthermore, based on the outfit ZSTD values, it was found that items with codes A1 and C4 had outfit ZSTD values > 2.0, and items coded B3, E3, and G2 had outfit ZSTD values < -2.0, so they were inappropriate and eliminated. Item B3 has an editorial, "*Saya mampu menyesuaikan diri dengan situasi di sekolah*". Item C4 has an editorial, "*Saya sering bercerita tentang sekolah saya kepada teman sepermainan di rumah*". Item E3 has an editorial, "*Saya aktif menjawab pertanyaan guru*". Item G2 has an editorial, "*Saya aktif menjawab pertanyaan guru*". The item fit test is then carried out by checking the Pt-Measure Corr value. The results show that the items coded AI, A3, B4, and D4 have a Pt-Measure Corr value < 0.4, so they were misclassified and eliminated. Item A3 has an editorial, "*Saya mampu memahami isi pelajaran dengan baik*". Item B4 has an editorial, "*Saya ingin menjadi siswa berprestasi*", and item D4 has an editorial, "*Saya bersedia memaafkan kesalahan teman*". In the item fit test stage, eight items were eliminated, and 20 items met all the required criteria.

Next, 20 items that passed the item fit test were examined using DIF analysis on gender and class identity data. DIF analysis was carried out by examining the DIF contrast value and Rasch–Welch p value. DIF analysis of gender data revealed 5 items with DIF contrast values > 0.5, namely, item B4 (1.31), item C3 (0.54), item D4 (0.71), item E1 (0.60), and item F4 (0.61). However, the Rasch–Welch p values for these five items were not significantly different (p>0.05). Therefore, these five items are maintained. Moreover, DIF analysis of class identity revealed that there was 1 item that had a DIF

contrast > 0.5, namely, item A2 (0.79). The Rasch–Welch p value for this item was statistically significant ($p<0.05$). Therefore, item A2 is eliminated.

Item A2 has an editorial, "*Saya mudah menemui guru selama di sekolah*". Thus, from item fit and DIF tests, it was found that nine out of 28 items with codes A1, A2, A3, B3, B4, C4, D4, E3, and G2 indicated a mismatch and were eliminated. Thus, 19 items met all the item quality test criteria. The analysis results of the 19 items are presented in Table 3 below.

Scale Evaluation Stage

Finally, the last stage is the scale evaluation stage. At this stage, the scale's quality is checked using unidimensionality, local independency, a rating scale, person separation, person measurement, person reliability, Cronbach's alpha, item separation, item measurement, item reliability, Wright map analysis, and the test information function (TIF). The scale quality analysis was carried out based on the answers from 189 respondents who passed the respondent quality test and 19 items that passed the item quality test.

The results of the dimensionality measurement show a raw variance value of 42.9%, or more than 40%, which indicates that the dataset belongs to the better category. The results of the Cronbach's alpha measurement show a value of 0.86 or more than 0.80, which is considered very good. The measurement results of unexplained variance in the 1st and 2nd contrasts show values of 5.2% and 4.7%, respectively, or do not exceed the 15% threshold, meaning that the variance that the instrument cannot explain meets the ideal requirements. This is also supported by the eigenvalue of unexplained variance in the first contrast of 1.93 (<2), which means that there is no substantive structure that disrupts the integrity of the instrument. Likewise, by examining the local independence analysis, the largest residual correlation values between items were obtained, namely, 0.38 and 0.16 (<0.70), which means that there is no dependence between items.

Table 3
List of Items that Passed the Item Fit Test and DIF

Item	Item Wording	Outfit		Pt-Measure Corr	Rasch-Welch *p*-value	
		MNSQ	ZSTD		Gender	Grade
A4	Saya mampu menjalin komunikasi yang baik dengan siapapun	0,84	-0,97	0,48	0,42	0,18
B1	Saya bersedia membantu teman yang mengalami kesulitan belajar	1,10	0,82	0,42	0,36	0,74
B2	Saya bersedia ditunjuk kelompok untuk mempresentasikan hasil kerja di kelas	0,99	-0,05	0,58	0,58	044
C1	Saya meminta bantuan guru bila menemui kesulitan belajar	0,92	-0,42	0,49	0,14	0,21
C2	Saya mampu bekerja kelompok di kelas dengan siapa saja	1,03	0,26	0,48	0,75	0,44
C3	Saya merasa betah bersekolah disini	0,80	-0,62	0,42	0,20	0,61
D1	Saya bersyukur dapat bersekolah disini	0,58	-1,17	0,41	0,86	0,49
D2	Saya merasa nyaman dengan perlakuan teman-teman	1,01	0,14	0,49	0,16	0,83
D3	Saya berbagi cerita dengan teman di saat senang ataupun sedih	1,11	0,95	0,57	0,50	0,36
E1	Saya bersemangat mengikuti pembelajaran di kelas	0,65	-1,35	0,49	0,14	0,27
E2	Saya mampu menjaga konsentrasi selama guru menjelaskan materi	0,92	-0,57	0,48	0,39	0,12
E4	Saya aktif mencatat penjelasan guru	0,86	-1,35	0,61	0,38	0,61
F1	Saya senang membantu guru menyiapkan pembelajaran di kelas	1,03	0,30	0,50	1,00	0,69
F2	Saya aktif mengerjakan tugas kelompok	0,80	-1,43	0,55	0,86	0,13
F3	Saya aktif mengikuti semua acara sekolah	0,96	-0,32	0,60	0,53	0,43
F4	Saya aktif menjalankan piket sekolah	0,92	-0,30	0,41	0,10	1,00
G1	Saya secara teratur dan disiplin belajar di rumah	0,89	-,93	0,52	0,72	0,71
G3	Saya bersedia ditunjuk oleh sekolah untuk mengikuti kompetisi	1,13	1,25	0,60	0,41	0,29
G4	Saya mengikuti berbagai pelatihan yang diselenggarakan sekolah secara rutin	1,11	1,02	0,57	0,63	0,07

Note:
A = Having dimension, B = Being dimension, C = Relating dimension, D = Feeling dimension, E = Thinking dimension, F = Functioning dimension, G = Striving dimension

An examination of the rating scale based on the Rasch-Andrich threshold value shows "NONE" or the 1st choice rating score (very unsuitable), logit -1.76 for the 2nd choice rating score (not appropriate), logit 0.61 for the score of choice 3 (appropriate), and logit 2.38 for the score of choice

4 (very suitable). The analysis results indicate that the respondents could declare the four ranking options (very inappropriate, not appropriate, appropriate, and very appropriate) valid or not confusing.

The analysis of respondents and items followed the evaluation stage items. The respondents were measured by person measurement, person separation, and person reliability. The results of the person measurement showed the highest value of 7.03 logit owned by the respondent with the code 015PF. This means that female respondents in grade 6 with serial number 15 gave more appropriate and very appropriate answers in the given scale instrument or had the highest level of student well-being compared to others. Meanwhile, the lowest value of -0.20 logit is owned by the respondent with the code 375LE. The respondents of 5th-grade male students with serial number 375 gave more inappropriate and highly inappropriate answers to the given scale instrument or had the lowest student well-being compared to the other respondents. Moreover, based on the average value obtained of 2.84 logit, 89 students had a well-being level above the average, and 100 students had a well-being level below the average.

Meanwhile, the results of the person separation measurement show a separation value of 2.12, and the results of the person reliability measurement show a value of 0.82, which means that the arrangement of the items successfully divides respondents into 2 groups with a good degree of reliability. This is also supported by the results of Cronbach's alpha measurements, which show a value of 0.86 or more than 0.80, which means that the interaction between respondents and the instrument has reliability, which is classified as very good.

After the respondent analysis at the scale evaluation stage was completed, item analysis was then carried out using item measures, item separation and item reliability. As a result, in examining the value of the item measure, the highest value was obtained at 1.84 logit for the item with the code G3. This means that item 3, which represents an indicator of the striving dimension, is the item on which it is most difficult for respondents to agree. Item G3 has the editorial "I am willing to be appointed by the school to take part in the competition". Meanwhile, the lowest value obtained was -2.37 logit for the item with code D1. This means that item 1, which represents an indicator of the

feeling dimension, is the item that is easiest for respondents to agree with. Item D1 has the editorial "I am grateful to be able to study here". In addition, based on the average value obtained for the 0.0 logit, 11 items (B1, B2, D2, D3, E2, E4, F1, F3, G2, G3, and G4) had above-average item difficulty levels. -average and 8 items (A4, C1, C2, C3, D1, E1, F2, F4) with item difficulty levels below average. Meanwhile, item separation showed a value of 6.93, and item reliability showed a value of 0.98, which means that the sample respondents were able to divide items into 7 levels with a very good degree of reliability.

The scale evaluation was continued with Wright Map analysis. As shown in Figure 1, the items that are most difficult to agree upon have a chance of being agreed upon by respondents with medium to high ability, and the items that are easiest to agree on have a chance of being agreed upon by all respondents, even respondents with high ability. lowest ability. This means that this instrument is suitable for students with medium to lower abilities or who are considered to have a medium to lower level of student well-being.

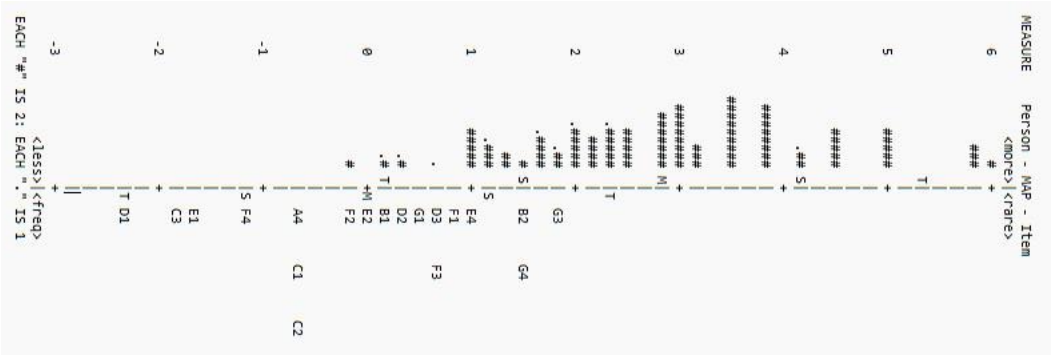

Figure 1. . Wright Map of The Student Well-being Scale for Elementary School

This is also supported by test information function (TIF) analysis. As shown in Figure 2, based on the TIF analysis, the instrument can reveal a large amount of information if it is applied to respondents with moderate to low abilities.

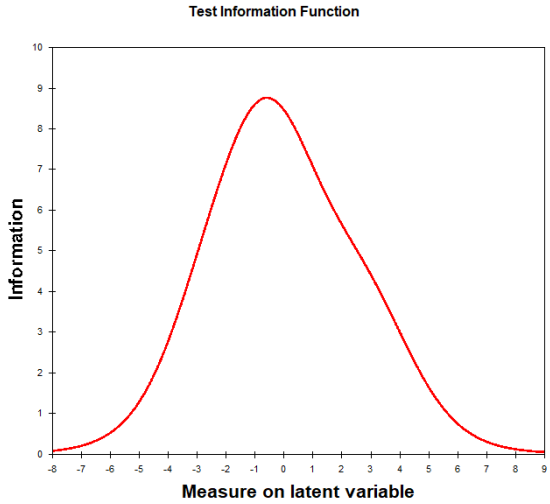

Figure 2. Test Information Function of The Student Well-being Scale for Elementary School

To make it easier to read the data at a glance, the results of the analysis at the scale evaluation stage are summarized in Table 4 as follows:

Table 4

Summary of the Scale Evaluation Stage Analysis Results

	Criterion	Value
Instrument	Alpha Cronbach value	0,86
	Raw variance	42,9%
	Unexplained variance in 1st contrast	5,2%
	Unexplained variance in 2nd contrast	4,7%
Respondent*	Separation value	2,12
	Reliability value	0,82
	Highest logit value	6,14
	Lowest logit value	0,10
Item**	Separation value	6,93
	Reliability value	0,98
	Highest logit value	1,84
	Lowest logit value	-2,37

Note: * 189 students; ** 19 items

Discussion

The development of items in this scale is based on the conceptual framework of the Student Well-being Model (SWBM) (Soutter et al., 2014). However, Wibowo et al. (2021) stated that student well-being needs to be understood in the educational context of a country. However, the Student Well-being Model (SWBM) is still worthy of being used as a conceptual basis for developing a universal scale because the structure of the conceptual model is able to explain the dynamics between the aspects that make it up. In addition, SWBM was developed based on two philosophical views, namely, hedonic and eudaimonic, which can represent the domains of students' subjective and psychological well-being. Another advantage is that SWBM is formulated from research conducted in school environments where the curriculum is specifically designed to realize student well-being so that it can comprehensively represent the meaning of well-being experienced by students.

In SWBM, the dynamics of student well-being are described as interactions between three-dimensional categories (assets, appraisals, and actions) within a complex system framework (Parrot, 2002). Assets are defined as external variables in the form of material and nonmaterial resources such as learning facilities and interpersonal and intrapersonal skills that students can access and utilize to achieve well-being. Assets consist of 3 dimensions, namely, having, being, and relating. Appraisals are defined as a form of affective and cognitive interpretation of the assets owned regarding how and why the dimensions of having, being, and relating are valuable for achieving well-being. Appraisals consist of 2 dimensions, namely, feeling and thinking. Actions are defined as motivations and actions that arise from the results of affective and cognitive assessments of the assets owned in achieving and maintaining well-being. Actions consist of 2 dimensions, namely, functioning and striving. In line with Soutter et al., (2014), the student well-being referred to in this study is the feeling of happiness and development experienced by students since they can utilize their assets to manage emotions and cognition to be positively involved in activities at school.

Furthermore, in this study, the seven dimensions of SWBM are defined as follows: (1) Having is a dimension related to skills and access to learning resources obtained by students either through effort or giving; (2) Being is a student's intrapersonal dimension that focuses on self-development trajectories, including who I am, what I am now, and what I want to be in the future; (3) Relating is a

student's interpersonal dimension that affects experiences, emotions, thoughts, and choices of actions to achieve prosperity; (4) Feelings are students' abilities to express and manage the emotional states experienced at school; (5) Thinking is a cognitive dimension that influences students' decisions to be actively involved in developing thinking skills in their efforts to achieve meaningful learning; (6) Functioning is a dimension of student behavior that shows active involvement in carrying out their functions as members of the learning community at school; and (7) Striving is a dimension of students' motivation and behavior in their efforts to achieve the desired goals.

In the item development stage, a review from 10 educational psychologists showed that all of the indicators compiled were appropriate or could represent each dimension, as referred to in the conceptual framework of the Student Well-Being Model. However, there are five items whose editorials are not in accordance with the language development stage of elementary school-age children. Thus, at this stage, the arrangement of indicators can be maintained, and only editorial improvements are made on five items. Expert review is an important part of the item development process, which aims to ensure that items reveal the desired construct (Haynes et al., 1995; Vucaj, 2022). However, Morgado et al. (2017) reminds us that despite the large contribution of expert reviews at the item development stage, it is also important to consider input from the target population through interviews or focus group discussions. Therefore, this study also interviewed ten students from the target population, intending to provide feedback on their understanding of the given instrument. As previously explained, ten students from the target population stated that all items in the instrument could be easily understood as worthy of being distributed to respondents.

Furthermore, in the scale development stage, an analysis of respondents' quality and item quality was carried out. The quality of the respondents was analyzed using the person fit test (Dabaghi et al., 2020) and scalogram (Linacre, 2024). The person fit test aims to determine the consistency of the respondent's item response pattern to several proposed models of valid item responses (Yu, 2020). The person fit test refers to the criteria of the outfit mean square (MNSQ), outfit z-standard (ZSTD) (Darman et al., 2024), and point measure correlation (Pt-Measure Corr) (Qudratuddarsi et al., 2022). The Guttman scalogram analysis in this study aimed to determine the pattern of unfair answers made by respondents (Linacre, 2024). At the research stage on the quality of the respondents, 225 of the

414 respondents were eliminated because misfits were detected or because they did not meet the three criteria of person fit. That is, the response patterns of the 225 respondents were inconsistent, unpredictable, and did not function in determining the quality of the items (Chan et al., 2021). Thus, from the respondent quality test stage, 189 respondents met the requirements to be used as material for analysis at the scale evaluation stage.

The item quality analysis was then performed using the item fit test (Davis & Boone, 2021) and DIF identification (Zhu & Aryadoust, 2022). Item fit analysis was carried out to determine the suitability of items, which showed the extent to which responses to certain items were consistent with how respondents responded to other items (Yu, 2020). Item quality analysis using the item fit test was carried out according to the following criteria: outfit mean square (MNSQ), fit z-standard (ZSTD) (Darman et al., 2024), and point measure correlation (Pt-Measure Corr) (Qudratuddarsi et al., 2022). The results of the item fit measurement showed that 8 out of 28 items were detected as misfits because they did not meet the three criteria and were therefore eliminated. This means that eight items do not match the desired model, cannot provide reasonable predictions, and do not function in revealing the desired construct. In addition, to ensure that the items on the scale did not contain gender or class identity bias, a DIF analysis was carried out (Zhu & Aryadoust, 2022). The results of the DIF analysis show that there is 1 item that contains class identity bias. It cannot accurately measure students' abilities at different grade levels, so they are eliminated. Thus, based on item quality analysis using item fit and DIF tests, 19 items were found to be unidimensional (they were able to reveal student well-being constructs) and did not contain gender and class identity bias, so they met the requirements to be used as material for analysis at the scale evaluation stage.

At the scale evaluation stage, the instrument was analyzed for unidimensionality, local independence, rating scale, person measure, person separation, person reliability, item measure, item separation, item reliability, Cronbach's alpha, Wright Map, and the test information function (TIF). The dimensionality test is carried out to ensure that all items in the instrument only measure one construct or meet the unidimensionality requirements (Aryadoust et al., 2021). Moreover, local independence analysis is carried out to ensure that there is no dependency between items in the instrument being tested (Aryadoust et al., 2021). The results of the unidimensionality measurement

show that the raw variance obtained is 42.9%, and the observed unexplained variance in the 1st contrast value is 5.8%, with an eigenvalue of 1.9 (<2). This means that the scale developed has succeeded in revealing only one construct (student well-being) with a relatively good degree. This is also supported by local independence analysis through examining the highest correlation value between items, which is less than 0.70, which means that there is no dependency between items; in other words, respondents do not immediately respond to an item based on their response to another item (Linacre, 2024).

Furthermore, an analysis was carried out using a rating scale that aimed to verify the accuracy of the assessment in representing a set of rating scales for all items (Bond & Fox, 2015). The results of the rating scale analysis showed that the 4 rating options (strongly disagree, disagree, agree, strongly agree) proved to be valid for respondents. This means it is not confusing for respondents.

Furthermore, the instrument was analyzed in terms of person measurement, person separation, and person reliability. The person-measure analysis aims to determine the level of student well-being experienced by each student (Rangka et al., 2023). Person measure analysis is very useful for classifying student well-being levels, so it can be used as a basis for formulating an action for low-well-being students. As revealed in this study, 89 students had a well-being level above average, and 100 students had a well-being level below average. Thus, most elementary school students in Malang city have not succeeded in achieving student well-being. The results of this person measure analysis need to be studied further along with the results of the item measure analysis to obtain a more complete explanation of the level and dynamics of student well-being experienced by students.

In addition, this instrument is also analyzed using person–person separation. Person separation analysis aims to determine how well respondents can spread among different groups (Davis & Boone, 2021). A greater separation value is followed by better instrument quality. The person separation value analysis revealed a value of 2.12. The value means that the instrument can be accepted because it can group respondents into two groups.

Next, the instrument was analyzed using people's reliability measurements to determine respondents' consistency in providing answers to the instrument (Davis & Boone, 2021). The results of the reliability analysis of people show that the obtained value is 0.82, which means that the reliability target of people in the good category is met. Thus, the consistency of respondents with this instrument can be considered reliable.

After the respondent quality analysis, the instrument analysis continued to include item quality analysis using item measures, item separation, item reliability, Wright maps, and test information functions (TIFs). Item separation analysis aims to determine to what extent the items are classified based on their difficulty level (Davis & Boone, 2021); the more significant the item value separation is, the better the quality of the items to be replicated in other samples (Bond & Fox, 2015).

The item measure analysis is conducted to determine the item's difficulty level. The results of the item measure analysis are helpful for identifying the dimensions of student well-being that are difficult for students to achieve, which is reflected in the items that students hardly agree on. As revealed in this study, the item coded G3 represents a striving dimension indicator, which is a challenging item that respondents quickly agree with. This result means that the educational process taken place in elementary schools in Malang city has not succeeded in maintaining the motivation and actions of students in their efforts to achieve student well-being. On the other hand, the item coded D 1, which represents the indicator of the feeling dimension, is the item the respondent most easily approves of. This means that the educational process in elementary schools in Malang city has succeeded in providing positive affective experiences for students. Meanwhile, when viewed from the average value of the item size, it appears that only eight items have a logit value below the average value or tend to be quickly approved by respondents. In contrast, the other 11 items have logit values above the average or tend to be challenging for respondents to agree with. Thus, primary school education services in Malang city have not succeeded in realizing student well-being, especially in the dimensions of being, feeling, thinking, functioning, and striving.

In more detail, the results of the item measure analysis show that on the dimension of being, respondents tend to find it challenging to give an agreeable answer to items coded B1 and B2. The

dimension of being represents students' ability to identify who they are, what they are like now, and what they want to be in the future. In other words, primary school education services in Malang city have not succeeded in helping students identify themselves, so it is not easy to develop a positive self-concept. In the feeling dimension, although there are items that most respondents quickly approved (items coded D1), there are items that tend to be difficult for respondents to agree with (items coded D2 and D3). The two items that are difficult for respondents to agree on are items that represent students' ability to manage their emotions. In the thinking dimension, although some items seem to be quickly approved by respondents (E1), most of the items (E2 and E4) tend to be challenging to agree with. Thus, primary school education services in Malang city have not succeeded in increasing students' emotional regulation abilities. According to previous item analysis, primary school education services in Malang city have failed to increase learning satisfaction and improve students' literacy skills. In the functioning dimension, respondents tend to find it difficult to provide answers to items coded F1 and F3. These two items are representations of students' abilities to carry out their functions as members of the learning community at school. This means that primary school education services in Malang city have not succeeded in increasing the active involvement of students in school activities. In the striving dimension, respondents tend to have difficulty providing answers to items coded G2, G3, and G4. The item represents students' motivation and actions to develop themselves to achieve the desired goals. This means that elementary school education services in the city of Malang have not succeeded in maintaining or increasing student motivation to develop their abilities to the fullest extent.

The results of the item measure analysis are then interpreted along with the person measure results that have been previously measured based on the SWBM conceptual framework. This is done to obtain a complete explanation of the dynamics of student well-being experienced by students. The item and person measure analysis results show that elementary school students in Malang City have sufficient learning resources and can use them optimally to have adverse cognitive abilities based on low motivation to develop themselves to the fullest.

Next, the instrument was examined using item separation analysis. Item separation analysis aims to determine the extent to which items can be grouped based on their level of difficulty, with reference

to the greater the item separation value, the better the quality of the items if replicated in other samples (Davis & Boone, 2021). The item separation analysis results show that the separation value obtained is 6.93, which means that this instrument is acceptable because it has 7 item groups (Bond & Fox, 2015).

In addition, the instrument was examined using the item reliability test. The item reliability test measures how well an item can be relied on in the same treatment from different samples (Davis & Boone, 2021). According to the analysis results, a very high reliability value of 0.98 was obtained. This value indicates that the target value of the item reliability was met, which indicates that the quality of the items in this instrument is classified as reliable to the respondents (Putra & Retnawati, 2020).

The instrument in this research was also examined for Cronbach's alpha values to measure the reliability of the instrument based on the respondent's interaction with the items as a whole (Akhtar & Sumintono, 2023). The resulting Cronbach's alpha value was 0.89, which means that the instrument developed was classified as reliable with a very good rating.

The quality of this instrument was also checked using Wright Map analysis to determine the relationship between the latent characteristics of respondents and the location of the items (Kim & Cho, 2022). As a result, the arrangement of goods is even. However, the location of the items is for respondents with medium to low ability levels. This means that the composition of the items is considered good for uncovering the construct of student well-being and is more suitable for measuring the level of student well-being in lower middle class elementary school students. This is also in line with the test information function (TIF) curve, where the peak of the curve is on the left (located at -1 logit) with wide coverage. This means that the instrument will reveal more information if used on students with a wide range of middle to lower levels of well-being (Putra & Retnawati, 2020).

Conclusion

The analysis of the Rasch Model on the student well-being scale for elementary schools shows that it meets all the requirements of good psychometric properties. The seven dimensions of the SWBM

are retained because they are proven to fit the desired model in the Rasch model analysis. Thus, this scale is appropriate for use as an assessment and research instrument to determine the level and dynamics of student well-being experienced by each student, including measuring educational programs to what extent elementary schools have succeeded in realizing student well-being. In addition, based on the analysis of the Rasch Model, it is also known that elementary education services in Malang City have not succeeded in realizing student well-being. Therefore, it is advisable to rearrange educational programs, especially those that can improve students' abilities in the dimensions of being, feeling, thinking, functioning, and striving. However, the use of this scale for assessments in lower grades of elementary school needs to be supported by further research.

Acknowledgment

We would like to thank the Ministry of Education, Culture, Research, and Technology for providing the BPP-DN S-3 Scholarship to the first author so that he could complete all stages of this research well.

References

Akhtar, H., & Sumintono, B. (2023). Rasch analysis of the International Personality Item Pool Big Five Markers Questionnaire: Is longer better? *Primenjena Psihologija*, *16*(1), 3–28. doi: 10.19090/pp.v16i1.2401

Armand, M. A., Biassoni, F., & Corrias, A. (2021). Sleep, Well-Being and Academic Performance: A Study in a Singapore Residential College. *Frontiers in Psychology*, *12*, 672238. doi: 10.3389/fpsyg.2021.672238

Aryadoust, V., Min, S., & Chen, X. (2024). Investigating differential item functioning across interaction variables in listening comprehension assessment. *Studies in Educational Evaluation*, *80*, 101322. doi: 10.1016/j.stueduc.2024.101322

Aryadoust, V., Ng, L. Y., & Sayama, H. (2021). A comprehensive review of Rasch measurement in language assessment: Recommendations and guidelines for research. *Language Testing*, *38*(1), 6–40. doi: 10.1177/0265532220927487

Bilz, L., Fischer, S. M., Hoppe-Herfurth, A.-C., & John, N. (2022). A Consequential Partnership: The Association Between Teachers' Well-Being and Students' Well-Being and the Role of Teacher Support as a Mediator. *Zeitschrift Für Psychologie*, *230*(3), 264–275. doi: 10.1027/2151-2604/a000497

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best Practices for Developing and Validating Scales for Health, Social, and Behavioral Research: A Primer. *Frontiers in Public Health*, *6*(149), 1–18. doi: 10.3389/fpubh.2018.00149

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch Model Fundamental Measurement in the Human Sciences* (Third Edition). Routledge.

Boone, Jr, H. N., & Boone, D. A. (2012). Analyzing Likert Data. *Journal of Extension*, *50*(2). doi: 10.34068/joe.50.02.48

Busseri, M. A., Sadava, S., Molnar, D., & DeCourville, N. (2009). A Person-Centered Approach to Subjective Well-Being. *Journal of Happiness Studies*, *10*(2), 161–181. doi: 10.1007/s10902-007-9072-3

Cárdenas, D., Lattimore, F., Steinberg, D., & Reynolds, K. J. (2022). Youth well-being predicts later academic success. *Scientific Reports*, *12*(1), 2134. doi: 10.1038/s41598-022-05780-0

Chan, S.-W., Looi, C.-K., & Sumintono, B. (2021). Assessing computational thinking abilities among Singapore secondary students: A Rasch model measurement analysis. *Journal of Computers in Education*, *8*(2), 213–236. doi: 10.1007/s40692-020-00177-2

Cohen, L., Manion, L., & Morrison, K. (2018). *Research Methods in Education*. Routledge. doi: 10.4324/9781315456539

Courtney, M. G. R., Hernández-Torrano, D., Karakus, M., & Singh, N. (2023). Measuring student well-being in adolescence: Proposal of a five-factor integrative model based on PISA 2018 survey data. *Large-Scale Assessments in Education*, *11*(1), 20. doi: 10.1186/s40536-023-00170-y

Dabaghi, S., Esmaielzadeh, F., & Rohani, C. (2020). Application of Rasch Analysis for Development and Psychometric Properties of Adolescents' Quality of Life Instruments: A Systematic Review. *Adolescent Health, Medicine and Therapeutics*, *11*, 173–197. doi: 10.2147/AHMT.S265413

Darman, D. R., Suhandi, A., Kaniawati, I., Samsudin, A., & Wibowo, F. C. (2024). Development and Validation of Scientific Inquiry Literacy Instrument (SILI) Using Rasch Measurement Model. *Education Sciences*, *14*(3), 322. doi: 10.3390/educsci14030322

Davis, D. R., & Boone, W. (2021). Using Rasch analysis to evaluate the psychometric functioning of the other-directed, lighthearted, intellectual, and whimsical (OLIW) adult playfulness scale. *International Journal of Educational Research Open*, *2*, 100054. doi: 10.1016/j.ijedro.2021.100054

Diener, E. (1984). Subjective well-being. *Psychological Bulletin*, *95*(3), 542–575. doi: 10.1037/0033-2909.95.3.542

Dunstone, E. K., Reynolds, K. J., & Cárdenas, D. (2024). The role of group memberships and school identification on student well-being. *British Journal of Social Psychology*, *63*(1), 403–428. doi: 10.1111/bjso.12685

Durrotunnisa, Pali, M., Atmoko, A., & Irtadji, Muh. (2018). Student Well-Being: Unraveling School Climate and Parenting Styles Influences. *International Journal of Humanities and Applied Social Science*, *3*(7).

Faizah, F., Rahma, U., Dara, Y. P., & Gunawan, C. L. (2020). School Well-Being Siswa Sekolah Dasar dan Siswa Sekolah Menengah Pertama Pengguna Sistem Full-Day School di Indonesia. *Jurnal Kajian Bimbingan dan Konseling*, *5*(1), 34–41. doi: 10.17977/um001v5i12020p034

FitzPatrick, S., Twohig, M., & Morgan, M. (2014). Priorities for primary education? From subjects to life-skills and children's social and emotional development. *Irish Educational Studies*, *33*(3), 269–286. doi: 10.1080/03323315.2014.923183

Fraillon, J. (2004). *Measuring student well-being in the context of Australian schooling: Discussion paper*.

Govorova, E., Benítez, I., & Muñiz, J. (2020). How Schools Affect Student Well-Being: A Cross-Cultural Approach in 35 OECD Countries. *Frontiers in Psychology*, *11*, 431. doi: 10.3389/fpsyg.2020.00431

Griet Vanwynsberghe, J. V. D., Gudrun Vanlaar, & Fraine, B. D. (2017). Long-term effects of primary schools on educational positions of students 2 and 4 years after the start of secondary education. *School Effectiveness and School Improvement*, *28*(2), 167–190. doi: 10.1080/09243453.2016.1245667

Haynes, S. N., Richard, D. C. S., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment*, *7*(3), 238–247. doi: 10.1037/1040-3590.7.3.238

Hidayah, N. H., Pali, M., Ramli, M., & Hanurawan, F. (2016). Students' Well-Being Assessment at School. *Journal of Educational, Health and Community Psychology*, *5*(1), 62. doi: 10.12928/jehcp.v5i1.6257

Holzer, J., Grützmacher, L., Lüftenegger, M., Prenzel, M., & Schober, B. (2024). Shedding light on relations between teacher emotions, instructional behavior, and student school well-being – Evidence from disadvantaged schools. *Learning and Instruction*, *92*, 101926. doi: 10.1016/j.learninstruc.2024.101926

Hossain, S., O'Neill, S., & Strnadová, I. (2023). What Constitutes Student Well-Being: A Scoping Review Of Students' Perspectives. *Child Indicators Research*, *16*(2), 447–483. doi: 10.1007/s12187-022-09990-w

Isoard-Gautheur, S., Ginoux, C., Petit, R., Clavier, V., Dias, D., Sarrazin, P., & Couturier, K. (2023). Relationships between food insecurity, physical activity, detachment from studies, and students' well-being: A prospective study. *Scandinavian Journal of Medicine & Science in Sports*, *33*(7), 1242–1253. doi: 10.1111/sms.14361

Karyani, U., Prihartanti, N., & Yuwono, S. (2015). The Dimensions of Student Well-being. *Perkembangan Manusia dan Kesejahteraan Psikologis*, 413–419.

Kellock, A. (2020). Children's well-being in the primary school: A capability approach and community psychology perspective. *Childhood*, *27*(2), 220–237. doi: 10.1177/0907568220902516

Kementerian Pendidikan dan Kebudayaan. (2020). *Peraturan Direktur Jenderal Guru dan Tenaga Kependidikan Kementerian Pendidikan dan Kebudayaan (Nomor 3813/B.B1/HK/2020) tentang Perubahan Kedua Atas Peraturan Direktur Jenderal Guru dan Tenaga Kependidikan (Nomor 2607/B.B1.3/HK/2018) tentang Petunjuk Teknis Penugasan Guru Sebagai Kepala Sekolah*. Kementerian Pendidikan dan Kebudayaan.

Khatri, P., Duggal, H. K., Lim, W. M., Thomas, A., & Shiva, A. (2024). Student well-being in higher education: Scale development and validation with implications for management education. *The International Journal of Management Education*, *22*(1), 100933. doi: 10.1016/j.ijme.2024.100933

Kim, S.-H., & Cho, D. (2022). Psychometric properties of Leisure Satisfaction Scale (LSS)-short form: A Rasch rating model calibration approach. *BMC Psychology*, *10*(1), 151. doi: 10.1186/s40359-022-00861-1

Konu, A., & Rimpela, M. (2002). Well-being in schools: A conceptual model. *Health Promotion International*, *17*(1), 79–87. doi: 10.1093/heapro/17.1.79

Kurniastuti, I., & Azwar, S. (2014). Constructtion of Student Well-being Scale for 4-6th Graders. *Jurnal Psikologi*, *41*(1), 1. doi: 10.22146/jpsi.6954

Linacre, J. M. (1999). Estimation Methods for Rasch Measures. *Journal of Outcome Measurement*, *3*, 381–405.

Linacre, J. M. (2024). *A user's guide to WINSTEPS® MINISTEP: Rasch-model computer programs. Program Manual 5.7.1*. Mesa Press.

Martanti, F., & Fatkhuronji, M. (2023). *Improving Student Well-Being Aspects through Teacher's Instructional Creativity in Islamic Elementary School*. *15*(2). doi: 10.18326/mudarrisa.v15i2.383

Morgado, F. F. R., Meireles, J. F. F., Neves, C. M., Amaral, A. C. S., & Ferreira, M. E. C. (2017). Scale development: Ten main limitations and recommendations to improve future research practices. *Psicologia: Reflexão e Crítica*, *30*(3), 1–20. doi: 10.1186/s41155-016-0057-1

Morinaj, J., & Held, T. (2023). Stability and change in student well-being: A three-wave longitudinal person-centered approach. *Personality and Individual Differences*, *203*, 112015. doi: 10.1016/j.paid.2022.112015

Morshed, M. B., Kulkarni, S. S., Saha, K., Li, R., Roper, L. G., Nachman, L., Lu, H., Mirabella, L., Srivastava, S., de Barbaro, K., de Choudhury, M., Plötz, T., & Abowd, G. (2022). Food, Mood, Context: Examining College Students' Eating Context and Mental Well-being. *ACM Transactions on Computing for Healthcare*, *3*(4), 38:1-38:26. doi: 10.1145/3533390

Nguyen, A. J., McDaniel, H., Braun, S. S., Chen, L., & Bradshaw, C. P. (2021). Contextualizing the Association Between School Climate and Student Well-Being: The Moderating Role of Rurality. *Journal of School Health*, *91*(6), 463–472. doi: 10.1111/josh.13026

Parrot, L. (2002). Complexity and The Limits of Ecological Engineering. *Transactions of the ASAE*, *45*(5). doi: 10.13031/2013.11032

Pustjens, H., Van de gaer, E., Van Damme, J., Onghena, P., & Van Landeghem, G. (2007). The short-term and the long-term effect of primary schools and classes on mathematics and language achievement scores. *British Educational Research Journal*, *33*(3), 419–440. doi: 10.1080/01411920701243677

Putra, M. D. K., & Retnawati, H. (2020). Rasch Analysis of an Indonesian version of the General Self-Efficacy Scale-12: A Comparison of Rating Scale Model (RSM) and Partial Credit Model (PCM). *Indonesian Journal of Educational Assesment*, *3*(1), 52. doi: 10.26499/ijea.v3i1.55

Qudratuddarsi, H., Hidayat, R., Shah, R. L. Z. B. R. M., Nasir, N., Imami, M. K. W., & Nor, R. B. M. (2022). Rasch Validation of Instrument Measuring Gen-Z Science, Technology, Engineering, and Mathematics (STEM) Application in Teaching during the Pandemic. *International Journal of Learning, Teaching and Educational Research*, *21*(6), 104–121. doi: 10.26803/ijlter.21.6.7

Rangka, I. B., Hidayah, N., Hanurawan, F., & Eva, N. (2023). A Rasch-based validation of the Indonesian version of the food neophobia scale. *Food Quality and Preference*, *109*, 104891. doi: 10.1016/j.foodqual.2023.104891

Rodríguez, S., González-Suárez, R., Vieites, T., Piñeiro, I., & Díaz-Freire, F. M. (2022). Self-Regulation and Students Well-Being: A Systematic Review 2010–2020. *Sustainability*, *14*(4), Article 4. doi: 10.3390/su14042346

Ryff, C. D., & Keyes, C. L. M. (1995). The structure of psychological well-being revisited. *Journal of Personality and Social Psychology*, *69*(4), 719–727. doi: 10.1037/0022-3514.69.4.719

Serinci, S., Çıkrıkçı, Ö., & Sarıkaya, Y. (2023). Investigation of relationships between academic self-efficacy, life goals and subjective well-being of middle school students. *British Journal of Guidance & Counseling*, *0*(0), 1–15. doi: 10.1080/03069885.2023.2285326

Soutter, A. K., O'Steen, B., & Gilmore, A. (2014). The student well-being model: A conceptual framework for the development of student well-being indicators. *International Journal of Adolescence and Youth*, *19*(4), 496–520. doi: 10.1080/02673843.2012.754362

Svane, D., Evans, N. (Snowy), & Carter, M.-A. (2019). Wicked wellbeing: Examining the disconnect between the rhetoric and reality of wellbeing interventions in schools. *Australian Journal of Education*, *63*(2), 209–231. doi: 10.1177/0004944119843144

Tesio, L., Caronni, A., Kumbhare, D., & Scarano, S. (2024). Interpreting results from Rasch analysis 1. The "most likely" measures coming from the model. *Disability and Rehabilitation*, *46*(3), 591–603. doi: 10.1080/09638288.2023.2169771

Vucaj, I. (2022). Development and initial validation of Digital Age Teaching Scale (DATS) to assess application of ISTE Standards for Educators in K–12 education classrooms. *Journal of Research on Technology in Education*, *54*(2), 226–248. doi: 10.1080/15391523.2020.1840461

Wang, M.-T., L. Degol, J., Amemiya, J., Parr, A., & Guo, J. (2020). Classroom climate and children's academic and psychological wellbeing: A systematic review and meta-analysis. *Developmental Review*, *57*, 100912. doi: 10.1016/j.dr.2020.100912

Wibowo, Y. S., Setiawati, F. A., Qodriah, S. R., & Ayriza, Y. (2021). Positive education: Exploring students' well-being framework in Indonesia. *Jurnal Cakrawala Pendidikan*, *40*(3), Article 3. doi: 10.21831/cp.v40i3.33530

Widodo, W., & Chotimah, C. (2023). Adaptasi dan Analisis Psikometri Skala Kompetensi Multikultural Calon Guru Menggunakan Pemodelan Rasch. *Jurnal Pendidikan dan Kebudayaan*, *8*(2), 153–172. doi: 10.24832/jpnk.v8i2.4228

Yu, C. H. (2020). How Rasch Modeling Can Simplify and Enhance Your Assessment. In *Rasch Measurement: Applications in Quantitative Educational Research* (pp. 47–73). Springer Nature. doi: 10.1007/978-981-15-1800-3_4

Zhu, X., & Aryadoust, V. (2022). An investigation of mother tongue differential item functioning in a high-stakes computerized academic reading test. *Computer Assisted Language Learning*, *35*(3), 412–436. doi: 10.1080/09588221.2019.1704788