MOJES MALAYSIAN ONLINE JOURNAL OF EDUCATIONAL SCIENCES JANUARY 2026, 14 (1)

INITIAL VALIDATION OF A NEWLY DEVELOPED SCHOOL-RELATED WELL-BEING INSTRUMENT: A RASCH MODEL APPROACH

Tie-Seng Te[1]
***Hutkemri Zulnaidi[1]**
Norfaezah Md Khalid[1]
[1] Faculty of Education, Universiti Malaya, 50603 Kuala Lumpur, Malaysia
**hutkemri@um.edu.my*

ABSTRACT

As adolescents spend a substantial amount of time in school, understanding their school-related well-being is essential for supporting their overall development. To facilitate this, a valid and reliable measurement instrument is required. This study employed a quantitative research design to evaluate the psychometric properties of the newly developed Students' School-related Well-being Instrument (SSWI). Data were collected from 100 secondary school students in Wilayah Persekutuan Kuala Lumpur, Malaysia, selected through simple random sampling. Rasch analysis was conducted using Winsteps (Version 5.3.3) to assess the instrument's validity and reliability. The results indicated that the SSWI demonstrates satisfactory psychometric properties. However, some limitations were noted in content representativeness. Further refinement of the instrument is recommended to enhance its ability to comprehensively assess students' school-related well-being.

Keywords: *Instrument evaluation, Messick's validity, Rasch analysis, students' well-being, secondary school.*

INTRODUCTION

For school-aged adolescents, well-being within the school environment is crucial, as they spend a significant portion of their day at school (Borgonovi & Pál, 2016; Raccanello et al., 2021). Due to the long hours spent in school, many of their daily interactions take place there (Stang-Rabrig et al., 2023). Those day-to-day interactions and experiences will subsequently have an impact on their well-being (Seligman et al., 2009). Students exhibiting a positive level of well-being in school are more likely to have high academic achievement (Hascher & Hagenauer, 2020; Holzer et al., 2022) and less likely to create problematic behaviours in school (Hascher & Hagenauer, 2020; Putwain et al., 2019).

In view of the importance of students' well-being in school, understanding this matter is important so that the relevant actions can be taken to further enhance it. For this purpose, a well-constructed measurement instrument is needed. However, the existing instruments are not robust enough to be adapted for that purpose. Even though various instruments have been developed and used by a number of researchers to measure students' well-being, there is a scarcity of measures specifically designed to evaluate well-being indicators within the school setting (Arslan et al., 2022; Burkett-McKee et al., 2021; Dias-Viana & Noronha, 2022). Since school presents a context with its own dynamics, general well-being measures may fail to reflect the specific dimensions of students' well-being within that setting (Joing et al., 2020).

For the existing instruments that are based on school context, they are mostly developed with Western populations, where an individualistic culture is apparent (Counted et al., 2024; Hossain et al., 2019, 2023; Kern et al., 2019; Susanto et al., 2023). Those instruments may not be an accurate measurement of students' well-being in Asian countries, such as Malaysia. This is because sources of well-being are closely tied to cultural differences (Hendriks et al., 2020; Layous et al., 2013; Uchida et al., 2004). For instance, in Western cultures where individualistic values are prevalent, having more autonomy or freedom is believed to lead to higher well-being of students (Ferguson et al., 2011; Lu & Gilmour, 2004; Pflug, 2009). Conversely, in Asian countries where collectivist cultures are dominant, meeting the expectations of the authoritative figure, such as teachers, might lead to higher school satisfaction (Hossain et al., 2019).

Given the limitations of the existing instruments, the researchers have developed the Students' School-related Well-being Instrument (SSWI) to measure school-related well-being among secondary school students in Malaysia. To ensure that the newly developed instrument could produce valid and reliable scores, this study was conducted to examine its psychometric properties. Specifically, this study aimed to evaluate the content, substantive, structural, and generalisability validity of the SSWI.

LITERATURE REVIEW

Theoretical Foundations of Students' Well-being in School Context

Positive psychology is defined as the scientific study of the strengths that enable individuals and communities to thrive (Jayawickreme et al., 2012; Seligman, 2011; Seligman & Csikszentmihalyi, 2000). Rather than preoccupation with repairing the worst things, positive psychology emphasises the development of positive qualities and optimal functioning (Seligman & Csikszentmihalyi, 2000). Within this perspective, the PERMA model (Seligman, 2011) provides a comprehensive framework for understanding well-being, comprising five core elements: positive emotion, engagement, relationships, meaning, and accomplishment. Extending this framework to younger populations, the EPOCH Model of Adolescent Well-Being (Kern et al., 2016) highlights engagement, perseverance, optimism, connectedness, and happiness as key contributors to adolescents' flourishing. Together, these models integrate both hedonic and eudemonic aspects of well-being, offering a multidimensional understanding of individuals' functioning (Huppert & So, 2013; Wentzel, 2024).

In addition to positive psychology frameworks, Self-Determination Theory (SDT) (Ryan & Deci, 2000) provides a complementary lens for understanding individuals' well-being. This theory posits that well-being is enhanced when three basic psychological needs, namely autonomy, competence, and relatedness, are fulfilled. Within the school context, these needs are reflected in students' sense of agency in learning (autonomy), perceived academic capability (competence), and quality of social relationships (relatedness). These dimensions align closely with key components identified in both the PERMA and EPOCH models, reinforcing their relevance in conceptualising students' well-being.

Drawing on these theoretical foundations, students' well-being can be understood as a multidimensional construct encompassing both emotional experiences and functional aspects within the school environment. The hedonic component refers to students' affective experiences, particularly the presence of positive emotions such as happiness, enjoyment, excitement, and comfort in school (Cavicchiolo et al., 2025; Hascher, 2007; Long et al., 2012; Putwain et al., 2021; Renshaw et al., 2015). The eudemonic component, in contrast, reflects students' effective functioning in school-related domains. A key aspect is social functioning, which includes social interactions or competence in interacting with others at school (Gorodzinsky et al., 2011). This dimension corresponds to the "relationships" component in PERMA (Seligman, 2011), "connectedness" in EPOCH (Kern et al., 2016), and "relatedness" in SDT (Ryan & Deci, 2000), highlighting its central role in students' well-being. In view of its significance, this aspect is included in many instruments pertaining to students' well-being (Conesa & Duñabeitia, 2021; Niclasen et al., 2018; Renshaw et al., 2015; Van Landeghem et al., 2002).

Beyond social functioning, school-related well-being also encompasses other school-related functioning elements, such as engagement, perseverance, and self-efficacy (Holzer et al., 2021; Hossain et al., 2023). Engagement refers to active involvement and interest in learning activities, while perseverance reflects sustained effort in the face of academic challenges. Self-efficacy represents a student's perceived ability to successfully perform school tasks. Collectively, these elements capture the extent to which students are functioning effectively and "doing well" in school (Huppert & So, 2013; Jayawickreme et al., 2012).

Based on the integration of the PERMA framework, the EPOCH model, and SDT, the present study conceptualises students' well-being as a multidimensional construct comprising positive emotional experiences, social functioning, and other school-related functioning. This theoretical grounding provides a clear foundation for the development and validation of the school-related well-being instrument and supports the interpretation of its construct validity.

Validation of Measurement Instrument Employing Rasch Analysis

Validation of the measurement instrument is important to ensure the scores obtained are valid and reliable (DeVellis & Thorpe, 2022). The Rasch model, which is based on Item Response Theory (IRT), offers a robust method for evaluating the psychometric properties of measurement instruments (Bond et al., 2021; Conrad et al., 2015; Khine, 2020; Sovey et al., 2022). Data from Rasch analysis can be used to establish various types of validity evidence within the Messick Validity Framework, including content validity, substantive validity, structural validity, and generalisability validity (Messick, 1989, 1995).

Content Validity

Content validity pertains to the relevance, representativeness, and technical quality of instrument items (Messick, 1989, 1995). Content relevance focuses on the alignment between items and the construct being measured (Ha, 2021), typically ensured during the development phase and via expert review (Blouin & Smith, 2020; Messick, 1995).

Representativeness evaluates whether the instrument adequately covers the construct's domain (Beglar, 2010; Messick, 1989, 1995). The Wright map is instrumental in assessing this, as it visualises the distribution of items across the construct continuum and helps identify gaps or redundancies (Makransky et al., 2015; Noroozi & Karami, 2024; Phipps, 2023). If the Wright map reveals sparsely populated regions, it suggests the need to develop additional items to improve coverage (Akour, 2022; Noroozi & Karami, 2024). Conversely, items aligned horizontally indicate redundancy and may be removed to enhance parsimony (Akour, 2022; Avinç & Doğan, 2024; Phipps, 2023). The item separation index further supports representativeness, with high values indicating the inclusion of items spanning a wide range of difficulty (Ha, 2021; Noroozi & Karami, 2024). The inclusion of sufficient items with varied difficulty levels is desirable as it contributes to a more comprehensive evaluation (Avinç & Doğan, 2024; Noroozi & Karami, 2024). In the context of psychological well-being, having sufficient items to differentiate well the respondents with low well-being is particularly important because those with low well-being are at risk for various problems and require necessary interventions (Gong et al., 2021).

The technical quality of items can be examined using item fit statistics (Beglar, 2010; Cheng et al., 2011; Fendrich et al., 2009; Greene et al., 2017; Ha, 2021; Makransky et al., 2015; Noroozi & Karami, 2024; Phipps, 2023). Items are considered to exhibit optimal fit if their mean square (MNSQ) values range between 0.60 and 1.40 (Wright & Linacre, 1994). Items with MNSQ values between 1.40 and 2.00 may still be acceptable, although less productive (Wright & Linacre, 1994). Furthermore, item fit can be assessed by examining item polarity, typically through the calculation of the Point-Measure Correlation (PTMEACorr) value (Fendrich et al., 2009; Noroozi & Karami, 2024; Sovey et al., 2022). Positive values suggest consistency with the intended construct, while zero or negative values indicate problematic items that may be inconsistent with the construct measured (Bond et al., 2021).

Substantive Validity

Substantive validity refers to the theoretical justification for observed response patterns (Messick, 1995). It focuses on the logical ordering of items and whether the item hierarchy aligns with theoretical expectations (Blouin & Smith, 2020; Cheng et al., 2011; Fendrich et al., 2009). Key sources of evidence include person fit statistics, the Wright map, and rating scale structure analysis.

Person fit statistics reveal whether respondents' response patterns conform to Rasch model expectations. A low number of misfitting respondents supports substantive validity, as it suggests consistent interpretation of item hierarchy (Fendrich et al., 2009; Noroozi & Karami, 2024; Phipps, 2023). Misfits may result from carelessness, guessing, item bias, or specialised knowledge, and are commonly identified using an infit MNSQ value greater than 2.00 (Phipps, 2023; Wolfe & Smith Jr., 2007).

The Wright map can also indicate substantive validity by confirming whether the item ordering matches theoretical expectations (Blouin & Smith, 2020; Cheng et al., 2011; Fendrich et al., 2009; Phipps, 2023). For constructs like school-related well-being, items may not follow a strict hierarchy, which is consistent with theory suggesting the components are interrelated rather than sequential (Buerger et al., 2023; Dias-Viana & Noronha, 2022; Furrer & Skinner, 2003; Stipek, 2002).

Rating scale structure analysis further assesses whether response categories behave as intended (Blouin & Smith, 2020; Makransky et al., 2015; Phipps, 2023). Several guidelines provided by Linacre (2002) were used to examine the functionality of the rating scales, which include: (i) each category should have at least 10 observations; (ii) average measures advance monotonically with category; and (iii) Outfit MNSQ values should be below 2.00. The rating probability curves also provide useful input, as each response option is expected to display a clear peak, suggesting that each is the most likely choice within a certain segment of the underlying trait (Bond et al., 2021).

Structural Validity

Structural validity is concerned with the hypothesised dimensionality of the data (Blouin & Smith, 2020; Fendrich et al., 2009; Ha, 2021; Phipps, 2023). In examining the dimensionality of a measurement instrument, the Principal Component Analysis (PCA) of the Rasch residuals helps in checking for violations of unidimensionality (Bond et al., 2021). There were several criteria to assess unidimensionality, which include: (i) the measuring dimension must have a substantial variance explained, ideally above 40 percent; (ii) the variance attributed to the first residual principal component should remain low, ideally below 15 percent; (iii) the ratio of explained Rasch variance to the first residual contrast should be at least 3:1; and (iv) the eigenvalue of the first contrast should be below 2.00 (Avinç & Doğan, 2024; Conrad et al., 2015; Linacre, 2016; Reckase, 1979). Additionally, good item fit statistics also support unidimensionality (Conrad et al., 2015). However, even when a dominant dimension exists, subdimensions may still emerge, indicating that item clusters may reflect specific facets of the construct (Akour, 2022; Lu & Yeo, 2015; Peng et al., 2025).

Generalisability Validity

Generalisability validity reflects the degree to which score interpretations remain consistent across various groups or settings (Blouin & Smith, 2020; Messick, 1995; Wasserman & Bracken, 2003). The evidence of generalisability validity can be obtained from differential item functioning (DIF) and internal consistency analysis (Cheng et al., 2011; Chong et al., 2022; Fendrich et al., 2009; Ha, 2021; Noroozi & Karami, 2024; Royal & Royal, 2017; Son & Ha, 2025). DIF occurs when distinct subgroups within the sample exhibit dissimilar scores on particular items, despite equal levels of latent trait (Makransky et al., 2015; Tennant & Conaghan, 2007). In DIF analysis, a DIF contrast of >0.64 and $p<0.05$ is used as an indicator of DIF (Zwick et al., 1999).

In addition, both person and item reliability indices also provide evidence for generalisability validity (Fendrich et al., 2009; Ha, 2021). These reliability indices range from 0 to 1 and are conceptually similar to Cronbach's alpha (Boone & Noltemeyer, 2017). For new scales, reliability coefficients above 0.70 are generally considered acceptable (Nunnally, 1978). Rasch also provides person and item separation indices to evaluate score precision (Bond et al., 2021; Boone et al., 2014). The person separation index evaluates how effectively a set of items differentiates among individuals, while the item separation index reflects how well the respondent sample distinguishes between the items (Boone & Noltemeyer, 2017). Higher values indicate greater generalisability. Specifically, for person separation, a value of 1.50 is acceptable, 2.00 is considered good, and 3.00 indicates excellent separation (Wright & Masters, 1982). For item separation, an index of 3.00 or greater is preferable (Linacre, 2016).

METHODOLOGY

Research Design
This initial validation study employed a quantitative research design to evaluate the psychometric properties of the SSWI using the Rasch measurement model.

Samples
The study involved 100 secondary school students aged 13 to 16 years old from Wilayah Persekutuan Kuala Lumpur, Malaysia. This age range was selected as it represents a developmental period characterised by heightened vulnerability to psychological and mental health challenges (Chin & Wu, 2020; Mohamed et al., 2026; Sabramani et al., 2021; Sahril et al., 2025). This sample size aligns with Linacre's (1994) guideline, which suggests that at least 100 participants are needed to ensure a 95% confidence level with a measurement precision of ±0.5 logit. However, it is important to note that this justification pertains primarily to the stability of item parameter estimates rather than to the generalisability validity of the findings. To ensure equal probability of selection, participants were selected using simple random sampling.

Instrument
The SSWI was developed to assess school-related well-being among Malaysian secondary school students. The development followed an exploratory sequential design (Creswell & Clark, 2018), beginning with qualitative data collection. Items were then generated based on findings from the initial qualitative phase. Subject matter experts in educational psychology, measurement and evaluation, and practising school counsellors reviewed the initial draft of the instrument. Based on their feedback, the instrument was refined. The pilot version comprised 36 items, each rated on a 5-point Likert scale, ranging from 1 "strongly disagree" to 5 "strongly agree".

Procedure
Before conducting the study, the researcher sought permission to conduct the research from the University of Malaya Research Ethics Committee (UMREC) (Reference number: UM.TNC2/UMREC_2598). In addition, the researcher obtained permission to conduct the study from the Ministry of Education Malaysia (MoE), the State Education Department of Wilayah Persekutuan Kuala Lumpur, and the respective school administrators. As this study involved secondary school students below 18 years old, the researcher obtained their parents' informed consents before administering the questionnaires. This step was in line with Standard 3.10(b) of the "Ethical Principles of Psychologists and Code of Conduct" (American Psychological Association, 2017), where, for persons who are legally incapable of giving informed consent, permission from a legally authorised person has to be obtained.

Data Analysis
To analyse the data, Rasch analysis was performed using Winsteps version 5.3.3 software. The analyses included examining dimensionality, item fit, person fit, the Wright map, DIF, response categories, and reliability.

RESULTS

Demographic Information of the Respondents
This study involved 100 secondary school students in Wilayah Persekutuan Kuala Lumpur, Malaysia. The respondents' demographic information is as illustrated in Table 1.

Table 1. *Demographic Information of Respondents*

	Frequency (*n*)	Percentage (%)
Gender		
Male	48	48.00
Female	52	52.00
Form		
1 (13 years old)	25	25.00

2 (14 years old)	25	25.00
3 (15 years old)	25	25.00
4 (16 years old)	25	25.00

As shown in Table 1, out of the 100 respondents, 48 (48%) were male, and 52 (52%) were female. In terms of form, there were 25 respondents (25%) from each form (Form 1 – Form 4).

Rasch Analysis

Dimensionality. The PCA was performed to examine dimensionality. Table 2 shows the PCA results for the SSWI.

Table 2. *Principal Component Analysis*

Standardised Residual Variance	Eigenvalue	Observed	Modelled
Total raw variance in observations	77.32	100%	100%
Raw variance explained by measures	41.32	53.4%	53.2%
Unexplained variance in 1st contrast	4.76	6.2%	

As presented in Table 2, the raw variance explained by the measures was 53.4%, satisfying one of the key unidimensionality criteria, which is that at least 40% of the variance should be accounted for by the measurement dimension (Avinç & Doğan, 2024; Conrad et al., 2015; Reckase, 1979). Secondly, the unexplained variance in the first contrast was 6.2%, which falls within the acceptable threshold of below 15% (Avinç & Doğan, 2024; Conrad et al., 2015). Additionally, the ratio between the explained measurement variance (53.4%) and the variance of the first residual contrast (6.2%) was approximately 8:1, exceeding the minimum recommended ratio of 3:1 (Conrad et al., 2015). However, the eigenvalue for the first contrast in the unexplained variance was 4.76, surpassing the commonly accepted threshold of 2.00 (Linacre, 2016).

Item Fit. Item fit was examined to determine how well the instrument items fit with the Rasch measurement model. Table 3 displays the fit statistics of the instrument items.

Table 3. *Item Fit Statistics*

Item	Measure	S.E.	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	PTMEA Corr
SW1	-0.17	0.16	1.08	0.60	1.09	0.62	0.63
SW2	-1.74	0.19	1.10	0.69	0.97	-0.01	0.57
SW3	0.12	0.16	0.98	-0.10	0.99	0.02	0.65
SW4	-0.47	0.17	0.85	-0.99	0.78	-1.31	0.72
SW5	0.02	0.16	1.04	0.31	1.04	0.32	0.66
SW6	-0.01	0.16	0.99	0.00	1.10	0.69	0.64
SW7	0.48	0.16	1.53	3.17	1.46	2.74	0.69
SW8	-0.44	0.17	1.67	3.80	1.66	3.33	0.58
SW9	-0.03	0.16	0.96	-0.26	0.93	-0.40	0.70
SW10	0.52	0.16	0.88	-0.80	0.90	-0.61	0.73
SW11	-0.17	0.16	1.21	1.40	1.32	1.86	0.56
SW12	-0.01	0.16	0.77	-1.62	0.71	-2.05	0.76
SW13	-1.26	0.18	0.97	-0.13	0.81	-0.85	0.66
SW14	-0.17	0.16	0.96	-0.25	0.96	-0.21	0.67
SW15	0.28	0.16	0.81	-1.32	0.77	-1.64	0.73
SW16	0.3	0.16	0.98	-0.10	0.89	-0.68	0.75
SW17	0.25	0.16	0.76	-1.74	0.87	-0.85	0.74
SW18	0.85	0.15	1.01	0.13	1.03	0.26	0.76
SW19	-0.44	0.17	1.22	1.41	1.26	1.46	0.67
SW20	0.4	0.16	0.81	-1.32	0.78	-1.53	0.79

Item	Measure	S.E.	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	PTMEA Corr
SW21	-0.53	0.17	1.00	0.05	0.95	-0.25	0.67
SW22	0.55	0.16	0.98	-0.10	0.97	-0.14	0.70
SW23	0.71	0.15	0.91	-0.58	0.89	-0.70	0.76
SW24	0.23	0.16	0.80	-1.36	0.80	-1.39	0.73
SW25	0.48	0.16	1.24	1.58	1.25	1.61	0.62
SW26	0.3	0.16	0.79	-1.51	0.81	-1.33	0.72
SW27	0.25	0.16	0.73	-1.99	0.71	-2.09	0.75
SW28	0.43	0.16	0.77	-1.62	0.75	-1.80	0.77
SW29	-0.41	0.17	1.28	1.80	1.25	1.44	0.61
SW30	-0.11	0.16	0.76	-1.73	0.75	-1.70	0.73
SW31	0.76	0.15	1.25	1.66	1.27	1.76	0.75
SW32	0.52	0.16	0.83	-1.18	0.81	-1.29	0.73
SW33	-0.3	0.17	0.89	-0.71	0.98	-0.06	0.69
SW34	0.28	0.16	1.01	0.11	0.97	-0.15	0.74
SW35	0.07	0.16	0.90	-0.67	0.85	-0.96	0.76
SW36	-1.56	0.19	1.14	0.96	0.97	-0.06	0.60

As shown in Table 3, all the items have infit and outfit MNSQ values which are within 0.60 – 1.40, except item SW7 (Infit MNSQ: 1.53; Outfit MNSQ: 1.46) and item SW8 (Infit MNSQ: 1.67; Outfit MNSQ: 1.66), which slightly exceeded the recommended upper bound for fit statistics. However, the misfit values remained below 2.00, suggesting that they do not substantially degrade the measurement system (Wright & Linacre, 1994). Moreover, all the items showed positive PTMEACorr values, indicating that they are positively aligned with the underlying construct.

Person Fit. Person fit was examined to identify individual response patterns that might have contributed to item misfit (Bond et al., 2021). Table 4 illustrates the person fit statistics.

Table 4. *Person Fit Statistics*

Person Code	Measure	S.E.	Infit MNSQ	Outfit MNSQ	Person Code	Measure	S.E.	Infit MNSQ	Outfit MNSQ
1	-0.16	0.24	1.49	1.49	**51**	-0.11	0.24	0.87	0.87
2	0.75	0.25	0.76	0.78	**52**	4.12	0.35	0.78	0.71
3	0.31	0.25	0.31	0.31	**53**	3.89	0.33	1.23	1.32
4	1.4	0.26	1.57	1.60	**54**	0.94	0.25	1.18	1.16
5	1.2	0.26	1.34	1.33	**55**	1.4	0.26	1.01	1.01
6	-0.28	0.24	1.16	1.19	**56**	-0.11	0.24	0.94	0.95
7	1.07	0.25	1.09	1.09	**57**	-0.76	0.23	1.29	1.30
8	0.01	0.24	0.67	0.67	**58**	2.08	0.26	0.92	0.92
9	2.01	0.26	1.18	1.18	**59**	0.13	0.25	1.34	1.36
10	3.68	0.32	1.09	1.06	**60**	5.34	0.53	1.00	1.11
11	3.68	0.32	0.86	0.84	**61**	3.3	0.3	1.31	1.25
12	2.58	0.27	0.63	0.62	**62**	3.13	0.29	1.10	1.11
13	4	0.34	0.89	0.88	**63**	0.44	0.25	0.80	0.82
14	1.07	0.25	0.74	0.75	**64**	0.13	0.25	1.45	1.39
15	5.09	0.48	0.96	0.92	**65**	-1.84	0.21	1.13	1.14
16	3.48	0.3	0.71	0.70	**66**	4.53	0.39	1.01	1.02
17	2.29	0.27	1.51	1.51	**67**	0.81	0.25	1.03	1.03
18	2.29	0.27	0.23	0.24	**68**	3.13	0.29	1.18	1.16
19	3.58	0.31	0.81	0.78	**69**	-0.5	0.23	0.98	1.00
20	1.2	0.26	0.90	0.89	**70**	1.01	0.25	0.82	0.82
21	-0.44	0.23	1.07	1.08	**71**	0.75	0.25	0.89	0.89
22	0.25	0.25	0.71	0.71	**72**	0.75	0.25	0.95	0.95
23	2.01	0.26	1.52	1.49	**73**	-0.11	0.24	1.12	1.12

Person Code	Measure	S.E.	Infit MNSQ	Outfit MNSQ	Person Code	Measure	S.E.	Infit MNSQ	Outfit MNSQ
24	0.56	0.25	1.31	1.29	**74**	3.22	0.29	0.80	0.77
25	2.43	0.27	0.77	0.81	**75**	2.15	0.27	0.68	0.68
26	1.14	0.26	0.98	0.97	**76**	1.33	0.26	1.36	1.34
27	1.27	0.26	1.00	1.01	**77**	1.6	0.26	0.86	0.87
28	0.31	0.25	1.20	1.19	**78**	-0.71	0.23	0.79	0.79
29	-1	0.22	0.99	1.03	**79**	-0.22	0.24	1.61	1.53
30	2.22	0.27	1.03	1.03	**80**	1.01	0.25	0.75	0.76
31	2.01	0.26	0.73	0.73	**81**	1.07	0.25	0.72	0.73
32	0.62	0.25	0.76	0.76	**82**	0.44	0.25	1.05	1.05
33	3.05	0.29	0.86	0.91	**83**	0.81	0.25	1.01	1.03
34	-0.55	0.23	0.98	1.00	**84**	0.44	0.25	1.44	1.40
35	4.69	0.41	0.84	0.72	**85**	-0.28	0.24	0.84	0.85
36	2.01	0.26	0.67	0.66	**86**	-0.16	0.24	1.19	1.26
37	2.97	0.28	1.10	1.15	**87**	2.01	0.26	1.08	1.08
38	5.65	0.6	0.97	0.96	**88**	6.09	0.72	0.95	0.70
39	2.22	0.27	0.73	0.74	**89**	2.43	0.27	1.26	1.25
40	1.87	0.26	0.65	0.65	**90**	8.03	1.83	1.00	1.00
41	1.4	0.26	0.82	0.82	**91**	1.8	0.26	1.07	1.07
42	1.01	0.25	0.78	0.81	**92**	3.48	0.3	0.69	0.68
43	3.39	0.3	1.22	1.21	**93**	4.12	0.35	0.81	0.75
44	3.48	0.3	0.87	1.10	**94**	2.58	0.27	1.20	1.17
45	5.09	0.48	0.93	0.89	**95**	2.01	0.26	1.32	1.30
46	1.33	0.26	1.02	1.02	**96**	4.12	0.35	0.94	0.89
47	1.94	0.26	0.96	0.95	**97**	4.87	0.44	0.99	0.99
48	3.13	0.29	0.74	0.70	**98**	0.81	0.25	1.23	1.23
49	2.29	0.27	0.98	0.98	**99**	1.14	0.26	0.80	0.79
50	2.51	0.27	0.77	0.75	**100**	2.08	0.26	0.76	0.75

Based on the analysis of person fit, four respondents were reported to have MNSQ more than 1.50, and two respondents had MNSQ less than 0.50. However, none of the respondents had MNSQ greater than 2.00, indicating that their responses did not distort the measurement (Wright & Linacre, 1994).

The Wright Maps. The Wright Map offers a clear visual representation of the relationship between persons and items using an equal-interval logit scale (Boone & Noltemeyer, 2017). Figure 1 displays the Wright map.

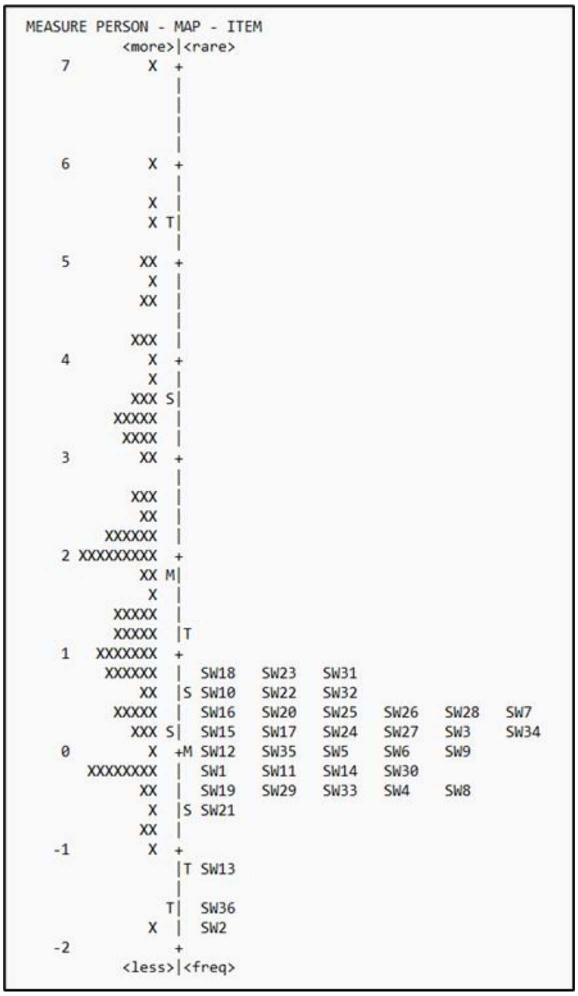

Figure 1. *The Wright Map*

As displayed in Figure 1, the mean endorsement score among respondents was greater than the mean item difficulty, implying that the items were generally easy for respondents to endorse. In addition, the Wright map reveals several instances of item overlap.

Differential Item Functioning (DIF). Differential item functioning (DIF) analysis was conducted to assess whether items function equivalently across gender groups. Table 5 presents the DIF results.

Table 5. *DIF Analysis for Gender*

Person	Obs-Exp Average	Person	Obs-Exp Average	DIF Contrast	Joint S.E.	Rasch-Welch			Item
						t	*d.f.*	*p*	
Male	0.01	Female	0.00	0.00	0.33	0.00	95	1.00	SW1
Male	0.05	Female	-0.04	-0.34	0.39	-0.87	94	0.39	SW2
Male	0.12	Female	-0.10	-0.56	0.32	-1.73	94	0.09	SW3
Male	-0.04	Female	0.04	0.22	0.34	0.66	96	0.51	SW4
Male	-0.07	Female	0.07	0.37	0.32	1.13	96	0.26	SW5
Male	-0.02	Female	0.02	0.10	0.32	0.32	95	0.75	SW6
Male	0.06	Female	-0.06	-0.29	0.31	-0.92	95	0.36	SW7
Male	-0.01	Female	0.01	0.06	0.34	0.17	95	0.86	SW8
Male	0.12	Female	-0.11	-0.59	0.33	-1.81	94	0.07	SW9
Male	-0.02	Female	0.02	0.11	0.31	0.36	95	0.72	SW10
Male	-0.06	Female	0.05	0.30	0.33	0.91	96	0.36	SW11
Male	0.07	Female	-0.06	-0.32	0.33	-0.99	95	0.33	SW12

Person	Obs-Exp Average	Person	Obs-Exp Average	DIF Contrast	Joint S.E.	Rasch-Welch			Item
						t	*d.f.*	*p*	
Male	-0.09	Female	0.08	0.55	0.36	1.52	96	0.13	SW13
Male	-0.04	Female	0.03	0.19	0.33	0.59	96	0.56	SW14
Male	-0.04	Female	0.04	0.18	0.32	0.58	96	0.56	SW15
Male	0.01	Female	-0.01	-0.07	0.32	-0.22	95	0.82	SW16
Male	0.08	Female	-0.07	-0.38	0.32	-1.18	95	0.24	SW17
Male	-0.03	Female	0.03	0.15	0.30	0.50	95	0.62	SW18
Male	-0.12	Female	0.11	0.62	0.34	1.84	96	0.07	SW19
Male	0.08	Female	-0.07	-0.35	0.32	-1.11	95	0.27	SW20
Male	-0.17	Female	0.15	0.91	0.34	2.67	96	0.01	SW21
Male	0.03	Female	-0.03	-0.13	0.31	-0.42	95	0.67	SW22
Male	0.06	Female	-0.05	-0.25	0.31	-0.80	95	0.42	SW23
Male	-0.02	Female	0.01	0.08	0.32	0.25	95	0.80	SW24
Male	-0.15	Female	0.14	0.69	0.31	2.21	96	0.03	SW25
Male	-0.01	Female	0.01	0.00	0.32	0.00	95	1.00	SW26
Male	0.02	Female	-0.01	-0.08	0.32	-0.24	95	0.81	SW27
Male	-0.02	Female	0.02	0.10	0.31	0.32	95	0.75	SW28
Male	0.09	Female	-0.08	-0.45	0.34	-1.34	94	0.18	SW29
Male	-0.02	Female	0.02	0.09	0.33	0.28	95	0.78	SW30
Male	-0.07	Female	0.07	0.33	0.31	1.07	96	0.29	SW31
Male	0.10	Female	-0.09	-0.48	0.31	-1.52	94	0.13	SW32
Male	0.06	Female	-0.05	-0.32	0.33	-0.95	95	0.34	SW33
Male	0.00	Female	0.00	0.00	0.32	0.00	95	1.00	SW34
Male	0.05	Female	-0.05	-0.26	0.32	-0.79	95	0.43	SW35
Male	0.01	Female	-0.01	-0.09	0.38	-0.23	95	0.82	SW36

As illustrated in Table 5, two items displayed DIF by gender, which were item SW21 (DIF contrast 0.91) and item SW25 (DIF contrast 0.69).

Reliability. Reliability and separation indices indicate the extent to which the scores can be consistently reproduced. The reliability and separation statistics are demonstrated in Table 6.

Table 6. *Reliability and Separation Statistics*

Parameter	Separation	Reliability
Person (100)	4.89	0.96
Item (36)	3.31	0.92

As illustrated in Table 6, person and item reliability indices from the Rasch Model analysis were 0.96 and 0.92, respectively. Both the person and item reliability were in the 0.90 range, suggesting that the item ordering and person measures would likely remain highly consistent if the SSWI were administered to similar samples (Bond et al., 2021). The person separation index was 4.89, suggesting that the instrument can reliably distinguish respondents into approximately five distinct groups. The item separation index of 3.31 indicates that the sample is sufficient to differentiate items into about three distinct difficulty levels.

Rating Scale Effectiveness. The effectiveness of the 5-point rating scale was analysed. Table 7 shows the rating scale effectiveness of the SSWI.

Table 7. *Rating Scale Effectiveness*

Category Label	Observed Count	Average Measure	Infit MNSQ	Outfit MNSQ	Threshold Calibration
1	36	-1.15	0.95	0.94	None
2	188	-0.42	0.90	0.88	-2.35

3	997	0.55	0.95	0.92	-1.58
4	1411	1.88	1.04	1.02	0.81
5	968	3.60	1.06	1.05	3.12

As demonstrated in Table 7, there were at least 10 observations for each response category. The average measures consistently increased across the categories. In terms of category fit, the outfit MNSQ for each category was less than 2.00. Those values met the guidelines for rating scale analysis as recommended by Linacre (2002). In addition, the response category probability curves were examined. Figure 2 illustrates the response category probability curves.

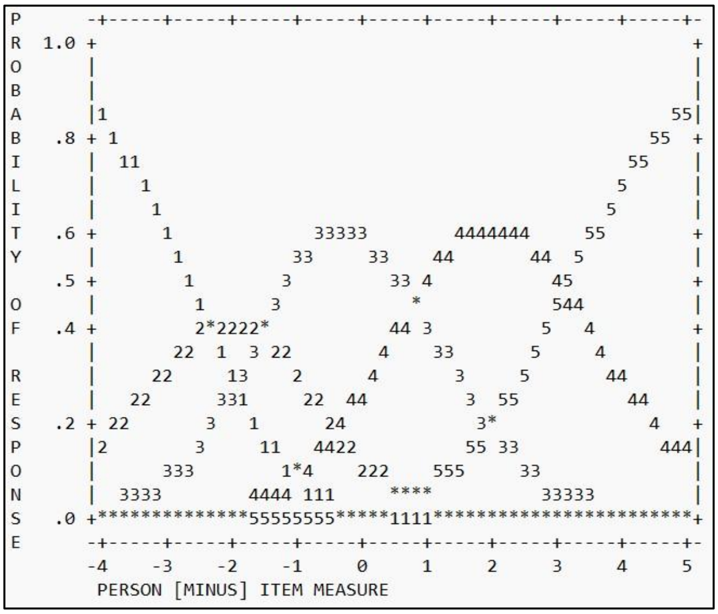

Figure 2. *Response Category Probability Curves*

As depicted in Figure 2, none of the categories was overshadowed by other categories. Based on the indicators explained above, the response category structure was acceptable.

DISCUSSION

This study aimed to examine the psychometric properties of the newly developed Students' School-related Well-being Instrument (SSWI). Through Rasch modelling, the instrument was assessed for its construct validity, focusing on content, substantive, structural, and generalisability components.

Content validity encompasses content relevance, representativeness, and technical quality (Messick, 1989, 1995). Regarding item quality, fit statistics revealed that 34 out of 36 items exhibited optimal fit, with MNSQ values ranging between 0.60 and 1.40 (Wright & Linacre, 1994). Two items (SW7 and SW8) showed misfit, indicating some degree of unexpected response variability. However, these values remained below 2.00, suggesting that they do not substantially degrade the measurement system (Wright & Linacre, 1994). Importantly, the two items capture students' functioning in school extracurricular activities and their relationship with schoolmates, which represent theoretically meaningful components of school well-being. Given that the misfit is moderate rather than severe, and the items remain conceptually relevant and positively correlated with the latent trait, they were retained in this initial validation. In addition, all items have positive PTMEACorr values, suggesting that the items were well aligned with the construct being measured (Bond et al., 2021). Collectively, these results support the technical quality and appropriateness of the SSWI items. Nevertheless, the misfitting items warrant further investigation in future studies, including potential rewording or refinement to enhance their measurement precision.

In terms of representativeness, the item separation index (3.31) was more than 3.00, indicating that a sufficient difficulty range of items has been included in the instrument (Linacre, 2016). However, inspection of the Wright map showed that the most difficult items failed to cover the higher end of the latent trait. Specifically, there was a notable gap between students with high agreeability and the most challenging items, suggesting insufficient differentiation among respondents with high levels of school-related well-being. While this is not a major flaw, given that distinguishing students with low well-being is more critical for early intervention (Gong et al., 2021), the lack of more challenging items constrains its ability to differentiate among higher well-being students. Therefore, future research should consider developing additional items targeting higher levels of the construct to improve scale targeting and enhance discrimination across the full spectrum of school-related well-being.

Furthermore, some items are positioned at similar logit locations, suggesting overlapping difficulty levels (Akour, 2022; Avinç & Doğan, 2024; Phipps, 2023). However, it is important to note that those items may capture distinct facets of school well-being across different subconstructs, such as affective, social, and other school-related functioning dimensions. Given that this study represents an initial validation of the instrument, retaining these items allows for a more comprehensive coverage of the construct.

Substantive validity of the SSWI was supported by person fit statistics. None of the respondents exhibited misfitting responses (MNSQ > 2.00), indicating that the item hierarchy was interpreted consistently across participants (Fendrich et al., 2009; Wolfe & Smith Jr., 2007). Additionally, the Wright map revealed no clear hierarchy among the affective, social, and other school-related functioning elements. Instead, items from these domains were dispersed across the continuum. This pattern aligns with theoretical perspectives that view the dimensions of students' well-being as interrelated rather than hierarchical (Buerger et al., 2023; Dias-Viana & Noronha, 2022; Furrer & Skinner, 2003; Stipek, 2002). Moreover, the examination on the 5-category rating scales showed that the rating scales function effectively, in accordance with the recommendation on rating scale evaluation by Linacre (2002). These findings support that the instrument's response structure reflects the underlying trait appropriately. However, although the categories met Linacre's criteria and no disordered thresholds or overshadowed categories were observed, the clustering of responses at higher levels suggests a potential ceiling tendency and may indicate limited differentiation between adjacent upper categories. Therefore, while the current five-category structure is retained for this initial validation due to its acceptable psychometric performance, future research may consider refining the upper categories, such as collapsing or rewording categories 4 and 5, to enhance discrimination among respondents with high levels of school-related well-being.

Next, the structural validity of the scale was examined using PCA of residuals. The results indicate that a dominant measurement dimension characterised the 36-item instrument. However, the eigenvalue of the first contrast (4.76) exceeded the suggested cutoff of 2.00, indicating the presence of secondary dimensions (Linacre, 2016). This suggests that, while a dominant latent trait underlies the instrument, the assumption of strict unidimensionality may not be fully met. Rather, the findings point to the coexistence of a general school-related well-being factor alongside potential subdimensions. Such a pattern is consistent with findings in Rasch measurement, where a strong general factor can coexist with secondary dimensions (Akour, 2022; Lu & Yeo, 2015; Peng et al., 2025). Importantly, this observation aligns with the theoretical conceptualisation of students' well-being as a multidimensional construct encompassing emotional, social, and other functional domains (Kern et al., 2016; Ryan & Deci, 2000; Seligman, 2011). Taken together, these findings support the use of the SSWI as a composite measure of students' school-related well-being, while acknowledging its multidimensional nature. Future research is recommended to further explore this structure using methods such as exploratory factor analysis (EFA) to clarify the interpretability and substantive meaning of these secondary dimensions.

Evidence supporting the generalisability of the SSWI was observed in the present study, as indicated by the DIF and reliability analyses (Cheng et al., 2011; Chong et al., 2022; Fendrich et al., 2009; Ha, 2021; Noroozi & Karami, 2024; Royal & Royal, 2017; Son & Ha, 2025). In terms of measurement invariance, DIF analysis by gender suggests that the majority of items functioned equivalently across groups. Only two items (SW21 and SW25) exhibited moderate DIF, indicating that responses to these items may differ slightly between gender groups despite similar levels of the underlying trait. Next, the item and person reliability indices were above 0.90, indicating strong internal consistency and suggesting that the scale would perform reliably across similar samples (Blouin & Smith, 2020; Cheng et al., 2011; Fendrich et al., 2009; Greene et al., 2017; Ha, 2021). The high item and person separation indices further supported this conclusion, as higher separation values reflect the instrument's ability to distinguish among different levels of students' school well-being. However, these findings should be interpreted with caution. The current evidence is based on data collected from a single geographical location, which may limit the broader generalisability of the instrument. As such, the results provide preliminary support for the SSWI's potential applicability rather than definitive evidence of generalisability across diverse populations.

IMPLICATIONS OF THE STUDY

The Rasch analysis employed in this study provides robust initial validity evidence for the newly developed SSWI. Overall, the findings suggest that the SSWI demonstrates satisfactory psychometric properties. However, enhancements are still needed, particularly concerning content representativeness. Practitioners and researchers using the SSWI are encouraged to stay informed about future revisions or refinements to the instrument, as ongoing validation efforts may further improve its utility and accuracy in assessing students' school-related well-being.

LIMITATIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

This study is not without limitations. First, the sample was limited to students from urban schools in Wilayah Persekutuan Kuala Lumpur. This restricts the generalisability of the findings and limits the ability to examine DIF across additional factors such as grade levels or other demographic subgroups. The limited sample size also constrains the interpretation of item difficulty and rating scale category functioning, particularly the tendency for responses to cluster at higher categories. Moreover, the sample size was not sufficient to conduct additional analyses, such as exploratory factor analysis (EFA) to further examine the instrument's structural validity. Future studies should include larger and more geographically and demographically diverse samples to enhance generalisability and allow more comprehensive psychometric evaluations.

Second, the external validity of the SSWI has not yet been established. The instrument was not evaluated against other theoretically related measures. Future research should assess convergent and discriminant validity by correlating the SSWI with established well-being instruments to further confirm its construct validity.

CONCLUSION

In conclusion, this study provides preliminary support for the psychometric soundness of the SSWI. Rasch analysis findings affirm the substantive, structural, and generalisability validity of the scale. However, improvements are warranted in the area of content representativeness, particularly to better capture the higher end of the school-related well-being continuum. Continued refinement and validation across more diverse populations will further strengthen the instrument's effectiveness and applicability in educational and psychological research contexts.

REFERENCES

Akour, M. M. (2022). Rasch rating scale analysis of the survey of attitudes toward statistics. *EURASIA Journal of Mathematics, Science and Technology Education*, *18*(12), em2190. https://doi.org/10.29333/ejmste/12646

American Psychological Association. (2017). *Ethical principles of psychologists and code of conduct* (2002, amended effective June 1, 2010, and January 1, 2017). https://www.apa.org/ethics/code/index.aspx

Arslan, G., Yıldırım, M., & Albertova, S. M. (2022). Development and initial validation of the Subjective Academic Wellbeing Measure: A new tool of youth wellbeing in school. *Journal of Positive School Psychology*, *6*(1), 3–11.

Avinç, E., & Doğan, F. (2024). Digital literacy scale: Validity and reliability study with the Rasch model. *Education and Information Technologies*, *29*(17), 22895–22941. https://doi.org/10.1007/s10639-024-12662-7

Beglar, D. (2010). A Rasch-based validation of the Vocabulary Size Test. *Language Testing*, *27*(1), 101–118. https://doi.org/10.1177/0265532209340194

Blouin, D., & Smith, E. V. (2020). Measuring the continuous quality improvement orientation of medical education programs: A validity study of the Baldrige "Are We Making Progress" questionnaire. *International Journal of Health Care Quality Assurance*, *33*(2), 158–171. https://doi.org/10.1108/IJHCQA-06-2019-0102

Bond, T. G., Yan, Z., & Heene, M. (2021). *Applying the Rasch Model: Fundamental measurement in the human sciences* (4th ed.). Routledge.

Boone, W. J., & Noltemeyer, A. (2017). Rasch analysis: A primer for school psychology researchers and practitioners. *Cogent Education*, *4*(1), Article 1416898. https://doi.org/10.1080/2331186X.2017.1416898

Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch analysis in the human sciences*. Springer. https://doi.org/10.1007/978-94-007-6857-4

Borgonovi, F., & Pál, J. (2016). *A framework for the analysis of student well-being in the PISA 2015 study* (Issue 140). https://doi.org/10.1787/5jlpszwghvvb-en

Buerger, S., Holzer, J., Yanagida, T., Schober, B., & Spiel, C. (2023). Measuring adolescents' well-being in schools: The adaptation and translation of the EPOCH Measure of Adolescent Well-being - A validation study. *School Mental Health*, *15*(2), 611–626. https://doi.org/10.1007/s12310-023-09574-1

Burkett-McKee, S., Knight, B. A., & Vanderburg, M. A. (2021). Psychological well-being of students with high abilities and their school's ecology: Is there a relationship? *Roeper Review*, *43*(3), 197–211. https://doi.org/10.1080/02783193.2021.1923593

Cavicchiolo, E., Manganelli, S., Lucidi, F., & Alivernini, F. (2025). Measuring emotional well-being at school among students with different characteristics: A population study. *Journal of Psychoeducational Assessment*, Article 07342829251354912. https://doi.org/10.1177/07342829251354912

Cheng, Y.-Y., Chen, L.-M., Liu, K.-S., & Chen, Y.-L. (2011). Development and psychometric evaluation of the school bullying scales: A Rasch measurement approach. *Educational and Psychological Measurement*, *71*(1), 200–216. https://doi.org/10.1177/0013164410387387

Chin, W. C., & Wu, S. L. (2020). The predicting effects of depression and self-esteem on suicidal ideation among adolescents in Kuala Lumpur, Malaysia. *Journal of Health and Translational Medicine*, *23*(1), 60–66. https://doi.org/10.22452/jummec.vol23no1.11

Chong, J., Mokshein, S. E., & Mustapha, R. (2022). Instrument validation based on the six aspects of Messick validity framework using the Rasch rating scale model. *Jurnal Penyelidikan Pendidikan*, *23*, 117–138.

Conesa, P. J., & Duñabeitia, J. A. (2021). The basic psychological needs in the classroom scale (BPN-CS). *Behavioral Sciences*, *11*(7), Article 96. https://doi.org/10.3390/bs11070096

Conrad, K. M., Conrad, K. J., Passetti, L. L., Funk, R. R., & Dennis, M. L. (2015). Validation of the full and short-form Self-Help Involvement Scale against the Rasch measurement model. *Evaluation Review*, *39*(4), 395–427. https://doi.org/10.1177/0193841X15599645

Counted, V., Cowden, R. G., & Lomas, T. (2024). Multidimensional flourishing in Africa: An intracontinental analysis of 38 well-being indicators in 40 countries. *Journal of Happiness Studies*, *25*(5), Article 51. https://doi.org/10.1007/s10902-024-00764-5

Creswell, J. W., & Clark, V. L. P. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE.

DeVellis, R. F., & Thorpe, C. T. (2022). *Scale development: Theory and applications* (5th ed.). SAGE.

Dias-Viana, J. L., & Noronha, A. P. P. (2022). Validity evidence for the School Subjective Well-Being Scale (Evidências de validade para a Escala de Bem-Estar Subjetivo Escolar). *Estudos de Psicologia (Campinas)*, *39*, Article e200220. https://doi.org/10.1590/1982-0275202239e200220

Fendrich, M., Smith Jr., E. V., Pollack, L. M., & Mackesy-Amiti, M. E. (2009). Measuring sexual risk for HIV: A Rasch scaling approach. *Archives of Sexual Behavior*, *38*(6), 922–935. https://doi.org/10.1007/s10508-008-9385-2

Ferguson, Y. L., Kasser, T., & Jahng, S. (2011). Differences in life satisfaction and school satisfaction among adolescents from three nations: The role of perceived autonomy support. *Journal of Research on Adolescence*, *21*(3), 649–661. https://doi.org/10.1111/j.1532-7795.2010.00698.x

Furrer, C., & Skinner, E. (2003). Sense of relatedness as a factor in children's academic engagement and performance. *Journal of Educational Psychology*, *95*(1), 148–162. https://doi.org/10.1037/0022-0663.95.1.148

Gong, Z., Xue, J., Han, Z., & Li, Y. (2021). Validation of the Chinese version of KIDSCREEN-10 Quality of Life Questionnaire: A Rasch model estimation. *Frontiers in Psychology*, *12*, Article 647692. https://doi.org/10.3389/fpsyg.2021.647692

Gorodzinsky, A. Y., Hainsworth, K. R., & Weisman, S. J. (2011). School functioning and chronic pain: A review of methods and measures. *Journal of Pediatric Psychology*, *36*(9), 991–1002. https://doi.org/10.1093/jpepsy/jsr038

Greene, L. M., Royal, K. D., Bradley, J. M., Lascelles, B. D. X., Johnson, L. R., & Hawkins, E. C. (2017). Severity of nasal inflammatory disease questionnaire for canine idiopathic rhinitis control: Instrument development and initial validity evidence. *Journal of Veterinary Internal Medicine*, *31*, 134–141. https://doi.org/10.1111/jvim.14629

Ha, H. T. (2021). A Rasch-based validation of the Vietnamese version of the Listening Vocabulary Levels Test. *Language Testing in Asia*, *11*(1), Article 16. https://doi.org/10.1186/s40468-021-00132-7

Hascher, T. (2007). Exploring students' well-being by taking a variety of looks into the classroom. *Hellenic Journal of Psychology*, *4*(3), 331–349.

Hascher, T., & Hagenauer, G. (2020). Swiss adolescents' well-being in school. *Swiss Journal of Educational Research*, *42*(2), 367–390. https://doi.org/10.24452/sjer.42.2.5

Holzer, J., Bürger, S., Lüftenegger, M., & Schober, B. (2022). Revealing associations between students' school-related well-being, achievement goals, and academic achievement. *Learning and Individual Differences*, *95*, Article 102140. https://doi.org/10.1016/j.lindif.2022.102140

Holzer, J., Bürger, S., Samek-Krenkel, S., Spiel, C., & Schober, B. (2021). Conceptualisation of students' school-related wellbeing: Students' and teachers' perspectives. *Educational Research*, *63*(4), 474–496. https://doi.org/10.1080/00131881.2021.1987152

Hossain, S., O'Neill, S. C., & Strnadová, I. (2019). What really matters for students' school satisfaction in Bangladesh? *Psychology in the Schools*, *56*(5), 670–689. https://doi.org/10.1002/pits.22226

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. https://doi.org/10.1007/s12187-022-09990-w

Huppert, F. A., & So, T. T. C. (2013). Flourishing across Europe: Application of a new conceptual framework for defining well-being. *Social Indicators Research*, *110*(3), 837–861. https://doi.org/10.1007/s11205-011-9966-7

Jayawickreme, E., Forgeard, M. J. C., & Seligman, M. E. P. (2012). The engine of well-being. *Review of General Psychology*, *16*(4), 327–342. https://doi.org/10.1037/a0027990

Joing, I., Vors, O., & Potdevin, F. (2020). The subjective well-being of students in different parts of the school premises in French middle schools. *Child Indicators Research*, *13*(4), 1469–1487. https://doi.org/10.1007/s12187-019-09714-7

Kern, M. L., Benson, L., Steinberg, E. A., & Steinberg, L. (2016). The EPOCH measure of adolescent well-being. *Psychological Assessment*, *28*(5), 586–597. https://doi.org/10.1037/pas0000201

Kern, M. L., Zeng, G., Hou, H., & Peng, K. (2019). The Chinese version of the EPOCH measure of adolescent well-being: Testing cross-cultural measurement invariance. *Journal of Psychoeducational Assessment*, *37*(6), 757–769. https://doi.org/10.1177/0734282918789561

Khine, M. S. (2020). Objective measurement in psychometric analysis. In M. S. Khine (Ed.), *Rasch measurement applications in quantitative educational research* (pp. 3–7). Springer. https://doi.org/10.1007/978-981-15-1800-3

Linacre, J. M. (1994). Sample size and item calibration (or person measure) stability. *Rasch Measurement Transactions*, *7*(4), 328. https://www.rasch.org/rmt/rmt74m.htm

Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement*, *3*(1), 85–106.

Linacre, J. M. (2016). *A user's guide to Winsteps/Ministep Rasch-Model Computer Programs* (5.3.3). Winsteps.com.

Long, R. F., Huebner, E. S., Wedell, D. H., & Hills, K. J. (2012). Measuring school-related subjective well-being in adolescents. *American Journal of Orthopsychiatry*, *82*(1), 50–60. https://doi.org/10.1111/j.1939-0025.2011.01130.x

Lu, L., & Gilmour, R. (2004). Culture and conceptions of happiness: Individual oriented and social oriented SWB. *Journal of Happiness Studies*, *5*, 269–291. https://doi.org/10.1007/s10902-004-8789-5

Lu, X., & Yeo, K. J. (2015). Psychometric properties of the Internet Addiction Test in a sample of Malaysian undergraduate students. *Psicología Educativa*, *21*(1), 17–25. https://doi.org/10.1016/j.pse.2015.03.001

Makransky, G., Rogers, M. E., & Creed, P. A. (2015). Analysis of the construct validity and measurement invariance of the Career Decision Self-Efficacy Scale: A Rasch model approach. *Journal of Career Assessment*, *23*(4), 645–660. https://doi.org/10.1177/1069072714553555

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). Macmillan.

Messick, S. (1995). Standards of validity and the validity of standards in performance assessment. *Educational Measurement: Issues and Practice*, *14*(4), 5–8. https://doi.org/10.1111/j.1745-3992.1995.tb00881.x

Mohamed, N. F., Jiun Ting, T., Ab Manan, N., Mohd Senari, I. F., Muhammad Firdaus Chan, M. F., Rahmatullah, B., Govindasamy, P., & Abdulla, K. (2026). Prevalence and predictors of social anxiety disorders among Malaysian secondary school students during the COVID-19 pandemic: Exploring the influence of internet gaming disorder and impulsivity. *Clinical Child Psychology and Psychiatry*, *31*(1), 69–83. https://doi.org/10.1177/13591045231206967

Niclasen, J., Keilow, M., & Obel, C. (2018). Psychometric properties of the Danish student well-being questionnaire assessed in >250,000 student responders. *Scandinavian Journal of Public Health*, *46*(8), 877–885. https://doi.org/10.1177/1403494818772645

Noroozi, S., & Karami, H. (2024). A Rasch-based validation of the University of Tehran English Proficiency Test (UTEPT). *Language Testing in Asia*, *14*(1), Article 18. https://doi.org/10.1186/s40468-024-00290-4

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). McGraw-Hill.

Peng, Y., Alias, B. S., Mansor, A. N., & Ab Hamid, M. R. (2025). Development and validation of the principals' digital leadership instrument using Rasch measurement model. *International Journal of Evaluation and Research in Education*, *14*(3), 1577–1589. https://doi.org/10.11591/ijere.v14i3.32214

Pflug, J. (2009). Folk theories of happiness: A cross-cultural comparison of conceptions of happiness in Germany and South Africa. *Social Indicators Research*, *92*(3), 551–563. https://doi.org/10.1007/s11205-008-9306-8

Phipps, J. (2023). The validation of two L2 self-efficacy instruments using Rasch analysis. *Research Methods in Applied Linguistics*, *2*(3), Article 100084. https://doi.org/10.1016/j.rmal.2023.100084

Putwain, D. W., Loderer, K., Gallard, D., & Beaumont, J. (2019). School-related subjective well-being promotes subsequent adaptability, achievement, and positive behavioural conduct. *British Journal of Educational Psychology*, *90*(1), 92–108. https://doi.org/10.1111/bjep.12266

Putwain, D. W., Stockinger, K., von der Embse, N. P., Suldo, S. M., & Daumiller, M. (2021). Test anxiety, anxiety disorders, and school-related wellbeing: Manifestations of the same or different constructs? *Journal of School Psychology*, *88*(2), 47–67. https://doi.org/10.1016/j.jsp.2021.08.001

Raccanello, D., Vicentini, G., Trifiletti, E., & Burro, R. (2021). A Rasch analysis of the School-related Well-being (SRW) scale: Measuring well-being in the transition from primary to secondary school. *International Journal of Environmental Research and Public Health*, *18*(1), Article 23. https://doi.org/10.3390/ijerph18010023

Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational Statistics*, *4*(3), 207–230.

Renshaw, T. L., Long, A. C. J., & Cook, C. R. (2015). Assessing adolescents' positive psychological functioning at school: Development and validation of the Student Subjective Wellbeing Questionnaire. *School Psychology Quarterly*, *30*(4), 534–552. https://doi.org/10.1037/spq0000088

Royal, K. D., & Royal, R. A. (2017). An evaluation of the psychometric properties of the Behavior Change Strategies for Healthy Eating Scale. *Journal of Nursing Measurement*, *25*(3), 411–420. https://doi.org/10.1891/1061-3749.25.3.411

Ryan, R. M., & Deci, E. L. (2000). Self-Determination Theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Sabramani, V., Idris, I. B., Ismail, H., Nadarajaw, T., Zakaria, E., & Kamaluddin, M. R. (2021). Bullying and its associated individual, peer, family and school factors: Evidence from Malaysian national secondary school students. *International Journal of Environmental Research and Public Health*, *18*(13), Article 7208. https://doi.org/10.3390/ijerph18137208

Sahril, N., Abd Hamid, H. A., Abd Razak, M. A., Sheikh Ilman, S. S., Noor Ansari, F., & Ahmad, N. A. (2025). Risk and protective factors associated with depressive symptoms among school-going adolescents in Malaysia: A cross-sectional study. *BMC Psychiatry*, *25*(1), Article 891. https://doi.org/10.1186/s12888-025-07319-9

Seligman, M. E. P. (2011). *Flourish: A visionary new understanding of happiness and well-being*. Simon & Schuster.

Seligman, M. E. P., & Csikszentmihalyi, M. (2000). Positive psychology. An introduction. *The American Psychologist*, *55*(1), 5–14. https://doi.org/10.1037/0003-066X.55.1.5

Seligman, M. E. P., Ernst, R. M., Gillham, J., Reivich, K., & Linkins, M. (2009). Positive education: Positive psychology and classroom interventions. *Oxford Review of Education*, *35*(3), 293–311. https://doi.org/10.1080/03054980902934563

Son, M., & Ha, M. (2025). Development of a digital literacy measurement tool for middle and high school students in the context of scientific practice. *Education and Information Technologies*, *30*(4), 4583–4606. https://doi.org/10.1007/s10639-024-12999-z

Sovey, S., Osman, K., & Mohd Matore, M. E. E. (2022). Rasch analysis for disposition levels of computational thinking instrument among secondary school students. *Eurasia Journal of Mathematics, Science and Technology Education*, *18*(3), Article em2088. https://doi.org/10.29333/ejmste/11794

Stang-Rabrig, J., Schwerter, J., Witmer, M., & McElvany, N. (2023). Beneficial and negative factors for the development of students' well-being in educational context. *Current Psychology*, *42*(35), 31294–31308. https://doi.org/10.1007/s12144-022-04067-x

Stipek, D. J. (2002). *Motivation to learn: From theory to practice* (4th ed.). Allyn & Bacon.

Susanto, H., Wedyaswari, M., & Dalimunthe, K. L. (2023). A content validity and cognitive interview to develop the HARMONI items: Instrument measuring student well-being in West Java, Indonesia. *Frontiers in Education*, *8*, Article 1132031. https://doi.org/10.3389/feduc.2023.1132031

Tennant, A., & Conaghan, P. G. (2007). The Rasch measurement model in rheumatology: What is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis & Rheumatism*, *57*(8), 1358–1362. https://doi.org/10.1002/art.23108

Van Landeghem, G., Van Damme, J., Opdenakker, M.-C., De Fraine, B., & Onghena, P. (2002). The effect of schools and classes on noncognitive outcomes. *School Effectiveness and School Improvement*, *13*(4), 429–451. https://doi.org/10.1076/sesi.13.4.429.10284

Wasserman, J. D., & Bracken, B. A. (2003). Pyschometric characteristics of assessment procedures. In J. R. Graham & J. A. Naglieri (Eds.), *Handbook of psychology: Assessment psychology (Volume 10)* (pp. 43–66). John Wiley & Sons.

Wentzel, K. (2024). Student well-being: In search of definitions, measures, and research designs. *Learning and Instruction*, *94*, Article 101990. https://doi.org/10.1016/j.learninstruc.2024.101990

Wolfe, E. W., & Smith Jr., E. V. (2007). Instrument development tools and activities for measure validation using Rasch models: Part II--validation activities. *Journal of Applied Measurement*, *8*(2), 204–234.

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

Wright, B. D., & Masters, O. N. (1982). *Rating scale analysis (Rasch measurement)*. MESA Press.

Zwick, R., Thayer, D. T., & Lewis, C. (1999). An empirical bayes approach to mantel-haenszel DIF analysis. *Journal of Educational Measurement*, *36*(1), 1–28. https://doi.org/10.1111/j.1745-3984.1999.tb00543.x