BMC Psychology

https://doi.org/10.1186/s40359-026-04237-7

Article in Press

Psychometric validation of the German Warwick-Edinburgh Mental Well-being Scale (WEMWBS) in an undergraduate health professions students cohort in Switzerland: a cross-sectional study

Received: 22 July 2025

Accepted: 20 February 2026

Published online: 28 February 2026

Cite this article as: Sterkele I., Koenig I., Rogan S. *et al.* Psychometric validation of the German Warwick-Edinburgh Mental Well-being Scale (WEMWBS) in an undergraduate health professions students cohort in Switzerland: a cross-sectional study. *BMC Psychol* (2026). https://doi.org/10.1186/s40359-026-04237-7

Iris Sterkele, Irene Koenig, Slavko Rogan, André Meichtry, Evert Zinzen & Mirjam Koerner

Psychometric Validation of the German Warwick-Edinburgh Mental Well-being Scale (WEMWBS) in an Undergraduate Health Professions Students Cohort in Switzerland: A Cross-sectional Study

Iris Sterkele[1,2], Irene Koenig[1], Slavko Rogan[1], André Meichtry[1], Evert Zinzen[2], and Mirjam Koerner[1]

[1] Bern University of Applied Sciences, School of Health Professions, Bern, Switzerland

[2] Vrije Universiteit Brussel, Department of Movement and Sport Sciences, Brussels, Belgium

Corresponding author: Iris Sterkele

Correspondence: Iris.Sterkele@bfh.ch

Abstract

Background: Mental well-being of health professions students is under pressure. Therefore, validated measurement tools that assess and monitor mental well-being in this population are in high demand. The Warwick-Edinburgh Mental Well-Being Scale (WEMWBS) is an internationally established measurement tool for the assessment of mental well-being (hedonic and eudaimonic) that has been validated in several populations, including students. However, to date not yet for German-speaking undergraduate health professions students in Switzerland.

Methods: A total of 290 undergraduate health professions students from a university of applied sciences in Switzerland participated in this cross-sectional study. The psychometric properties of the German WEMWBS were evaluated at both the scale and item levels by applying a combined approach using Classical Test Theory (CTT) and Item Response Theory (IRT) analyses.

Results: The three-factor model showed the best fit in the Swiss German context with acceptable to good internal consistency for all subscales and excellent internal consistency for the entire scale. IRT analyses using the Graded Response Model indicated moderate to good item discrimination and well-ordered thresholds across the latent trait continuum. The scale showed high measurement precision at a broad range of trait levels, supporting its reliability and psychometric adequacy.

Conclusion: The German WEMWBS demonstrated strong psychometric properties for assessing mental well-being among Swiss health professions students. It offers a practical tool for routine monitoring, enabling institutions to identify students with low mental well-being and guide early support efforts.

Trial registration: Clinical trial number not applicable.

Keywords: Mental Well-being, Warwick-Edinburgh Mental Well-Being Scale, Item Response Theory, Psychometric properties, health professions students

Background

Students' mental well-being (MWB), especially in health-related fields, is an increasing concern for educators, healthcare institutions, and policymakers. Health professions students often face intense academic demands, ethical dilemmas, and tough clinical training in practice – all of which contribute to increase level of psychological distress (1). This is particularly alarming, given that research across various countries indicates a decline in the MWB of university students, including those in health professions (2–7). MWB is generally understood as a concept that represents the positive dimension of mental health, beyond the absence of mental illness, as it is shaped by both external circumstances and how we respond to them (8). It is conceptualized as a multifaceted construct that comprises two distinct elements: the hedonic element, which is concerned with the subjective experience of happiness and life satisfaction; and the eudaimonic element, which is concerned with psychological functioning and self-realisation (9). High levels of mental wellbeing support students in maintaining-active engagement in both academic studies and clinical training. Moreover, research indicates that higher levels of MWB and life satisfaction are associated with beneficial outcomes, including increased study engagement, improved academic performance, and achievement, higher academic satisfaction, and better physical health (10–13). Given its importance, assessing students MWB is essential. As a result, there is a strong demand for validated instruments that accurately measure MWB outcomes. Internationally recognized tools are particularly valuable, as they allow for meaningful cross-country comparisons. One of the most widely applied questionnaires is the Warwick-Edinburgh Mental Well-being Scale (WEMWBS). The original WEMWBS was developed in 2006 in English as part of the Scottish Government's National Programme for improving mental health and well-being. It

was commissioned by NHS Health Scotland and developed in collaboration with researchers from the universities of Warwick and Edinburgh (14). The WEMWBS has been translated into multiple languages and has undergone cross-cultural validation in numerous countries (15), primarily among adult populations aged 18 to 75 (16). However, fewer studies have focused on university students' samples from Europe, Asia, and the Middle East (9,17–22). Even fewer studies have focused on samples of undergraduate health professions students, one with a sample of Slovenian and one with a sample of Chinese nursing students (23,24). Beyond nursing students, research involving students from other health disciplines such as midwifery, physiotherapy or nutrition and dietetics studying at universities of applied sciences remains rare. This group is unique because they face a dual academic-clinical workload and will play a critical role in future health systems, making their MWB essential for both educational success and patient care. Moreover, most studies examining the psychometric properties of the WEMWBS in university students have utilized Classical Test Theory (CTT), while relatively few have applied alternative methodologies such as Item Response Theory (IRT). Among these, various IRT models have been employed, including the Rasch model (25), it's extensions such as the Rating Scale Model (23), and the Graded Response Model (26). Combining CTT and IRT offers complementary insights: CTT provides test-level metrics such as reliability and item-total correlations, while IRT delivers detailed item-level analysis across the latent trait continuum. This combined approach enables a more comprehensive evaluation of measurement quality. The original WEMWBS has been translated into German and cross-culturally validated using a community-based sample of adults in Austria (16) and in Germany (27). In contrast to other validation studies e.g., (20), that have supported a unidimensional factor structure of the original WEMWBS, evidence for unidimensionality could not be clearly replicated in the German-speaking context. Instead, the data provided evidence for a multidimensional latent structure, supporting bifactor and three factor models

(16,27). However, no psychometric validation of the German WEMWBS has yet been conducted among undergraduate health professions students. To address this gap, the present study examines the psychometric properties of the German WEMWBS using CTT and IRT analyses. The objective of this study is to evaluate the structural validity, reliability, and item-level- measurement properties of the scale for assessing MWB among Swiss German -speaking- undergraduate health professions students.

Methods

Study design and participants

A cross-sectional study design was used to validate the German version of the WEMWBS among Swiss undergraduate students in health professions. Participants were bachelor's students enrolled in nursing, physiotherapy, nutrition and dietetics, or midwifery programmes at our School of Health, part of a University of Applied Sciences in the German-speaking region of Switzerland, across two campuses. N= 780 undergraduate health professions students at our School of Health were invited to participate in the study. The final sample consisted of n = 290 students who completed the survey. Sample size adequacy was based on established guidelines recommending 5-10 participants per item and at least 100-300 participants for factor analysis (28). For Graded Response Models (GRM), large samples are required; poor parameter estimation of some category thresholds has been reported with 350 participants, and recommendations of approximately 500 participants have been made for a 25-item assessment (29,30). The sample was broadly representative in terms of age and gender, and partially representative regarding year of study and health-related disciplines. However, institutional diversity of the national student population may not be fully reflected.

Data collection

The data were collected in June 2024 through an online survey using Evasys (Evasys GmbH, Lüneburg, Germany). Evasys is a fully automated survey and exam solution for online, paper and hybrid use (https://evasys.de/en/). The online surveys included sociodemographic and professional characteristics (age, gender, degree programme, semester, and study site), and the German WEMWBS (16). Study participants were invited via email, and faculty members also informed students about the study during class sessions. The email included information about the study purpose, informed consent, and the survey link.

Materials

Warwick-Edinburgh Mental Well-being Scale (WEMWBS)

The German WEMWBS is a 14-item questionnaire of MWB focusing entirely on positive aspects and covering both feelings and functioning aspects of mental health (16). Study participants are asked to describe their experiences over a two-week reference period using a 5-point Likert scale ranging from 1 = "none of the time" to 5 = "all of the time". The total score of the WEMWBS is calculated as the sum of all items. The range of the WEMWBS is from 14 to 70 points with a higher score indicating a higher level of MWB (9,16). The German version of the WEMWBS has previously been used in community samples in Germany and Austria, supporting its suitability for validation in a Swiss German-speaking student population (16,27).

Ethical considerations

A clarification of responsibilities for this study (registration number: Req-2024-01645) was obtained from the Ethics Committee Bern, and it was determined that this study does not require any additional ethical approval procedures under the Swissethics guidelines. Data were collected anonymously and treated confidentially. The non-personal data was stored in a

data repository. It may be used for research purposes in compliance with the applicable data protection guidelines.

Statistical analyses

Descriptive statistics

A missing data analysis was conducted by frequency analysis and descriptive statistics (means, standard deviations, and skewness) were computed for each item and total WEMWBS score to examine item and total score distributions and detect potential floor or ceiling effects. Item-total correlations were calculated to assess item discrimination.

Reliability

The internal consistency of the scale was estimated using Cronbach's alpha and model-based McDonald's omega coefficients to provide a more robust estimate of reliability, particularly given the ordinal nature of the data. Values ≥ 0.70 were considered acceptable (31).

Structural validity

To assess the structural validity of the German WEMWBS, confirmatory factor analyses (CFA) were conducted for different models (one-factor, three-factor, and bifactor), as suggested by previous studies (16,27). The CFA was estimated using the robust weighted least squares estimator (WLSMV), appropriate for ordinal data. Model fit was evaluated using multiple indices, including the chi-square (χ^2), degree of freedom (df), the Comparative Fit Index (CFI), the Root Mean Square Error of Approximation (RMSEA), and the Standardized Root Mean Square Residual (SRMR). We followed the guidelines proposed by Hu and Bentler (32) that models with a CFI of ≥ 0.95, RMSEA of ≤ 0.06 and an SRMR of ≤ 0.08 are representative of well-fitting models (32). Additionally, we inspected the data structure through an exploratory bifactor analysis (EBFA) before performing the confirmatory bifactor

analysis (CBFA) as proposed by Reise (33) to verify the data structure and to identify potential cross-loadings that could distort parameter estimates in a confirmatory framework. Given the ordinal 5-point Likert scale, a polychoric correlation matrix was used for the EBFA. Minimum residual factor analysis (minres) with bifactor rotation was applied to explore the factor structure of the bifactor model. A factor loading of ≥ 0.4 was set as the cut-off point for item selection. The Explained Common Variance (ECV) was computed to evaluate the dominance of the general factor. An ECV > 0.70 suggests that most common variance is attributable to the general factor (34).

Item response theory analyses

To further assess item functioning and scale precision of the WEMWBS in an undergraduate health professions student sample, an Item Response Theory (IRT) analysis was conducted using the GRM, suitable for ordered Likert response formats (35). Prior to conducting the GRM analysis, key assumptions unidimensionality, local independence, and monotonicity of item response functions were tested. Unidimensionality was evaluated using model fit statistics (χ^2, CFI, RMSEA, SRMR) and inspection of eigenvalues. Local independence was assessed using residual correlations (Yen's Q3), while values >0.20 indicating potential local dependence. Within the framework of the GRM analysis, item discrimination (a) and threshold (b) parameters were estimated to assess how effectively each item differentiates individuals along the latent trait continuum. In addition, item probability functions (IPFs), item information functions (IIFs), and test information function (TIF) were examined to evaluate the scale's precision across varying levels of the latent trait (θ). IRT-based reliability was assessed by examining the conditional reliability function derived from the TIF. The corresponding marginal reliability coefficient was computed, reflecting the average reliability across the latent trait continuum.

All analyses were conducted in R (version 4.5.0) using the following packages: “psych” for descriptive statistics and reliability, “GPArotation” for EFA with polychoric correlations, “lavaan” for CFA, “SemTools” for reliability and “mirt” for the IRT-based GRM.

Results

Descriptive results

Out of 780 undergraduate health professions students n=290 students (37%) participated in the survey. A total of five responses (2% of data points) to different items of the WEMWBS were missing. The sample consisted of 40% physiotherapy, 30% nursing, 17% midwifery, and 13% nutrition and dietetics students. Most of the participants were female (84%), with a smaller proportion of males (15%) and a small number of individuals who identified as diverse (1%). The average age of the study participants was 25 years, SD age 4.67 (Table 1). Overall, study participants scored the WEMWBS items predominately “positive” with item means between 2.7 (item 3) to 3.9 (item 12) and standard deviations between 0.7 (item 14) and 1.0 (item 13). The means, standard deviations, skewness, and corrected item–total correlations for each of the 14 WEMWBS items (n= 290) are displayed in Table 3. The total WEMWBS score is based on cases that completed the WEMWBS in full, ranging between 23 and 68 with a mean (SD) of 48.62 (8.23). The sample distribution was slightly negative skewed (skew = -0.3, kurtosis = -0.17) but showed neither floor nor ceiling effects (Figure 1).

Table 1. Sociodemographic characteristics of study participants

Variables	Categories	n	%	Mean (SD)
Gender	Male	42	14.5	
	Female	243	84.4	
	Diverse	3	1.0	
	Missing	2	0.1	
Age (years)				24.5

				(4. 67)
Bachelor's degree programme	Physiotherapy	115	40	
	Nursing	87	30.0	
	Midwifery	48	16.6	
	Nutrition and dietetics	39	13.5	
Semester	First semester	3	1.0	
	Second semester	77	26.8	
	Third semester	16	5.6	
	Fourth semester	100	34.8	
	Fifth semester	6	2.1	
	Sixth semester	85	29.6	
	Missing	3	0.1	
Study site	Basel	74	25.5	
	Bern	215	74.1	
	Missing	1	0.4	

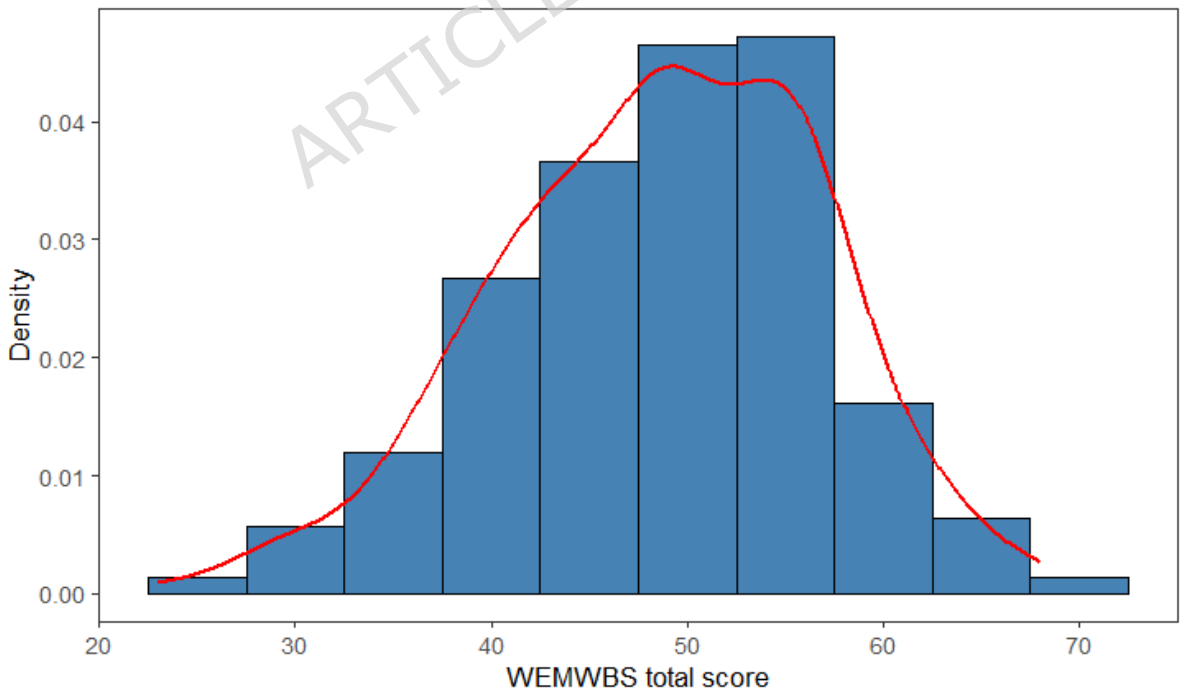


Figure 1. WEMWBS total score distribution

Reliability

Table 2 presents the internal consistency of the WEMWBS, assessed using Cronbach's alpha and model-based McDonald's omega coefficients. The results indicated excellent Cronbach's alpha and Mc Donald's omega coefficients for the one-factor model (α= 0.91 and ω = 0.91). For the three-factor model high reliability was exhibited for Factor 1 (α= 0.80 and ω = 0.80) and Factor 3 (α= 0.85 and ω = 0.86), and acceptable reliability for Factor 2 (α= 0.75 and ω = 0.76). In addition, the total reliability across all items, reflecting the internal consistency of the full scale, was high (α= 0.91 and ωt = 0.92). In the bifactor model, high reliability was observed for the general factor (ωH = 0.91), indicating that 91% of the variance in the total score is attributable to the general factor. In contrast, the group factor showed low reliability (ωHS = 0.29), meaning that 29% of the variance in the subscale score, beyond the general factor, is attributable to the specific group factor.

Structural validity

Confirmatory factor analyses

A comparison of fit indices for the tested one-factor (M1), three-factor (M2), and bifactor (M3) models, as well as their modified versions (M1a, M2a), is presented in Table 2. Among the tested models, the three-factor model (M2) showed best overall model fit, although the RMSEA was slightly higher than desired (χ 2 = 174.807 (df = 74), standard CFI = 0.99, robust CFI = 0.89, standard RMSEA = 0.069 (90% CI: 0.056, 0.083), robust RMSEA = 0.112 (90% CI: 0.097, 0.127), and standard SRMR = 0.076). The correlation matrix showed high to extremely high correlation between the factors (F1 and F2: r = 0.80, F1 and F3: r = 0.91, F2 and F3: r = 0.7), indicating a strong common underlying structure. Allowing one cross-loading (item 14 on factor 2) further improved model fit (M2a), reducing RMSEA under the recommended threshold. The bifactor model (M3), specified as one general factor and one specific group factor, showed good model fit indices, although RMSEA was above the

recommended threshold. The CBFA produced warnings indicating the information matrix could not be inverted, resulting in no standard errors or robust fit indices being available. The initial one-factor model (M1) also showed good fit according to most indices, but demonstrated the highest standard and robust RMSEA values, which were above the recommended threshold. Modification indices suggested several error covariances among items sharing similar content domains. Adding four error covariances (items 9-12, 1-2, 4-9, and 1-10) stepwise substantially improved model fit (M1a), reducing RMSEA from >0.10 to approximately 0.06.

Table 2. Model comparison resulting from CFA and internal consistency

Model	**χ 2(df)**	**Standard CFI**	**Robust CFI**	**Standard RMSEA [90% CI]**	**Robust RMSEA**	**SRMR**	**Cronbach's Alpha (α)**	**McDonald's Omega**
One-factor (M1)	308.980 (77)	0.98	0.82	0.103 [0.091, 0.115]	0.139 [0.125, 0.154]	0.078	0.91	ω= 0.91
One-factor with 4 error covariances (M1a)	149.757 (73)	0.99	0.89	0.061 [0.047, 0.075]	0.108 [0.093, 0.124]	0.057	0.91	ω= 0.89
Three-factor (M2)	174.807 (74)	0.99	0.89	0.069 [0.056, 0.083]	0.112 [0.097, 0.127]	0.076	F1: 0.80 F2: 0.75 F3: 0.85 Total: 0.91	F1: ω= 0.80 F2: ω = 0.76 F3: ω= 0.86 ωt= 0.92
Three-factor with cross-loading of item 14 on factor 2 (M2a)	145.214 (73)	0.99	0.91	0.059 [0.045, 0.073]	0.105 [0.090, 0.120]	0.054	F1: 0.80 F2: 0.80 F3: 0.85 Total: 0.91	F1: ω= 0.67 F2: ω= 0.67 F3: ω= 0.86 ωt= 0.92
Bifactor (M3)	273.91 (75)	0.98		0.097 [0.085, 0.109]		0.074		G: ωH= 0.91 S1: ωHS= 0.29 ωt= 0.92

χ2: chi-square test statistic, df: degrees of freedom, RMSEA: Root Mean Standard Error of Approximation, CFI: Comparative Fit Index, CI: confidence interval, SRMR: Standardised

Root Mean Square Residual, F1: hedonic factor, F2: interpersonal relationship factor, F3: eudaimonic factor, ω: McDonald's omega G: general factor, S1: group factor 1, , ωH: McDonald's omega hierarchical, ωHS: McDonald's omega hierarchical specific, ωt: McDonald's omega total

Exploratory bifactor analysis

Data were considered suitable for EBFA as the KMO sampling adequacy index yielded overall MSA = 0.92 and Bartlett's Sphericity test was statistically significant (χ^2= 2429.398, df=91, p <0.001). The parallel analysis indicated four factors with observed eigenvalues (6.72, 0.75, 0.33, 0.24) exceeding those from random data (0.51, 0.32, 0.25, 0.19). The EBFA with four factors showed that all items loaded meaningfully on Factor 1 (loadings 0.6 to 0.8), while only a few items showed substantial loadings on secondary factors. One item (item 4) loaded strongly on Factor 2 (loading 0.8), two items (item 1 and 2) loaded strongly on Factor 3 (loadings 0.47 and 0.6), and one item (item 12) loaded on Factor 4 (loading 0.46) (Table 3). The cumulative variance explained was 0.64. The ECV was 0.77, indicating that the general mental well-being factor accounted for 77% of the common variance. Although four factors were retained for exploratory purposes, the dominance of the general factor, reflected by a substantially larger first eigenvalue relative to all subsequent eigenvalues, combined with an ECV > 0.70 and strong item-loadings on Factor 1 supports treating the scale as essentially unidimensional.

Table 3. Descriptive statistics and exploratory bifactor analysis

Item	Short item wording	Mean (SD)	Skew	Corrected item-total correlation	Factor loading			
					Factor 1	Factor 2	Factor 3	Factor 4

1	Optimistic	3.4 (0.86)	-0.3	0.61	0.6		0.6	
2	Useful	3.4 (0.92)	-0.3	0.54	0.6		0.5	
3	Relaxed	2.7 (0.91)	0.1	0.62	0.7			
4	Interested in other people	3.9 (0.82)	-0.7	0.56	0.6	0.8		
5	Energy to spare	2.9 (0.9)	0.1	0.65	0.7			
6	Dealing with problems	3.5 (0.8)	-0.5	0.69	0.7			
7	Thinking clearly	3.5 (0.85)	-0.4	0.73	0.8			
8	Feeling good about myself	3.5 (0.87)	-0.3	0.74	0.8			
9	Close to other people	3.7 (0.87)	-0.5	0.63	0.7			
10	Confident	3.4 (0.89)	-0.3	0.74	0.8			
11	Able to make up mind	3.8 (0.87)	-0.7	0.62	0.6			
12	Loved	3.9 (0.89)	-0.7	0.55	0.7			-0.46
13	Interested in new things	3.4 (1.0)	-0.3	0.61	0.6			
14	Cheerful	3.5 (0.7)	-0.3	0.71	0.8			
Cumulative variance explained					0.5	0.53	0.59	0.64

Graded response model analysis

Assumption testing supported the application of the GRM, as factor analytic results indicated essential unidimensionality of the scale. Although the data are slightly multidimensional, the

GRM performs well because the factors are strongly correlated, which aligns with Reise et al.'s (33) assertion that unidimensional IRT models can be robustly applied to multidimensional data when multiple latent traits are moderately intercorrelated or when a strong general factor underlies the data. No substantial violations of local independence or monotonicity were observed (all items revealed Loevinger $H > 0.3$), confirming that the data met the necessary conditions for GRM estimation.

Discrimination and threshold parameters

The GRM analysis demonstrated item discrimination parameters (a) ranging from 1.2 to 2.4, indicating moderate to high discrimination across items. Items 6, 7, 8, 10, and 14 demonstrated particularly strong discrimination (a > 2.0), suggesting high ability to distinguish between trait levels. Threshold parameters (b1–b4) were consistently ordered and increase across all items, indicating that higher response categories corresponded to increasing levels of the latent construct. They ranged from approximately -2 to -4.6 SD below the mean (b1) to around +1 to +3.4 SD above the mean (b4), with many items, except items 3 and 5, clustering forward the lower end of the logit scale. This indicate that the items are relatively easy, with many items measuring small amounts of the trait and few (item 3 and 5) that are endorsed only by individuals who have high MWB level. Most Items (items 1, 2, 6, 7, 8, 10, 13, and 14 showed threshold parameters close to zero at b3, indicating that the transition between response categories 3 ("some of the time") and 4 ("often") occurs at average levels of the latent trait. This suggests that these items are well suited to discriminate among individuals with mid-range MWB levels.

Item probability and information functions

Figure 2 shows item probability (IPF) and item information (IIF) functions. The IPFs showed that as the latent trait (θ) increases, individuals are more likely to choose higher response categories. The gradual transition from one response option to next cross the spectrum of the latent trait indicates that the items are capturing incremental increases in the MWB level, which is consistent with the gradual increase in threshold parameters (b-parameters). For most items (item 1, 3, 4, 6, 7, 10, 11 and 14) response category 4 ("often") has the highest probability of endorsement at trait levels around or slightly above the mean. This suggests that individuals with average to slightly higher levels of MWB are most likely to choose this response option. The IIFs showed that most items provide the highest amount of information near the centre to the left side of the latent trait distribution, suggesting that the scale is most precise for screening students at risk of low or average levels of MWB. Items 5, 6, 7, 8, 10, and 14 provided the highest amount of information and exhibited multiple peaks across different levels of the latent trait. This suggests that these items contribute most to the scale's measurement precision and discriminate well across several parts of the latent trait continuum.

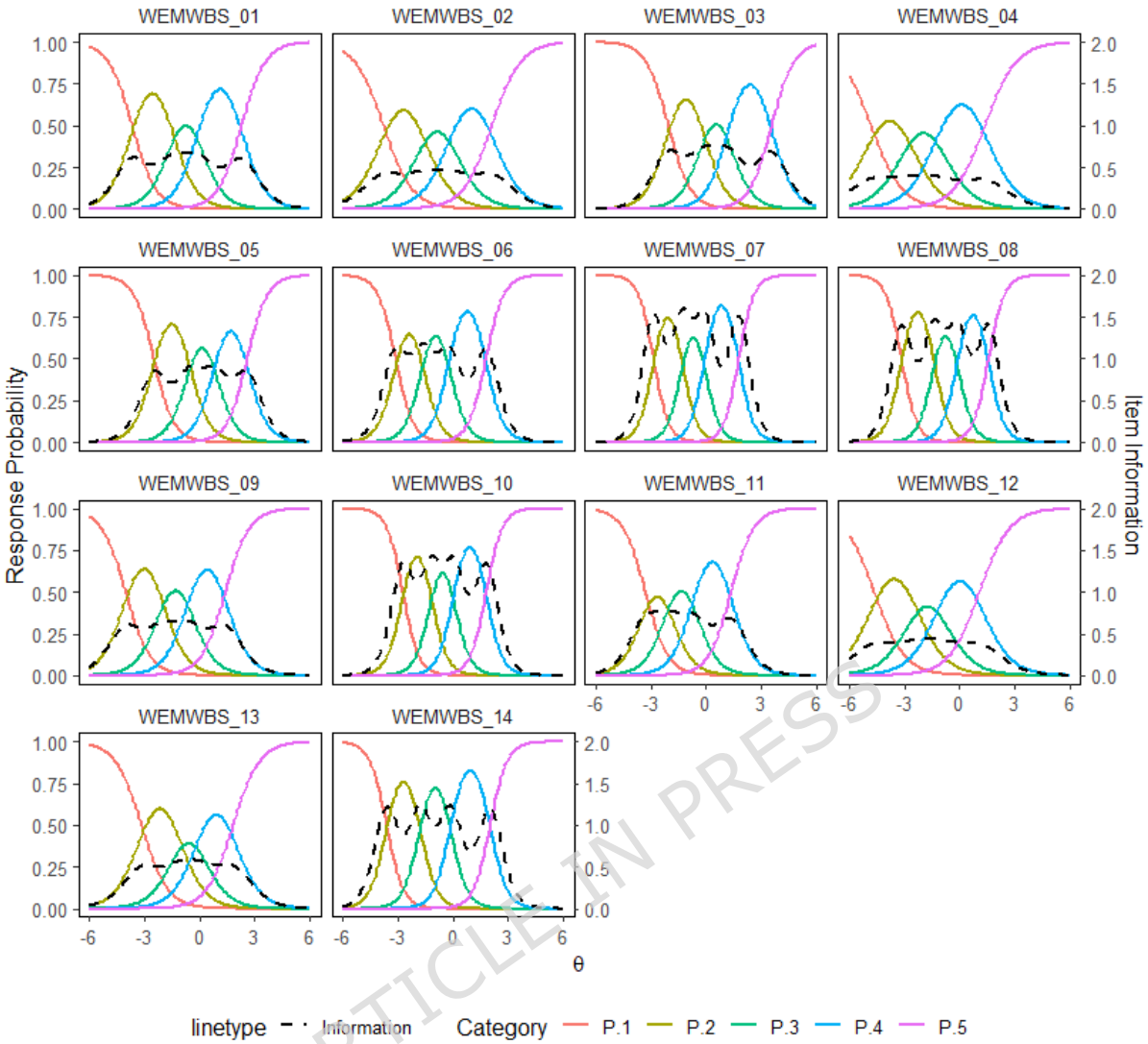
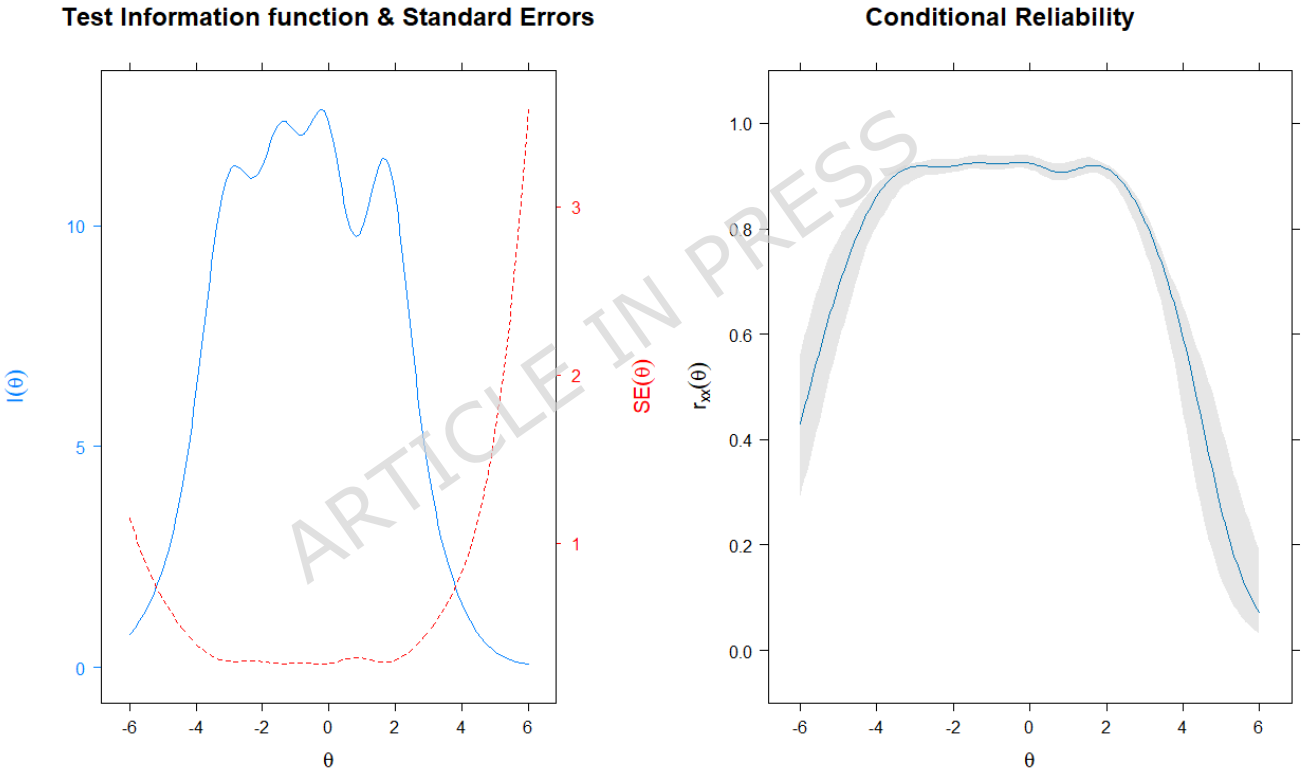


Figure 2. Item Probability and Item Information functions of the WEMWBS scale

Test information function and conditional reliability

Figure 3 shows on the left side of the figure the test information function (TIF), and on the right side the conditional reliability function. The TIF showed that the scale was most precise for respondents with trait levels (θ) between approximately –4 SD and +2 SD, reflecting that the WEMWBS is informative for measuring a group of individuals with a wide range of WEMWBS levels. The marginal reliability coefficient of the scale was 0.92 (95% CI: 0.90, 0.93), indicating excellent overall measurement reliability across the latent trait continuum.

The WEMWBS can measure individuals with trait levels between approximately −4 SD and +3 SD with optimal reliability, that is, rxx ≥ 0.90. This means that the WEMWBS is most reliable for participants with low to high levels of MWB, but it becomes less reliable when measuring participants with extremely low (θ < −4 SD) or very high levels of MWB (θ > +3 SD). The empirical reliability of the ability estimates was 0.92, indicating excellent measurement precision within the sample of German-speaking undergraduate health professions students.



θ: Latent Trait, I: Test Information, SE: Standard Error, r_{xx}: Conditional Reliability

Figure 3. Test Information Curve and Conditional Reliability of the WEMWBS Scale

Discussion

This study aimed to evaluate the psychometric properties of the German version of the WEMWBS at both the scale and item levels by applying a combined approach using CTT and

IRT analyses in a sample of undergraduate health professions students in Switzerland. Overall, the students reported moderate levels of MWB. The IRT analyses showed moderate to strong item discrimination, with items 5, 6, 7, 8, 10, and 14 providing the most information. The scale predominantly assesses low to average levels of MWB, while still offering precise measurement across a wide range of the trait. These findings underscore its suitability for use among students with varying levels of MWB. In the CFA, the three-factor models with highly intercorrelated factors demonstrated a slightly better fit, particularly in terms of RMSEA, compared to both the bifactor model, which comprises a strong general factor and one specific group factor, and the one-factor models. Although the three-factor and bifactor models suggest some degree of multidimensionality, the high intercorrelations among the factors in the three-factor model, along with the high values of both ECV and omega hierarchical in the bifactor model, further support the dominance of the general factor. This indicates that, despite the multidimensional structure, the scale can be interpreted as essentially unidimensional for practical purposes (33). Similar evidence of moderate to good fit for a correlated three-factor model has been reported in previous studies, including validations of the Greek WEMWBS in Greek university students (22), the original WEMWBS in a Pakistani sample in the United Kingdom (36), the Finnish version in the general Finnish population (37), and the German WEMWBS in a German general population sample (27). However, for the German WEMWBS the correlated three-factor model did not show the best fit in earlier studies; instead, the bifactor model with a dominant general factor and three group factor showed slightly better fit indices in general population samples from Germany and Austria along with good internal consistencies: a McDonald's omega hierarchical of 0.81 in the Austrian sample (16) and a Cronbach's alpha of 0.95 in the German sample (27). Furthermore, the bifactor model was also supported in a study of adolescent school pupils in Ireland using the original WEMWBS (38). Although, some variability in model fit indices,

most validation studies involving diverse populations—including university and health professions students—have supported a unidimensional factor structure of the WEMWBS (15,19,23,24,26,39–41). Some studies reported suboptimal RMSEA values, but improved model fit by allowing correlated error terms (18,20,42). However, introducing such correlations can artificially enhance fit at the expense of theoretical clarity and replicability (43). In the current study, the CBFA produced estimation warnings, likely due to poorly defined group factors, each comprising only one or two items. This limitation prevented the inversion of the information matrix, thereby obstructing the computation of robust fit indices (e.g., RMSEA, CFI) and standard errors, including p-values for factor loadings. As a result, the interpretability of the model is limited, and it cannot be considered well-identified (44).

The findings of our IRT analyses were largely consistent with those reported by Marmara et al. (26). They also employed the GRM to evaluate the psychometric properties of the WEMWBS in a sample of English-speaking adults from the United States of America, United Kingdom, Ireland, Australia, New Zealand, and Canada. Both our study and the study by Marmara et al. (26) found also moderate to strong item discrimination, and items 3 and 5 emerged as the most difficult, indicating endorsement primarily by individuals with higher levels of MWB. In both samples, response category 4 (“often”) had the highest probability of endorsement across most levels of the latent trait. Additionally, items 8, 10, and 14 contributed the most information. However, differences emerged in the distribution of threshold (b) parameters. In contrast to our findings, Marmara et al. (26) reported negative values only for the first and second threshold parameters (b_1 and b_2). The predominantly negative thresholds observed in the present study suggest that several items were generally easy to endorse, even among respondents with low to average levels of MWB. This indicates that individuals did not require particularly high levels of MWB to agree with the positively framed WEMWBS items. This pattern is consistent with previous findings by Deng et al.

(45). Several factors may contribute to lower endorsement thresholds. The positive wording of the WEMWBS items may encourage higher agreement; social desirability may lead respondents to overreport their MWB, and cultural interpretations of MWB may make some statements appear more normative or broadly desirable. Regarding the mean WEMWBS score in our sample, it aligns with results from European university students (2,4,46), who generally show lower levels of MWB compared to the general population within European countries (16,47). Taken together, these psychometric results support the German WEMWBS as a valid and precise tool of MWB among Swiss undergraduate health professions students. Its confirmed unidimensionality simplifies practical use in educational settings. Institutions can rely on a single total score for routine monitoring students' MWB. Integrating the WEMWBS into MWB surveys enables longitudinal tracking, and scores can be interpreted using distribution-based- categories (±1 SD) as recommended by the WEMWBS developers (49). While these categories are not clinical thresholds, they can help institutions identify students whose MWB falls clearly below the typical range and may benefit from early support.

Limitations of the study

While this study provides important insights, several limitations must be considered. First, a sample size of n=290 participants might be insufficient for GRM, particularly for complex models with several parameters, such as bifactorial models (29). To ensure stable estimates a larger sample size (n=500) might be more appropriate (48). Second, the use of a sample drawn from a single health professions school and the gender imbalance (84% female participants) may limit the generalizability of the findings due to reduced heterogeneity. However, this limitation is partially mitigated by the inclusion of participants from two different study sites and multiple health professions. Third, the study did not assess measurement invariance, it remains uncertain whether the scale operates equivalently across relevant subgroups (e.g. gender). This limits the interpretability of group comparisons, as

differences in scale scores cannot be unequivocally interpreted as true differences in well-being. Finally, marginal and empirical reliability measures are specific to certain trait levels and do not provide a global or generalizable estimate of reliability.

Recommendation for further research

To enhance generalizability of the findings, future research should include more diverse samples from multiple health professions institutions and regions and test for measurement invariance across groups (e.g. gender, language or region). Moreover, validation in student populations from other German-speaking countries (e.g., Austria, Germany) is recommended to establish the scale's applicability beyond the Swiss sample. Furthermore, test-retest reliability, and responsiveness are recommended to test the stability of the German WEMWBS over time and its sensitivity to change. Finally, longitudinal studies should examine predictive validity, such as whether MWB scores predict academic outcomes or burnout risk.

Conclusions

The present study makes a valuable contribution to the field of higher education by validating the German version of the WEMWBS for use among undergraduate students in health professions. Through rigorous psychometric evaluation using both CTT and IRT, the findings support the scale's reliability and structural validity in this context as did other studies in other languages e.g., (26). The findings highlight the potential of the German WEMWBS as a practical tool for early identification of students at risk of low MWB. Integrating this scale into routine surveys can facilitate systematic monitoring of MWB, guide targeted initiatives, and ultimately strengthen student resilience and academic success. Its use enables educators, healthcare institutions, and researchers to implement timely support strategies and preventive interventions.

List of abbreviations

If abbreviations are used in the text they should be defined in the text at first use, and a list of abbreviations should be provided.

a	Difficulty parameter
b	Threshold parameter
α	Cronbach's alpha
CBFA	Confirmatory Bifactor Analysis
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit Index
χ2	Chi-square
CI	Confident Interval
CTT	Classical Test Theory
DIF	Differential Item Function
df	Degrees of freedom
EBFA	Exploratory Bifactor Analysis
ECV	Explained Common Variance
G	General factor
GRM	Graded Response Model
IRT	Item Response Theory
IIF	Item Information Function
IPF	Item Probability Function
KMO	Kaiser-Meyer-Olkin
ω	McDonald's omega
ωH	McDonald's omega hierarchical
ωHS	McDonald's omega hierarchical specific
ωtotal	McDonald's omega for the total score
MWB	Mental well-being
MSA	Measure of Sampling Adequacy
n	number
RMSEA	Root Mean Standard Error of Approximation
rxx	Marginal reliability
S	Specific group factor
SD	Standard Deviation
TIF	Test Information Function

SRMR	Standardized Root Mean Square Residual
WEMWBS	Warwick-Edinburgh Mental Well-being Scale
WLSMV	Weighted Least Squares Mean and Variance Adjusted

Declarations

Ethics approval and consent to participate

This study was conducted in accordance with the Declaration of Helsinki. The Cantonal Ethics Committee Bern conducted a clarification of responsibilities for this study (Req-2024-01645) and determined that no additional ethical approval procedures were required under the Swissethics guidelines. All participants provided written informed consent prior to their inclusion in the study.

Consent for publication

Not applicable

Availability of data and materials

The dataset generated and analysed during the current study, excluding any personal or identifiable information, is available in the OLOS repository, https://doi.org/10.34914/olos:qkab3rbfvbft3jfgliuxqecmru.

Competing interests

The authors declare that they have no competing interests.

Funding

This research received no external funding.

Authors' contributions

S. I.: Conceptualization, Methodology, Data Collection, Data Analysis, Writing - Original Draft
R. S.: Data Collection, Review and Editing
K. I.: Review and Editing
M.A.: Methodology and Data Analysis
Z.E.: Review and Editing
K. M.: Conceptualization, Review and Editing

Acknowledgements

Not applicable

References

1. Mladen S, Loughan A, Kinser P, Crawford M, Jones A, Edwards S, u. a. An Analysis of Psychological Distress Profiles and their Correlates in Interdisciplinary Health-care Professional Students. Glob Adv Health Med. 1. Januar 2019;8:2164956119879872.

2. Slimmen S, Timmermans O, Mikolajczak-Degrauwe K, Oenema A. How stress-related factors affect mental wellbeing of university students A cross-sectional study to explore the associations between stressors, perceived stress, and mental wellbeing. PLOS ONE. 7. November 2022;17(11):e0275925.

3. Cilar L, Barr O, Štiglic G, Pajnkihar M. Mental well-being among nursing students in Slovenia and Northern Ireland: A survey. Nurse Educ Pract. August 2019;39:130–5.

4. Cameron DM, Muratore F, Tower M, Eades CE, Evans JMM. Exploration of health and health behaviours of undergraduate nursing students: a multi-methods study in two countries. Contemp Nurse. 2. November 2022;58(5–6):473–83.

5. Nayak BS, Sahu PK. Socio-demographic and educational factors associated with Depression, Anxiety and Stress among Health Professions students. Psychol Health Med. 21. April 2022;27(4):848–53.

6. Heinrichs K, Lehnchen J, Burian J, Deptolla Z, Heumann E, Helmer S, u. a. Mental and physical well-being among students in Germany: results from the StudiBiFra study. J Public Health [Internet]. 23. September 2024 [zitiert 15. Januar 2025]; Verfügbar unter: https://doi.org/10.1007/s10389-024-02348-2

7. Stallman HM. Psychological distress in university students: A comparison with general population data. Aust Psychol. 1. Dezember 2010;45(4):249–57.

8. World Health Organization. Promoting mental health: Concepts, emerging evidence, practice. Geneva; 2005. 288 S.

9. Tennant R, Hiller L, Fishwick R, Platt S, Joseph S, Weich S, u. a. The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): development and UK validation. Health Qual Life Outcomes. 27. November 2007;5:63.

10. Franzen J, Jermann F, Ghisletta P, Rudaz S, Bondolfi G, Tran NT. Psychological Distress and Well-Being among Students of Health Disciplines: The Importance of Academic Satisfaction. Int J Environ Res Public Health. 23. Februar 2021;18(4):2151.

11. Wei H, Dorn A, Hutto H, Webb Corbett R, Haberstroh A, Larson K. Impacts of Nursing Student Burnout on Psychological Well-Being and Academic Achievement. J Nurs Educ. Juli 2021;60(7):369–76.

12. Brooker A, Vu C. How do University Experiences Contribute to Students' Psychological Wellbeing? Stud Success. 16. Oktober 2020;11(2):99–108.

13. Herbert C. Enhancing Mental Health, Well-Being and Active Lifestyles of University Students by Means of Physical Activity and Exercise Research Programs. Front Public Health. 25. April 2022;10:849093.

14. Stewart-Brown S, Taggart F, Parkinson J. Warwick-Edinburgh Mental Well-being Scale (WEMWBS) - User guide – Version 2 [Internet]. The University of Warwick, Warwick Medical School and NHS Health Scotland, Herausgeber. 2015. Verfügbar unter: https://warwick.ac.uk/fac/sci/med/research/platform/wemwbs/using/register/resources/wemwbs_user_guide_jp_02.02.16.pdf

15. Koushede V, Lasgaard M, Hinrichsen C, Meilstrup C, Nielsen L, Rayce SB, u. a. Measuring mental well-being in Denmark: Validation of the original and short version of the Warwick-Edinburgh mental well-being scale (WEMWBS and SWEMWBS) and cross-cultural comparison across four European settings. Psychiatry Res. Januar 2019;271:502–9.

16. Lang G, Bachinger A. Validation of the German Warwick-Edinburgh Mental Well-Being Scale (WEMWBS) in a community-based sample of adults in Austria: a bi-factor modelling approach. J Public Health. April 2017;25(2):135–46.

17. Ramadhana MR. Psychometric Evaluation of the Warwick-Edinburgh Mental Well-Being Scale on Indonesian Students. J -Nafs Kaji Penelit Psikol. 20. Dezember 2023;8(2):182–98.

18. Fung S fu. Psychometric evaluation of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS) with Chinese University Students. Health Qual Life Outcomes. Dezember 2019;17(1):46.

19. Konaszewski K, Niesiobędzka M, Surzykiewicz J. Factor structure and psychometric properties of a Polish adaptation of the Warwick–Edinburgh Mental Wellbeing Scale. Health Qual Life Outcomes. 2. März 2021;19(1):70.

20. Zayed K, Omara E, Al-Shamli A, Al-Rawahi N, Haramlah AA, Al-Attiyah AA, u. a. A validation study of the Arabic version of the Warwick-Edinburgh Mental Well-being scale among undergraduate students. BMC Psychol. 17. November 2023;11(1):399.

21. Lopez EKN, Johnson JH, Cunningham EP, Warshaw J, Baker EW. A unique and effective method of anatomy education: Small-group learning with prosected plastinated specimens. J Dent Educ. April 2022;86(4):425–36.

22. Petrogiannis K, Sangeorzan I, Andriopoulou P. Validation of the Greek Warwick-Edinburgh Mental Well-Being Scale (WEMWBS) in a Mature Student Community-Based Sample. Advers Resil Sci. Dezember 2024;5(4):351–68.

23. Cilar L, Pajnkihar M, Štiglic G. Validation of the Warwick-Edinburgh Mental Well-being Scale among nursing students in Slovenia. J Nurs Manag. 29. Juli 2020;jonm.13087.

24. Dong A, Chen X, Zhu L, Shi L, Cai Y, Shi B, u. a. Translation and validation of a Chinese version of the Warwick–Edinburgh Mental Well-being Scale with undergraduate nursing trainees. J Psychiatr Ment Health Nurs. November 2016;23(9–10):554–60.

25. Stewart-Brown S, Tennant A, Tennant R, Platt S, Parkinson J, Weich S. Internal construct validity of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS): a Rasch analysis using data from the Scottish Health Education Population Survey. Health Qual Life Outcomes. 19. Februar 2009;7(1):15.

26. Marmara J, Zarate D, Vassallo J, Patten R, Stavropoulos V. Warwick Edinburgh Mental Well-Being Scale (WEMWBS): measurement invariance across genders and item response theory examination. BMC Psychol. 18. Februar 2022;10(1):31.

27. Peitz D, Thom J, Walther L, Hoelling H, Cohrdes C. Validation of the Warwick-Edinburgh Mental Well-Being Scale for the Mental Health Surveillance (MHS) of German adults. Health Qual Life Outcomes. 26. Oktober 2024;22(1):92.

28. Knol DL, Mokkink LB, Terwee CB, de Vet HCW. Measurement in Medicine: A Practical Guide [Internet]. Cambridge: Cambridge University Press; 2011. (Practical Guides to Biostatistics and Epidemiology). Verfügbar unter: https://www.cambridge.org/core/product/8BD913A1DA0ECCBA951AC4C1F719BCC5

29. Reise SP, Yu J. Parameter Recovery in the Graded Response Model Using MULTILOG. J Educ Meas. 1990;27(2):133–44.

30. Embretson SE, Reise SP. Item response theory for psychologists. Mahwah, N.J: L. Erlbaum Associates; 2000. 371 S. (Multivariate applications book series).

31. Cheung GW, Cooper-Thomas HD, Lau RS, Wang LC. Reporting reliability, convergent and discriminant validity with structural equation modeling: A review and best-practice recommendations. Asia Pac J Manag. 1. Juni 2024;41(2):745–83.

32. Hu L tze, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Struct Equ Model. 1999;6(1):1–55.

33. Reise SP. The Rediscovery of Bifactor Measurement Models. Multivar Behav Res. 1. September 2012;47(5):667–96.

34. Rodriguez A, Reise-,Steven P., and Haviland MG. Applying Bifactor Statistical Indices in the Evaluation of Psychological Measures. J Pers Assess. 3. Mai 2016;98(3):223–37.

35. Zein RA, Akhtar H. Getting started with the graded response model: An introduction and tutorial in R. Int J Psychol. Februar 2025;60(1):e13265.

36. Taggart F, Friede T, Weich S, Clarke A, Johnson M, Stewart-Brown S. Cross cultural evaluation of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS) --a mixed methods study. Health Qual Life Outcomes. 27. Februar 2013;11:27.

37. Sarasjärvi KK, Elovainio M, Appelqvist-Schmidlechner K, Solin P, Tamminen N, Therman S. Exploring the structure and psychometric properties of the Warwick-Edinburgh Mental Well-Being Scale (WEMWBS) in a representative adult population sample. Psychiatry Res. Oktober 2023;328:115465.

38. Shannon S, Breslin G, Prentice G, Leavey G. Testing the factor structure of the Warwick-Edinburgh Mental Well-Being Scale in adolescents: A bi-factor modelling methodology. Psychiatry Res. November 2020;293:113393.

39. Bass M, Dawkin M, Muncer S, Vigurs S, Bostock J. Validation of Warwick-Edinburgh Mental Well-being Scale (WEMWBS) in a population of people using Secondary Care Mental Health Services. J Ment Health. 3. Juli 2016;25(4):323–9.

40. Lloyd K, Devine P. Psychometric properties of the Warwick–Edinburgh Mental Well-being Scale (WEMWBS) in Northern Ireland. J Ment Health. Juni 2012;21(3):257–63.

41. Al-Wattary N, Mollazehi M, Al-Jubouri AM, Al-Qahtani NA, Jolo LK, Badr AA, u. a. Validation and psychometric properties of the Arabic version of the Warwick–Edinburgh Mental Well-being Scale (WEMWBS) on a sample of students in Qatar. Ment Health Sci. Juni 2023;1(2):67–72.

42. Smith ORF, Alves DE, Knapstad M, Haug E, Aarø LE. Measuring mental well-being in Norway: validation of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS). BMC Psychiatry. 12. Mai 2017;17(1):182.

43. Hermida R. The Problem of Allowing Correlated Errors in Structural Equation Modeling: Concerns and Considerations. Comput Methods Soc Sci. 14. September 2015;3:1–17.

44. Pretorius TB, Padmanabhanunni A. Dimensionality in confirmatory factor analysis is not in the eye of the beholder: Ancillary bifactor statistical indices illuminate dimensionality and reliability. Int J Psychol. Februar 2025;60(1):e13266.

45. Deng W, Carpentier S, Blackwood J, Van de Winckel A. Rasch validation of the Warwick-Edinburgh Mental Well-Being Scale (WEMWBS) in community-dwelling adults. BMC Psychol. 17. Februar 2023;11(1):48.

46. Evans JMM, Andreis F, Cameron DM, Eades CE. How does the self-reported health of undergraduate nursing students change during their degree programme? Survey results from a Scottish University. BMC Nurs. 17. März 2021;20(1):44.

47. Yadav P, Torre JA de la, Bakolis I, Bartoll X, Kogel CC, Farran JC, u. a. Internal structure, reliability and cross-cultural validity of the Warwick-Edinburgh Mental Wellbeing Scale in three European populations. BMJ Ment Health [Internet]. 11. März

2025 [zitiert 15. Dezember 2025];28(1). Verfügbar unter: https://mentalhealth.bmj.com/content/28/1/e301433

48. Streiner DL, Norman GR, Cairney J. Health measurement scales : a practical guide to their development and use. Fifth edition. New York, New York: Oxford University Press; 2015.

49. Warwick Medical School. *Warwick-Edinburgh Mental Well-being Scale (WEMWBS): User guidance*. https://warwick.ac.uk/services/innovations/wemwbs/how/ Accessed 16. Dezember 2025.