

Ensuring the Validity of Assessing Student Well-Being in Paper-and-Pencil and Web-Based Formats: The Measurement Invariance of the MHC-SF

Journal of Psychoeducational Assessment
2026, Vol. 0(0) 1–16
© The Author(s) 2026



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/07342829261423745
journals.sagepub.com/home/jpa



Norman B. Mendoza¹ , Jet Uy Buenconsejo² , Shen Ba¹ , and Zi Yan¹ 

Abstract

The increasing use of digital assessment tools in education necessitates rigorous psychometric evaluation of to support their equivalence across administration formats. This study investigated the structural validity and measurement invariance of the Mental Health Continuum-Short Form (MHC-SF), a widely used measure of student well-being, across paper-and-pencil and web-based formats. Participants included 1,297 secondary school students (59.75% female; $M_{\text{age}} = 14.57$, $SD = 1.23$) from public schools in the Philippines, who answered the MHC-SF administered in English. Using confirmatory factor analysis, we examined competing measurement models and found superior fit for the theoretically-derived three-factor structure (emotional, social, and psychological well-being) compared to alternative models. Multigroup invariance testing demonstrated measurement invariance between paper-and-pencil ($n = 663$) and web-based ($n = 634$) administration modes, with invariant factor structure (configural), factor loadings (metric), intercepts (scalar), and residual variances (strict). The findings corroborate the internal validity of the MHC-SF and demonstrates its utility for assessing adolescent students' well-being across different administration formats. This study provides empirical support on the viability of the MHC-SF as a web-based tool for assessing well-being, which is crucial for timely intervention. Critically, the equivalence of the MHC-SF across administration methods has important implications for the growing integration of technology in education, especially in resource-constrained contexts.

¹Department of Curriculum and Instruction, Faculty of Human Development, The Education University of Hong Kong, Hong Kong SAR, China

²Department of Applied Social Sciences, Faculty of Health and Social Sciences, The Hong Kong Polytechnic University, Kowloon, Hong Kong SAR

Corresponding Author:

Norman B. Mendoza, The Education University of Hong Kong, 10 Lo Ping Road, Tai Po, New Territories, Hong Kong SAR, China.

Email: nbmendoza@eduhk.hk

Keywords

mental health continuum-short form, measurement invariance, paper-and-pencil, web-based, assessment, well-being, adolescents, Philippines

The increasing integration of technology in educational settings has opened up new possibilities for assessing student well-being and learning outcomes. Online administration of psychological instruments offers potential advantages in terms of efficiency, accessibility, and data management compared to traditional paper-and-pencil methods (Davies et al., 2017; Howell et al., 2010; Lazarova et al., 2021; Van de Looij-Jansen & De Wilde, 2008). However, as assessments transition to digital formats, it is crucial to rigorously examine how different modes of administration may impact the psychometric properties of instruments and measures, as well as the interpretation of scores (Cai et al., 2008; Fouladi et al., 2002; Kim et al., 2022; Meng et al., 2023; Vecchione et al., 2012).

Establishing measurement invariance is essential to ensure that an instrument measures the same constructs in the same way across different groups or conditions, such as paper-and-pencil vs. online administration (Chen, 2007; Maassen et al., 2023; Vecchione et al., 2012). Without evidence of invariance, observed differences between groups or formats could be due to measurement bias rather than true differences in the underlying constructs of interest (Maassen et al., 2023; Millsap, 2011; Vecchione et al., 2012). As educational institutions increasingly rely on technology-based tools to support both student learning and well-being, demonstrating the psychometric equivalence of measures across traditional and virtual contexts is a fundamental imperative (Davies et al., 2017; Kim et al., 2022; Lazarova et al., 2021; Wasil et al., 2020). This study examines the factorial validity and measurement invariance of the Mental Health Continuum-Short Form (MHC-SF; Keyes, 2002), as a measure of positive well-being, in paper-and-pencil and web-based formats among secondary school students in the Philippines.

The Mental Health Continuum-Short Form (MHC-SF)

The Mental Health Continuum-Short Form (MHC-SF; Keyes, 2002) is a widely used instrument that assesses emotional, social, and psychological dimensions of well-being. It has demonstrated robust psychometric properties across diverse cultural contexts. In China, Guo et al. (2015) validated the scale among 5,399 adolescents (ages 12–18), finding excellent fit for the three-factor model (CFI = .936, RMSEA = .086, SRMR = 0.057) with evidence for internal consistency ($\alpha = .89-.94$). Their findings supported using MHC-SF scores for identifying flourishing and languishing youth in school-based screening programs. Lupano Perugini et al. (2017) established measurement invariance across gender and age groups in Argentina ($N = 1,300$ adults), demonstrating that MHC-SF scores can validly support cross-group comparisons for research and intervention evaluation. Karaś et al. (2014) validated the Polish adaptation among 2,115 individuals, establishing scalar invariance across gender and educational groups. Žemojtel-Piotrowska et al.'s (2018) 38-nation study ($N = 8,066$) provided the most comprehensive cross-cultural evidence, demonstrating configural and metric invariance across all countries, indicating that the three-factor structure and factor loadings remain consistent globally. These validation studies collectively support the MHC-SF for: (1) screening and identifying students requiring mental health support, (2) evaluating intervention effectiveness through pre-post comparisons, (3) conducting cross-cultural and developmental research, and (4) monitoring population-level well-being trends.

The MHC-SF is grounded on the dual-continua model of mental health (Keyes, 2005), which posits that positive mental health is not merely the absence of psychopathology but a distinct dimension encompassing emotional, psychological, and social well-being. This model has substantial psychoeducational relevance. MHC-SF has been shown to distinguish between flourishing and languishing students (Keyes, 2007) and can determine profiles of students who may be at risk for academic difficulties based on their scores in each dimension (Suldo & Shaffer, 2008). MHC-SF profiles have also been shown to predict student engagement and GPA (Antaramian et al., 2010), which has significant implications for educators in addressing needs of students across all levels of well-being (Chan et al., 2022). This multidimensional perspective on well-being has gained considerable empirical support and has been influential in guiding research and intervention efforts aimed at promoting positive mental health (Keyes, 2007; Lamers et al., 2015; Petrillo et al., 2015). These findings suggest that capturing emotional, social, and psychological dimensions provides valuable information for educational intervention and support beyond what mental health symptom screening alone can offer.

Two important observations arise from the existing literature. First, the MHC-SF has been implemented in both formats and research has shown evidence of its reliability and validity, and research supports its theoretical foundations and practical applications (e.g., Bernardo et al., 2022; Mendoza et al., 2023; Mendoza & Yan, 2023); however, no published studies have directly examined its measurement invariance across paper-and-pencil and online administration formats (Lamborn et al., 2018; Luijten et al., 2019; Lupano Perugini et al., 2017). Given the increasing reliance on online administration methods, particularly during the COVID-19 pandemic (e.g., Kim et al., 2022; Meng et al., 2023), establishing the equivalence of the MHC-SF across these modes is critical to ensure the validity of research findings and the appropriateness of intervention decisions based on the instrument (see Cockerham et al., 2021).

Second, while the MHC-SF has been validated in various cultural contexts, there is a paucity of research examining its psychometric properties in lower middle-income countries such as the Philippines. The Philippines, with its large population and growing mental health concerns (Lally et al., 2019; Perlas et al., 2022), represents an important context for investigating the applicability and invariance of the MHC-SF. Establishing its measurement equivalence across administration modalities in this setting would not only contribute to the MHC-SF's cross-cultural validation but also provide valuable insights into the assessment of well-being in resource-limited contexts.

Paper-and-Pencil versus Online Administration and the MHC-SF

Previous investigations of well-being and mental health instruments such as the Satisfaction With Life Scale, Positive and Negative Affect Schedule, Subjective Happiness Scale, and Strengths and Difficulties Questionnaire have shown that the online administration mode can yield similar results to paper-based administration (Van de Looij-Jansen & De Wilde, 2008; Howell et al., 2010). However, other studies have noted potential differences between these two modes of administration. For example, Fouladi et al. (2002) and McCoy et al. (2004) found that online surveys can introduce biases and may affect stability of the participants' responses, although the magnitude of these effects was generally small. Relatedly, Crawley et al. (2000) noted that online administration of quality-of-life instruments was preferred by both patients and research coordinators as it was found to be more comfortable, efficient, and easier to use.

To date, no published study has directly examined these modes of administration for the MHC-SF, despite its validation across several countries (Żemojtel-Piotrowska et al., 2018) and groups (e.g., Lamborn et al., 2018; Luijten et al., 2019; Lupano Perugini et al., 2017). Given the MHC-SF's multidimensional structure and its grounding in the dual-continua model of mental health (Keyes, 2005), it is essential to establish the invariance of the scale's factor structure and psychometric

properties across paper-and-pencil and online administration methods. This would provide evidence for the robustness of the underlying theoretical framework and support the valid use of the MHC-SF in various research and educational settings, regardless of the assessment medium.

The Present Study

Building upon the existing literature and addressing the identified research gaps, the present study aims to evaluate the psychometric properties and measurement invariance of the English version of the MHC-SF across paper-and-pencil and online administration methods. We hypothesize that the three-factor structure of emotional, social, and psychological well-being will demonstrate a good fit to the data and show invariance across the two administration modes. Testing this hypothesis will provide evidence for the valid use of the MHC-SF in online formats, with implications for educational research and practice in diverse contexts, particularly in lower middle-income countries like the Philippines.

This study has three main objectives. First, by focusing on a sample of secondary school students, we aim to shed light on the MHC-SF's utility in assessing well-being among adolescents, a population facing unique mental health challenges (Rapee et al., 2019; Salmela-Aro & Tuominen-Soini, 2010). Establishing measurement invariance across administration methods in this age group would support the MHC-SF's use in school-based well-being screening and intervention programs that increasingly incorporate online platforms (Dowdy et al., 2015; Solmi et al., 2022). Second, by conducting the study in the Philippines, we seek to expand the cross-cultural applicability of the MHC-SF and its utility in resource-limited settings. The findings may inform well-being assessment and intervention efforts in contexts with limited mental health infrastructure and resources (Lally et al., 2019; Perlas et al., 2022). Finally, we aim to provide evidence for the robustness of the dual-continua model of mental health (Keyes, 2005) by examining the invariance of the MHC-SF's factor structure and psychometric properties across administration methods. This addresses the need for research on the MHC-SF's factorial validity across traditional and web-based modes of administration, particularly in the context of a lower middle-income country facing the challenges posed by the COVID-19 pandemic. By addressing these gaps, this study contributes to the understanding of the MHC-SF, its use in diverse contexts and modalities, its implications for educational research and practice.

Methods

Participants and Procedures

Selection and Recruitment. Data collection occurred during two time periods, with paper-and-pencil administration in January 2020 and online administration in May 2020. In the Philippines, no educational disruptions due to COVID-19 cases have been enforced in January 2020. Nationwide closures began in mid-March, and by May 2020, schools had already transitioned to remote learning. Rather than viewing this as a limitation, we consider it an ecologically valid test of measurement equivalence under naturally occurring conditions. This temporal difference provides an opportunity to examine measurement equivalence under substantially different contexts, testing whether the MHC-SF maintains psychometric integrity across both format and contextual changes.

Schools were selected through stratified random sampling from the Department of Education's (DepEd) regional database of public secondary schools, stratified by size and urban/rural location. Classrooms within each school were randomly selected for participation. All participating public secondary schools served predominantly lower-income communities. Individual student

socioeconomic status was documented through receipt of government assistance: 68.4% of students across both samples ($n = 887$ of 1,297) were recipients of the Pantawid Pamilyang Pilipino Program (4Ps), the Philippines' national conditional cash-transfer program targeting households classified as poor or near-poor based on program eligibility criteria tied to the national poverty line (approximately USD215 monthly household income at the time of data collection). Specifically, 67.9% of the paper-and-pencil sample ($n = 450$ of 663) and 68.9% of the online sample ($n = 437$ of 634) received government assistance.

Inclusion criteria for participants were: (1) enrollment in Grades 7–10 and (2) sufficient English comprehension as verified by teachers. For participants under 18 years old, written parental consent was obtained via consent forms sent home with students 1 week prior to data collection. Students provided written assent on the day of assessment. For 18-year-old participants, direct informed consent was obtained. No students were excluded based on academic performance, socioeconomic status, or other characteristics.

For paper-and-pencil administration, 796 of 1,200 invited students from one large urban public secondary school provided consent (66.3% response rate). For online administration, 675 of 850 students invited across five public secondary schools within the same district consented to participate (79.4% response rate). The higher online response rate likely reflects different recruitment contexts (single school vs. multiple schools) rather than format preference. Analysis of available administrative data indicated no significant differences between participants and non-participants in age (available for all invited students) or grade level distribution ($\chi^2 = 2.34$, $p = .504$).

Administration Procedures. The MHC-SF was administered in English. All participants were enrolled in English-medium instruction schools where all subjects are taught in English. Participants' English proficiency was evidenced by successful completion of at least 6 years of English-medium instruction, English language grades of "satisfactory" or above in the previous academic year, and teacher verification of adequate comprehension skills for survey completion.

Paper-and-pencil administration occurred in classroom settings during regular school hours with 25–35 students per session (see Mendoza & Yan, 2021). Trained research assistants provided standardized verbal instructions following a scripted protocol. They responded to student questions but did not assist with item interpretation. No time limits were imposed; completion typically required 12–18 minutes. Online administration was conducted using the Qualtrics platform. Students answered the survey in school computer laboratories during designated class periods. Technical support staff were available for platform-related issues (e.g., login problems) but did not assist with survey content. Completion time ranged from 10 to 15 minutes.

Both formats included: (1) welcome screen/page with study overview, (2) informed consent/assent process, (3) demographic questions (age, gender, grade level), (4) MHC-SF items in fixed order, and (5) closing acknowledgment. Item wording, response scales, and sequencing were identical across formats, although the digital format enabled forced responses to minimize missing data. All consent materials described the study purpose, procedures, voluntary participation, confidentiality protections, and available mental health referral resources. The study was approved by the Human Research Ethics Committee of the first author's affiliated university (Ref. no.: 2019-2020-0152).

Participants. The paper-and-pencil dataset initially contained 796 responses. Missing data occurred primarily due to incomplete surveys (e.g., skipped items). Analysis revealed that 94 respondents (11.8%) had missing data exceeding 5% of MHC-SF items and were excluded from analysis to maintain data quality. The remaining 702 cases had missing data on 1–4 items (representing 0.5%–3.6% of responses per case). We tested whether data were missing completely

at random (MCAR) using Little's MCAR test ($\chi^2 = 87.34$, $df = 94$, $p = .665$), confirming that missingness was not systematically related to observed variables. This supports the use of multiple imputation procedures.

Multiple imputation with chained equations (MICE; van Buuren & Groothuis-Oudshoorn, 2011) was performed using the mice package in R (version 3.14.0), generating 20 imputed datasets. The imputation model included all 14 MHC-SF items, including demographic variables (age, gender, grade level) as auxiliary variables. Predictive mean matching was used for continuous variables. Results were pooled across imputations following Rubin's rules. After outlier removal using the Mahalanobis distance rule (see Bedrick et al., 2000), the final paper-and-pencil sample consisted of 663 cases.

The online dataset ($n = 675$) contained no missing data due to forced-response format in the Qualtrics platform, which required responses to all items before proceeding. After outlier removal ($n = 41$), the final online sample consisted of 634 cases.

The final analytical dataset consisted of 663 from the paper-and-pencil administration and 634 for the online administration ($N = 1,297$; 59.75% females). The mean age of the respondents was 14.57 years old ($SD = 1.23$). Students in Grade 10 comprised the largest percentage of participants ($n = 309$; 23.59%), followed by students from Grade 9 ($n = 275$; 20.99%).

Instruments

The Mental Health Continuum-Short Form (MHC-SF; Keyes, 2002; e.g., Mendoza & Yan, 2023) is a 14-item measure of general well-being, with items pertaining to emotional (3 items; e.g., "satisfied with life"), social well-being (5 items; e.g., "that you had something important to contribute to society"), and psychological (6 items; e.g., "that your life has a sense of direction or meaning to it"). Respondents rated the items from 1 (*never*) to 6 (*everyday*). The Cronbach's alpha for emotional, social, and psychological well-being subscales for this study were $\alpha = 0.82$, $\alpha = 0.87$, and $\alpha = 0.89$, respectively. The Cronbach's alpha of the full MHC-SF is $\alpha = 0.94$.

Data Analysis

To determine the best-fitting factor structure of the MHC-SF, we conducted confirmatory factor analysis (CFA) using the lavaan package (Rosseel, 2012). We tested three factor structures: (1) a unidimensional model included all items in one general latent factor with all 13 MHC-SF items, (2) a two-factor model with emotional (3 items) and eudaemonic well-being (11 items) as latent variables, and (3) a three-factor model with emotional (3 items), social (5 items), and psychological (6 items) well-being as latent variables. In the CFA, we employed the Weighted Least Squares with Mean and Variance (WLSMV) adjusted estimator for all CFA models (e.g., Valdez & Mendoza, 2024). This estimator has demonstrated advantages over both Weighted Least Squares and Maximum Likelihood (WLSML) methods (Flora & Curran, 2004). We assessed CFA models using the Satorra-Bentler scaled test statistic and compared models using the chi-square tests ($SB\chi^2$). To determine goodness-of-fit, we used several fit indices, including the Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), and standardized root mean square residual (SRMR). Following conventional guidelines, models with CFI and TLI values exceeding 0.90 and RMSEA and SRMR values below 0.08 were considered to have good fit to the data (Hu & Bentler, 1999).

Consequently, to investigate the measurement invariance of the best-fitting model, we conducted multigroup CFA using the equaltestMI package in R (Jiang & Mai, 2020). The equivalence testing required dividing the sample into two groups based on the administration procedures: paper-and-pencil and web-based administration. We first established the initial multigroup CFA model for each group, allowing all factor loadings, uniqueness, and residuals to be freely estimated

across groups (e.g., Mendoza et al., 2024; Mendoza & Yan, 2025). Subsequently, we performed configural, metric, scalar, and strict invariance tests by progressively constraining the factor structure, factor loadings, intercepts, and errors, respectively. To determine measurement invariance, we followed Chen's (2007) recommendations for samples greater than 300: a change of CFI (Δ CFI) that is less than or equal to .01, supplemented by a change of RMSEA (Δ RMSEA) that is less than or equal to .15 or a change in SRMR (Δ SRMR) that is less than or equal to .03 will indicate invariance. All analyses were conducted in R (R Core Team, 2016).

Results

Comparing the Factor Structures of the MHC-SF

Table 1 presents the descriptive statistics and reliability coefficients of the MHC-SF and its three dimensions. Table 2 presents the results of the CFA and chi-square difference test of the three competing models. All measurement models showed adequate model fit, with increasing model fit as the number of factors increased. The three-factor model (see Figure 1) exhibited the highest CFI and TLI values and lowest RMSEA and SRMR values. The superiority of the three-factor model is also supported by the chi-square difference tests. Using robust likelihood ratio tests, the data suggests that the two-factor model provided a significantly better fit than the unidimensional model. However, the three-factor model showed a significantly better fit to the data compared to the other models. For the three-factor model, item factor loadings for emotional, social, and psychological well-being factors ranged from .68 to .84, .71 to .79, and .72 to .81, respectively.

Administration Invariance of the MHC-SF Through Multigroup CFA

We used the best-fitting model of the MHC-SF (i.e., three-factor model) to test the measurement invariance between two scale administration procedures: paper-and-pencil and online. Preliminary analyses confirmed no significant differences between administration groups on demographic or socioeconomic characteristics: age ($t = -1.43$, $p = .154$), gender distribution ($\chi^2 = 0.13$, $p = .716$), grade level distribution ($\chi^2 = 3.21$, $p = .523$), or government assistance receipt ($\chi^2 = 0.84$, $p = .359$). These findings support the interpretation that observed measurement equivalence reflects format robustness rather than sample differences.

The three-factor model of the MHC-SF exhibited measurement invariance across administration procedures (see Table 3). Specifically, between paper-and-pencil ($n = 663$) and online ($n = 634$) administration, the configural model which constrains the three-factor structure across groups had excellent fit indices. The model fit of the metric invariance which constrains all factor loadings

Table 1. Overall bivariate correlations, descriptive, and internal consistencies

	1	2	3	4
1. MHC-SF total	(0.92)			
2. Emotional WB	.83***	(0.77)		
3. Social WB	.88***	.63***	(0.80)	
4. Psychological WB	.93***	.70***	.69***	(0.89)
Mean	54.01	11.79	17.79	24.42
Standard deviation	13.99	3.43	5.54	6.74
Skewness	-.02	-.09	.20	-.22
Kurtosis	2.32	2.16	2.35	2.28

Table 2. Construct validation comparing fit indices of the models

MHC factor structure	$SB\chi^2$	df	CFI	TLI	RMSEA	SRMR	χ^2 difference vs hypothesized model
Model 1: 1-factor MHC	739.946	77	0.991	0.992	0.044	0.043	---
Model 2: 2-factor MHC	540.218	76	0.993	0.995	0.037	0.036	$\chi^2_{diff}(2) = 3354.7, p < 0.001$
Model 3: 3-factor MHC	346.400	74	0.997	0.996	0.028	0.028	$\chi^2_{diff}(3) = 4757.5, p < 0.001$

across groups had no meaningful decrement from that of the fit indices of the configural model. Similarly, the fit indices of the scalar invariance constraining all item intercepts were well within the invariance thresholds in comparison to the metric model. Finally, the strict invariance tests which constrain errors likewise did not have a significant decrement in its fit indices compared to the scalar invariance. Taken altogether, the results provide evidence of measurement invariance for the three-factor MHC-SF despite increasing levels of model constraints.

Post-hoc Analysis of Contextual Effects

Following the establishment of strict measurement invariance, we conducted post-hoc analyses to examine whether the temporal context affected mean well-being levels (see [Supplemental Table 2](#)). This analysis addresses a distinct question from measurement invariance: while invariance examines whether the measurement model functions equivalently, mean comparisons test whether actual well-being levels differed between contexts. Mean comparisons revealed significant differences across all MHC-SF scales with medium effect sizes (Cohen, 1988).

Item-level descriptive statistics ([Supplemental Table 1](#)) show that students in May 2020 reported lower means across all 14 items. These patterns are consistent with documented pandemic impacts on adolescent mental health, including increased anxiety, social isolation, and disrupted routines during early COVID-19 school closures (e.g., Cockerham et al., 2021; Lee, 2020). However, we acknowledge that mean differences may also reflect unmeasured school-level

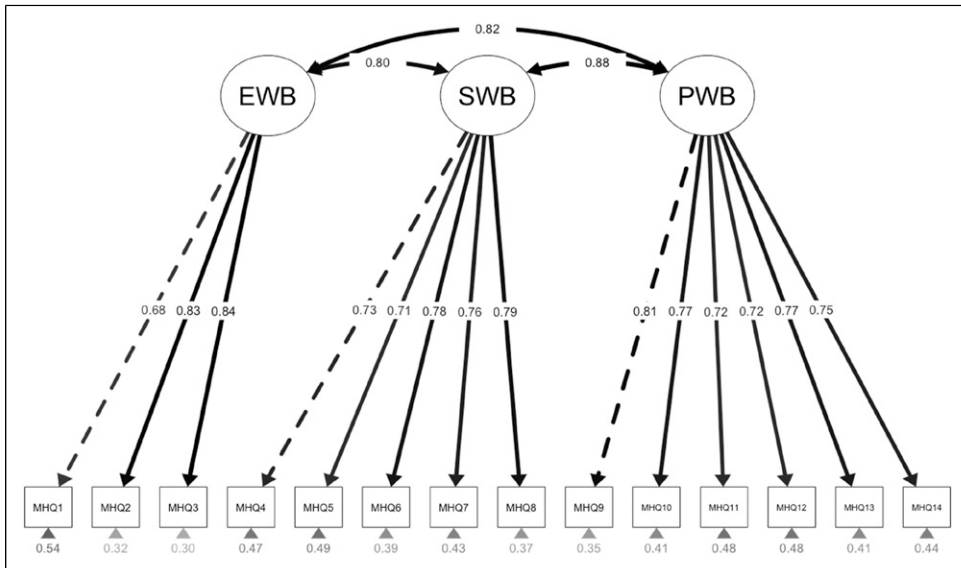


Figure 1. The three-factor structure of the MHC-SF

Table 3. Multigroup CFA for measurement invariance of the 3-factor model across two administration methods

Model constraints	SB χ^2	df	CFI	TLI	RMSEA	SRMR	Δ CFI	Δ RMSEA	Δ SRMR	Invariant
3-factor model	346.40	74	0.997	0.996	0.028	0.028	---	---	---	---
Across administration										
Paper-and-pencil only (<i>n</i> = 663)	207.451	74	0.994	0.995	0.028	0.035	---	---	---	---
Online only (<i>n</i> = 634)	189.273	74	0.997	0.998	0.027	0.027	---	---	---	---
Configural invariance	396.429	148	0.996	0.997	0.028	0.031	0.000	0.000	−0.003	Invariant
Metric invariance	336.902	159	0.994	0.995	0.033	0.040	0.002	−0.005	−0.009	Invariant
Scalar invariance	661.814	170	0.985	0.986	0.053	0.054	0.009	−0.020	−0.014	Invariant
Strict invariance	704.75	184	0.985	0.985	0.054	0.058	0.000	−0.001	−0.004	Invariant

differences in student well-being, as paper–pencil administration occurred in one school while online administration occurred across five different schools within the same district.

Discussion

This study evaluated the psychometric properties and measurement invariance of the Mental Health Continuum-Short Form (MHC-SF; [Keyes, 2002](#)) administered through paper-and-pencil and online methods among a large sample of Filipino secondary school students. Consistent with previous literature ([Franken et al., 2018](#); [Guo et al., 2015](#); [Karaš et al., 2014](#); [Lamers et al., 2011](#)), the findings supported the three-factor structure of the MHC-SF comprising emotional, social, and psychological well-being dimensions. Additionally, the three-factor model exhibited significantly better fit compared to one-factor and two-factor models, showing evidence of construct validity.

The three-factor structure’s robustness across paper-and-pencil and web-based administration procedures suggests that the MHC-SF’s conceptualization of well-being as a multidimensional construct ([Keyes, 2002](#)) holds true regardless of the assessment medium. This finding aligns with the dual-continua model of mental health ([Keyes, 2005](#)), which posits that positive mental health is not merely the absence of psychopathology but a distinct dimension encompassing emotional, psychological, and social well-being. The invariance of the three-factor structure across administration methods supports the notion that these well-being components are fundamental to the conceptual framework underlying the MHC-SF and are not influenced by measurement administration.

Notably, we found evidence of strict measurement invariance between paper-and-pencil and online administration, indicating comparable reliability and validity of scores across these modes. This finding suggests the MHC-SF can be validly applied through various formats, supporting its use in adapted contexts necessitated by changes in educational practices and environments ([Kim et al., 2022](#); [Meng et al., 2023](#)). Our results address the gap noted by [Lamborn et al. \(2018\)](#), [Luijten et al. \(2019\)](#), and [Lupano Perugini et al. \(2017\)](#) by providing the first evidence of the scale’s measurement invariance when transitioning between traditional and virtual administration formats.

Our design presents both a limitation and a strength. We cannot isolate format effects from temporal effects, as data collection occurred during different contexts (pre-COVID vs. during COVID). Post-hoc analyses revealed substantial mean-level differences with medium effect sizes

($d = 0.43\text{--}0.63$) across all well-being dimensions. However, this addresses an important practical question: Does the MHC-SF maintain measurement equivalence when administered under substantially different contextual conditions? Strict measurement invariance, despite these substantial differences provides particularly robust evidence. Future research employing counter-balanced designs would complement our findings by isolating format effects under controlled conditions.

As educational institutions increasingly rely on technology-based tools to support student learning and well-being, demonstrating the psychometric equivalence of measures across paper-and-pencil and online contexts is a fundamental concern (Davies et al., 2017; Kim et al., 2022; Lazarova et al., 2021; Wasil et al., 2020). Our results suggest the MHC-SF can be used interchangeably between these modes without introducing measurement bias, consistent with some previous studies on the online adaptation of well-being scales (Howell et al., 2010; Van de Looij-Jansen & De Wilde, 2008). Our study extends this research by establishing strict invariance for the multidimensional MHC-SF and sampling students from an under-represented, lower-middle-income country.

Implications

The study's findings have important theoretical and practical implications. From a theoretical perspective, the invariance of the MHC-SF's three-factor structure across paper-and-pencil and web-based administration supports the robustness and generalizability of the dual-continua model (Keyes, 2005) and the multidimensional nature of well-being. We demonstrated strict measurement invariance of the MHC-SF across paper-and-pencil and online administration formats. This finding addresses a critical gap in literature and has three important implications. First, the MHC-SF functions equivalently across formats in that observed scores reflect the same underlying constructs regardless of administration method. Second, educators and researchers can confidently transition between formats or compare scores obtained via different methods without introducing measurement bias. Third, this evidence supports flexible, resource-adaptive assessment strategies where schools can select administration formats based on available infrastructure rather than psychometric concerns.

The robustness of our invariance findings is particularly noteworthy given the substantial contextual differences between samples. Post-hoc analyses revealed that students assessed online in May 2020 (during COVID school closures, from five schools) reported significantly lower well-being than students assessed via paper-pencil in January 2020 (pre-COVID, from one school), with medium effect sizes across all subscales ($d = 0.43\text{--}0.63$). These differences may reflect temporal effects related to the pandemic, school-level differences in student well-being, or a combination of both factors. The demonstration of strict measurement invariance despite these meaningful differences (which may stem from temporal, school-level, or combined effects) provides particularly robust evidence for the scale's psychometric stability by functioning equivalently even when comparing groups that differ substantially in their actual well-being levels.

Practically, the establishment of strict measurement invariance for the MHC-SF across administration methods has significant implications for educational and psychological research and practice. The findings indicate that the MHC-SF can be confidently used in online formats without sacrificing psychometric integrity, which is particularly valuable in light of the increasing reliance on digital tools in educational and mental health settings (Davies et al., 2017; Lazarova et al., 2021; Wasil et al., 2020). This flexibility allows for the valid assessment of student well-being across various contexts and facilitates the transition to remote learning and psychological service delivery when necessary, as evidenced during the COVID-19 pandemic (Kim et al., 2022; Meng

et al., 2023). Moreover, the results, along with previous research highlighting the impact of the pandemic on adolescent well-being and social interactions (Cockerham et al., 2021), underscore the importance of identifying and addressing student needs in online education to foster positive mental health.

These findings extend beyond well-being assessment to inform the broader transition toward digital educational measurement. As schools increasingly implement computer-based testing for achievement and aptitude assessment, establishing measurement equivalence across formats becomes paramount. Our methodological approach (i.e., multigroup confirmatory factor analysis testing configural through strict invariance) provides a framework applicable to other educational measures. The demonstration that a multidimensional psychological instrument maintains psychometric integrity across formats suggests similar robustness may be achievable for cognitive and achievement measures, though empirical verification remains necessary for each instrument. Educational psychologists and assessment specialists should prioritize invariance testing when transitioning established measures to digital formats, ensuring that technological modernization does not compromise measurement validity.

Our findings also illustrate an important methodological principle: measurement invariance testing and mean comparison serve distinct but complementary purposes. Invariance testing examines whether the measurement model (factor structure, loadings, intercepts, and residuals) functions equivalently across groups, while mean comparisons examine whether groups differ on the construct being measured. These are separate questions. The common misconception that groups must have similar means for invariance testing to be valid or meaningful would paradoxically eliminate the primary purpose of invariance methodology, that is, enabling valid cross-group comparisons when groups are expected to differ. Our study demonstrates this whereby we established measurement equivalence first, which then provided the psychometric foundation to confidently interpret substantial mean differences as reflecting genuine contextual effects (whether temporal, school-level, or combined) rather than measurement bias introduced by format or assessment conditions.

Lastly, the study's focus on Filipino secondary school students addresses the need for culturally-sensitive mental health assessment tools in lower middle-income countries. The MHC-SF's psychometric robustness in this context highlights its potential as a valuable instrument for assessing well-being among adolescents in diverse cultural settings, particularly those that may have limited access to mental health resources. The findings can inform the development and evaluation of interventions aimed at promoting positive mental health and well-being among students in the Philippines and other similar contexts.

Limitations and Future Research Directions

While this study provides validity evidence and measurement invariance of the MHC-SF across administration methods, some limitations should be considered.

First, our design confounds administration format with both temporal context and school-level factors, precluding definitive attribution of mean differences to any single source. Paper-pencil administration occurred in one school in January 2020 (pre-COVID), while online administration occurred across five schools in May 2020 (during COVID). Although observed mean differences align with expected pandemic impacts on adolescent well-being, they may also reflect unmeasured school-level differences or the interaction of multiple factors. Importantly, the demonstration of strict measurement invariance remains valid regardless of the source of mean differences, confirming that the MHC-SF functions equivalently across these varied conditions. This represents a particularly stringent test of measurement stability under naturally occurring, real-world conditions.

Demonstrating strict invariance with meaningful mean differences illustrates the fundamental purpose of invariance testing. Without establishing equivalence of the measurement model, we could not determine whether lower well-being scores in May 2020 reflected actual decreases in well-being versus measurement artifacts introduced by format changes, temporal instability, or their interaction. Our findings confirm that observed differences reflect genuine changes during the pandemic transition. While controlled experimental designs isolating format effects would complement our findings, the co-occurrence of format and substantial contextual differences provides ecologically valid evidence for measurement robustness under real-world implementation conditions where multiple factors inevitably vary.

Additional limitations include reliance on self-report data, which may be subject to response biases. Multi-informant designs could provide more comprehensive assessment. The cross-sectional nature limits examination of within-person changes over time. Longitudinal designs would allow investigation of measurement stability and predictive validity across administration modes over multiple time points. Experimental and mixed-methods designs enable a more nuanced understanding of the factors influencing well-being and inform development of targeted interventions. Qualitative approaches, such as interviews or focus groups, could provide valuable insights into students' experiences and perceptions of well-being, complementing the quantitative findings from the MHC-SF. Examining MHC-SF's psychometric properties across various administration methods such as technology-based formats and digital platforms would further support the scale's adaptability and validity in the rapidly evolving landscape of psychological service delivery. While our sample represents resource-limited contexts in the Philippines, replication in other lower middle-income countries and diverse cultural contexts would strengthen evidence for cross-cultural applicability. Finally, although our focus was on adolescent well-being, examining invariance across different age groups (children, adults) would provide evidence for the MHC-SF's developmental applicability across the lifespan.

Conclusion

Our findings advance the psychometric literature on well-being assessment in educational contexts by establishing the MHC-SF's measurement invariance across administration formats among Filipino secondary school students. While previous validations demonstrated the instrument's utility across cultural contexts (Franken et al., 2018; Guo et al., 2015; Karás et al., 2014), this study contributes to psychoeducational assessment research through specific ways. First, it provides empirical evidence supporting flexible use of the MHC-SF across paper-and-pencil and web-based formats, addressing a key methodological concern in flexible and resource-adaptive assessment implementation. Second, it offers validity evidence from a non-Western educational system, supporting cross-cultural applicability while highlighting context-sensitive measurement considerations. Finally, the findings inform the feasibility of sustainable school-based mental health monitoring that accommodate shifts in assessment modality without compromising measurement precision. These findings suggest that schools can implement hybrid well-being assessment approaches, which are particularly valuable for developing evidence-based well-being support systems across diverse educational contexts. Future research should examine the MHC-SF's utility within integrated school-based assessment programs, specifically focusing on early identification and intervention planning applications.

Acknowledgment

We would like to thank Ms. Jelli Grace C. Luzano, LifeRisksPH, and Dr. Ericson Cabrera for their support in data collection and feedback of this paper.

ORCID iDs

Norman B. Mendoza  <https://orcid.org/0000-0003-0344-0709>

Jet Uy Buenconsejo  <https://orcid.org/0000-0003-3777-8601>

Zi Yan  <https://orcid.org/0000-0001-9305-884X>

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data Availability Statement

All pertinent data, materials, and codes for analysis will be made available by the authors upon written request by any party.

Supplemental Material

Supplemental material for this article is available online.

References

- Antaramian, S. P., Huebner, E. S., Hills, K. J., & Valois, R. F. (2010). A dual-factor model of mental health: Toward a more comprehensive understanding of youth functioning. *American Journal of Orthopsychiatry*, *80*(4), 462–472. <https://doi.org/10.1111/j.1939-0025.2010.01049.x>
- Bedrick, E. J., Lapidus, J., & Powell, J. F. (2000). Estimating the mahalanobis distance from mixed continuous and discrete data. *Biometrics*, *56*(2), 394–401. <https://doi.org/10.1111/j.0006-341X.2000.00394.x>
- Bernardo, A. B. I., Mendoza, N. B., Simon, P. D., Cunanan, A. L. P., Dizon, J. I. W. T., Tarroja, M. C. H., Ma Balajadia-Alcala, A., & Saplala, J. E. G. (2022). Coronavirus pandemic anxiety scale (CPAS-11): Development and initial validation. *Current Psychology*, *41*(8), 5703–5711. <https://doi.org/10.1007/s12144-020-01193-2>
- Cai, H., Lin, Y., Wu, Q., Yan, L., & Huang, X. (2008). Examining the measurement invariance between paper-and-pencil and internet-administered tests in China. *Acta Psychologica Sinica*, *40*(2), 228–239. <https://doi.org/10.3724/SP.J.1041.2008.00228>
- Chan, M. K., Furlong, M. J., Nylund-Gibson, K., & Dowdy, E. (2022). Heterogeneity among moderate mental health students on the mental health continuum-short form (MHC-SF). *School Mental Health*, *14*(2), 416–430. <https://doi.org/10.1007/s12310-021-09476-0>
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*(3), 464–504. <https://doi.org/10.1080/10705510701301834>
- Cockerham, D., Lin, L., Ndolo, S., & Schwartz, M. (2021). Voices of the students: Adolescent well-being and social interactions during the emergent shift to online learning environments. *Education and Information Technologies*, *26*(6), 7523–7541. <https://doi.org/10.1007/s10639-021-10601-4>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Crawley, J., Kleinman, L., & Dornitz, J. (2000). User preferences for computer administration of quality-of-life instruments. *Drug Information Journal*, *34*(1), 137–144. <https://doi.org/10.1177/009286150003400119>
- Davies, E. B., Craven, M. P., Martin, J. L., & Simons, L. (2017). Proportionate methods for evaluating a simple digital mental health tool. *Evidence-Based Mental Health*, *20*(4), 112–117. <https://doi.org/10.1136/eb-2017-102755>

- Dowdy, E., Furlong, M., Raines, T. C., Boverly, B., Kauffman, B., Kamphaus, R. W., Dever, B. V., Price, M., & Murdock, J. (2015). Enhancing school-based mental health services with a preventive and promotive approach to universal screening for complete mental health. *Journal of Educational and Psychological Consultation*, 25(2-3), 178–197. <https://doi.org/10.1080/10474412.2014.929951>
- Flora, D. B., & Curran, P. J. (2004). An Empirical Evaluation of Alternative Methods of Estimation for Confirmatory Factor Analysis With Ordinal Data. *Psychological Methods*, 9(4), 466–491. <https://doi.org/10.1037/1082-989X.9.4.466>
- Fouladi, R. T., McCarthy, C. J., & Moller, N. P. (2002). Paper-and-pencil or online? Evaluating mode effects on measures of emotional functioning and attachment. *Assessment*, 9(2), 204–215. <https://doi.org/10.1177/10791102009002011>
- Franken, K., Lamers, S. M. A., Ten Klooster, P. M., Bohlmeijer, E. T., & Westerhof, G. J. (2018). Validation of the mental health continuum-short form and the dual continua model of well-being and psychopathology in an adult mental health setting. *Journal of Clinical Psychology*, 74(12), 2187–2202. <https://doi.org/10.1002/jclp.22659>
- Guo, C., Tomson, G., Guo, J., Li, X., Keller, C., & Söderqvist, F. (2015). Psychometric evaluation of the mental health continuum-short form (MHC-SF) in Chinese adolescents – a methodological study. *Health and Quality of Life Outcomes*, 13, 198. <https://doi.org/10.1186/s12955-015-0394-2>
- Howell, R. T., Rodzon, K. S., Kurai, M., & Sanchez, A. H. (2010). A validation of well-being and happiness surveys for administration via the internet. *Behavior Research Methods*, 42(3), 775–784. <https://doi.org/10.3758/BRM.42.3.775>
- Hu, L. T., & Bentler, P. M., analysis: Conventional criteria versus new alternatives. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Jiang, G., & Mai, Y. (2020). equaltestMI: Examine measurement invariance via equivalence testing and projection method. *R Package Version 0.6.1*. <https://CRAN.R-project.org/package=equaltestMI>
- Karaś, D., Ciecuch, J., & Keyes, C. L. M. (2014). The polish adaptation of the mental health continuum-short form (MHC-SF). *Personality and Individual Differences*, 69, 104–109. <https://doi.org/10.1016/j.paid.2014.05.011>
- Keyes, C. L. M. (2002). The mental health continuum: From languishing to flourishing in life. *Journal of Health and Social Behavior*, 43(2), 207–222. <https://doi.org/10.2307/3090197>
- Keyes, C. L. M. (2005). Mental illness and/or mental health? Investigating axioms of the complete state model of health. *Journal of Consulting and Clinical Psychology*, 73(3), 539–548. <https://doi.org/10.1037/0022-006X.73.3.539>
- Keyes, C. L. M. (2007). Promoting and protecting mental health as flourishing: A complementary strategy for improving national mental health. *American Psychologist*, 62(2), 95–108. <https://doi.org/10.1037/0003-066X.62.2.95>
- Kim, J., Moon, K., Lee, J., Jeong, Y., Lee, S., & Ko, Y. G. (2022). Online learning performance and engagement during the COVID-19 pandemic: Application of the dual-continua model of mental health. *Frontiers in Psychology*, 13, Article 932777. <https://doi.org/10.3389/fpsyg.2022.932777>
- Lally, J., Tully, J., & Samaniego, R. (2019). Mental health services in the Philippines. *BJPsych International*, 16(3), 62–64. <https://doi.org/10.1192/bji.2018.34>
- Lamborn, P., Cramer, K. M., & Riberdy, A. (2018). The structural validity and measurement invariance of the mental health continuum – Short form (MHC-SF) in a large Canadian sample. *Journal of Well-Being Assessment*, 2(1), 1–19. <https://doi.org/10.1007/s41543-018-0007-z>
- Lamers, S. M. A., Westerhof, G. J., Bohlmeijer, E. T., ten Klooster, P. M., & Keyes, C. L. M. (2011). Evaluating the psychometric properties of the mental health continuum-short form (MHC-SF). *Journal of Clinical Psychology*, 67(1), 99–110. <https://doi.org/10.1002/jclp.20741>

- Lamers, S. M. A., Westerhof, G. J., Glas, C. A. W., & Bohlmeijer, E. T. (2015). The bidirectional relation between positive mental health and psychopathology in a longitudinal representative panel study. *The Journal of Positive Psychology, 10*(6), 553–560. <https://doi.org/10.1080/17439760.2015.1015156>
- Lazarova, E., Mora, S., Armanino, D., Poire, A., Furlani, F., & Giacomini, M. (2021). A web-based tool to support a personalized therapeutic path through the use of psychological tests. *Studies in Health Technology and Informatics, 285*, 211–216. <https://doi.org/10.3233/SHTI210600>
- Lee, J. (2020). Mental health effects of school closures during COVID-19. *The Lancet Child & Adolescent Health, 4*(6), 421. [https://doi.org/10.1016/S2352-4642\(20\)30109-7](https://doi.org/10.1016/S2352-4642(20)30109-7)
- Luijten, C. C., Kuppens, S., van de Bongardt, D., & Nieboer, A. P. (2019). Evaluating the psychometric properties of the mental health continuum-short form (MHC-SF) in Dutch adolescents. *Health and Quality of Life Outcomes, 17*(1), 157. <https://doi.org/10.1186/s12955-019-1221-y>
- Lupano Perugini, M. L., de la Iglesia, G., Castro Solano, A., & Keyes, C. L. M. (2017). The mental health continuum-short form (MHC-SF) in the Argentinean context: Confirmatory factor analysis and measurement invariance. *Europe's Journal of Psychology, 13*(1), 93–108. <https://doi.org/10.5964/ejop.v13i1.1163>
- Maassen, E., D'Urso, E. D., van Assen, M. A. L. M., Nuijten, M. B., De Roover, K., & Wicherts, J. M. (2023). The dire disregard of measurement invariance testing in psychological science. *Psychological Methods*. Advance online publication. <https://doi.org/10.1037/met0000624>
- McCoy, S., Marks, P. V., Carr, C. L., & Mbarika, V. W. (2004). Electronic versus paper surveys: Analysis of potential psychometric biases. *Proceedings of the 37th Annual Hawaii International Conference on System Sciences, 8*, 8. <https://doi.org/10.1109/HICSS.2004.1265634>
- Mendoza, N. B., Frondoza, C. E., Dizon, J. I. W. T., & Buenconsejo, J. U. (2024). The factor structure and measurement invariance of the PHQ-4 and the prevalence of depression and anxiety in a Southeast Asian context amid the COVID-19 pandemic. *Current Psychology, 43*, 22019–22028. <https://doi.org/10.1007/s12144-022-02833-5>
- Mendoza, N. B., King, R. B., & Haw, J. Y. (2023). The mental health and well-being of students and teachers during the COVID-19 pandemic: Combining classical statistics and machine learning approaches. *Educational Psychology, 43*(5), 430–451. <https://doi.org/10.1080/01443410.2023.2226846>
- Mendoza, Norman B., & Yan, Zi (2021). Validation of a Subject-Specific Self-assessment Practices Scale (SaPS) Among Secondary School Students in the Philippines. *Journal of Psychoeducational Assessment, 39*(4), 481–493. <https://doi.org/10.1177/0734282921994374>
- Mendoza, N. B., & Yan, Z. (2023). Exploring the moderating role of well-being on the adaptive link between self-assessment practices and learning achievement. *Studies in Educational Evaluation, 77*, Article 101249. <https://doi.org/10.1016/j.stueduc.2023.101249>
- Mendoza, N. B., & Yan, Z. (2025). From beliefs to behaviors: Conceptualizing and assessing students' practices that reflect a growth mindset. *Social Psychology of Education, 28*(1), 73. <https://doi.org/10.1007/s11218-025-10032-w>
- Meng, W., Yu, L., Liu, C., Pan, N., Pang, X., & Zhu, Y. (2023). A systematic review of the effectiveness of online learning in higher education during the COVID-19 pandemic period. *Frontiers in Education, 8*, Article 1334153. <https://doi.org/10.3389/educ.2023.1334153>
- Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. Routledge.
- Perlas, A. P., Ruiz, M. A. B., Perlas, C. J. S., & de Guzman, A. B. (2022). The Filipino college students' mental health during the COVID-19 pandemic: A phenomenological inquiry. *Journal of Loss and Trauma, 27*(3), 205–218. <https://doi.org/10.1080/15325024.2021.1886966>
- Petrillo, G., Capone, V., Caso, D., & Keyes, C. L. M. (2015). The mental health continuum-short form (MHC-SF) as a measure of well-being in the Italian context. *Social Indicators Research, 121*(1), 291–312. <https://doi.org/10.1007/s11205-014-0629-3>
- Rapee, R. M., Oar, E. L., Johnco, C. J., Forbes, M. K., Fardouly, J., Magson, N. R., & Richardson, C. E. (2019). Adolescent development and risk for the onset of social-emotional disorders: A review and

- conceptual model. *Behaviour Research and Therapy*, 123, Article 103501. <https://doi.org/10.1016/j.brat.2019.103501>
- R Core Team. (2016). R: A language and environment for statistical computing [computer software]. <https://www.R-project.org/>
- Rosseel, Yves (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2), 1–36. <https://doi.org/10.18637/jss.v048.i02>
- Salmela-Aro, K., & Tuominen-Soini, H. (2010). Adolescents' life satisfaction during the transition to post-comprehensive education: Antecedents and consequences. *Journal of Happiness Studies*, 11(6), 683–701. <https://doi.org/10.1007/s10902-009-9156-3>
- Solmi, M., Radua, J., Olivola, M., Croce, E., Soardo, L., Salazar de Pablo, G., Il Shin, J., Kirkbride, J. B., Jones, P., Kim, J. H., Kim, J. Y., Carvalho, A. F., Seeman, M. V., Correll, C. U., & Fusar-Poli, P. (2022). Age at onset of mental disorders worldwide: Large-scale meta-analysis of 192 epidemiological studies. *Molecular Psychiatry*, 27(1), 281–295. <https://doi.org/10.1038/s41380-021-01161-7>
- Suldo, S. M., & Shaffer, E. J. (2008). Looking beyond psychopathology: The dual-factor model of mental health in youth. *School Psychology Review*, 37(1), 52–68. <https://doi.org/10.1080/02796015.2008.12087908>
- Valdez, J. P. M., & Mendoza, N. B. (2024). Digital learning for preschools: Validation of basic ICT competence beliefs of preschool teachers in Hong Kong and the Philippines. *Education and Information Technologies*, 29(14), 1–15. <https://doi.org/10.1007/s10639-024-12591-5>
- van Buuren, S., & Groothuis-Oudshoorn, K. (2011). Mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, 45(3), 1–67. <https://doi.org/10.18637/jss.v045.i03>
- Van de Looij-Jansen, P. M., & De Wilde, E. J. (2008). Comparison of web-based versus paper-and-pencil self-administered questionnaire: Effects on health indicators in Dutch adolescents. *Health Services Research*, 43(5 Pt 1), 1708–1721. <https://doi.org/10.1111/j.1475-6773.2008.00860.x>
- Vecchione, M., Alessandri, G., & Barbaranelli, C. (2012). Paper-and-pencil and web-based testing: The measurement invariance of the Big Five personality tests in applied settings. *Assessment*, 19(2), 243–246. <https://doi.org/10.1177/1073191111419091>
- Wasil, A. R., Gillespie, S., Patel, R., Petre, A., Venturo-Conerly, K. E., Shingleton, R. M., Weisz, J. R., & DeRubeis, R. J. (2020). Reassessing evidence-based content in popular smartphone apps for depression and anxiety: Developing and applying user-adjusted analyses. *Journal of Consulting and Clinical Psychology*, 88(11), 983–993. <https://doi.org/10.1037/ccp0000604>
- Žemojtel-Piotrowska, M., Piotrowski, J. P., Osin, E. N., Ciecuch, J., Adams, B. G., Ardi, R., Bălătescu, S., Bogomaz, S., Bhomi, A. L., Clinton, A., de Clunie, G. T., Czarna, A. Z., Esteves, C., Gouveia, V., Halik, M. H. J., Hosseini, A., Khachatryan, N., Kamble, S. V., Kawula, A., ... Maltby, J. (2018). The mental health continuum-short form: The structure and application for cross-cultural studies-A 38 nation study. *Journal of Clinical Psychology*, 74(6), 1034–1052. <https://doi.org/10.1002/jclp.22570>