Rasch-Based Evaluation of Rating Scale Functioning in a PERMA Flourishing Instrument for University Students

[1]Nadia Aulia Nadhirah, [2]Faraz Maulana Gunawan, [3]Nur'aini Hasna Hamidah, [4]Sudaryat Nurdin Akhmad, [5]Syamsu Yusuf, [6]M. Solehuddin, [7]Ilfiandra
[1,2,3,4,5,6,7]Universitas Pendidikan Indonesia, Bandung
[1]nadia.aulia,nadhirah@upi.edu, [2]maulanafaraz@upi.edu, [3]nuraini.hasna18@upi.edu,
[4]sunurakhmad@upi.edu, [5]syamsu@upi.edu, [6]msolehuddin@upi.edu, [7]ilfiandra@upi.edu

Abstract

The assessment of psychological well-being is becoming increasingly important in the context of higher education to support the development of students' potential. This study aims to develop and test the validity and reliability of a rating scale based on the PERMA model (Positive Emotion, Engagement, Relationships, Meaning, Accomplishment) in measuring flourishing or positive functioning among students. The research employs a quantitative survey method involving 305 students from various study programs at Universitas Pendidikan Indonesia. Data were collected using the PERMA model flourishing instrument. Data analysis was conducted using the Rasch Model, focusing on Rating Scale Measurement. The results indicate that the original 0-10 flourishing instrument scale was simplified to a 0-5 scale, with the rating scale functioning effectively. These findings support the use of the PERMA model flourishing instrument as a valid and reliable measurement tool in the context of higher education. This research has implications for the design of intervention strategies by educators and counselors to enhance student well-being.

Keywords: *Flourishing scale, PERMA, Rasch, Rating scale, University student.*

INTRODUCTION

The concept of flourishing has become a central focus in the study of psychological well-being, particularly in higher education, which is often regarded as a critical period in individual development. Flourishing, as described by Keyes (2002), is a state of optimal well-being in which individuals are not only free from illness or psychological issues but are also able to achieve their full potential emotionally, socially, and psychologically. In the context of higher education, flourishing is increasingly important as students face a variety of unique challenges. The university experience represents a critical transition to adulthood, during which students are required to manage academic, social, and personal responsibilities simultaneously (Stallman, 2010; Duffy et al., 2016). Today's students are confronted with demands that extend beyond academic achievement; they are also expected to develop 21st-century life skills, such as critical thinking, creativity, collaboration, and digital literacy (Greenhow & Lewin, 2016). In this context, flourishing serves as a vital foundation to support students in overcoming these challenges and reaching their full potential.

Student well-being in higher education cannot be measured solely by a single dimension, such as academic success or emotional happiness. Students face a variety of interconnected academic, social, and personal challenges, necessitating a holistic approach to comprehensively understand their well-being. In this context, the PERMA model,

introduced by Martin Seligman, serves as a relevant framework for measuring multidimensional well-being (Seligman, 2011). Higher education is a complex environment where students encounter academic, social, and personal challenges. The PERMA model provides a relevant measurement tool to evaluate how each dimension of flourishing contributes to overall student success (Kern et al., 2015). This model allows for a deeper assessment of student experiences, helping them identify areas for improvement and develop strategies to enhance their well-being (Hone et al., 2015).

The multidimensional measurement of student well-being using the PERMA model holds significant urgency in supporting a more holistic higher education process. Higher education is not merely an arena for honing academic skills but also an environment that shapes students' character, social skills, and emotional readiness. Therefore, the PERMA model, which measures flourishing through five key dimensions: positive emotion, engagement, relationships, meaning, and accomplishment, provides a structured and relevant framework for understanding student well-being in depth. The dimension of positive emotion refers to an individual's ability to experience feelings of happiness, while the engagement dimension describes the extent to which individuals are involved in activities that provide satisfaction. The relationships dimension emphasizes the importance of positive social connections, the meaning dimension pertains to the search for and understanding of life purposes that provide meaning and direction, and the accomplishment dimension relates to the ability to achieve personal or professional goals that create a sense of pride and self-satisfaction ((Ryan & Deci, 2000; Fredrickson, 2001; Steger et al., 2006). The PERMA model, with its five dimensions, offers a more comprehensive framework compared to previous well-being measurement models, which tended to focus on one or two aspects, such as life satisfaction or the absence of psychological problems (Ryff & Keyes, 1995). This multidimensional approach ensures that individual well-being is not measured from a single perspective but encompasses various factors that influence life as a whole.

The PERMA model developed by Seligman (2011) has spurred research to create measurement tools that can comprehensively evaluate multidimensional well-being. One of the most well-known instruments is the PERMA-Profiler, developed by Butler dan Kern (2016). Research by Kern et al. (2015) demonstrated that the PERMA-Profiler has strong internal consistency, with alpha coefficients exceeding 0.70 for each of its dimensions, making it one of the reliable tools for measuring multidimensional well-being. In addition to the PERMA-Profiler, other scales that adapt the PERMA model have also begun to be developed for specific population needs. For instance, Khaw dan Kern (2014) adapted the PERMA model to measure well-being in the Malaysian cultural context, demonstrating that this model is flexible enough to be applied across various cultural backgrounds. Despite the effectiveness of PERMA-based scales, the complexity and length of the items that need to be answered often pose barriers for students to participate in measuring their well-being.

Therefore, the need for a simple, valid, and reliable scale is crucial to enhance participation and the accuracy of well-being measurements.

Although the PERMA model has been widely recognized as an effective framework for measuring flourishing and multidimensional well-being, there are several gaps in the research that need to be addressed to enhance the understanding and application of this scale in the context of higher education. These gaps include aspects such as the validity and reliability of the scale, cultural context appropriateness, and limitations in practical application within educational environments. Many studies have developed PERMA-based scales; however, not all these scales have been thoroughly tested for their validity and reliability. Some studies indicate that while the scales can measure the dimensions of PERMA, there are concerns regarding the consistency of results obtained from various student populations (Hone et al., 2015). Most research using the PERMA scale originates from Western cultural contexts, which may not fully reflect the experiences of students in other countries or cultures. Well-being and flourishing can be influenced by different cultural factors, making it essential to adapt and test the PERMA scale in diverse cultural contexts (Diener et al., 2010). Research by Khaw dan Kern (2014) shows that while the PERMA framework can be applied in cross-cultural contexts, some dimensions require adaptation for local relevance. However, studies on cultural adaptation remain limited, particularly in the context of developing countries or collectivist cultures like Indonesia, where social relationships are prioritized. Many studies tend to focus on one or two dimensions of the PERMA model, such as positive emotions or achievement, without considering the interactions among all dimensions. This can lead to an incomplete understanding of how various aspects of well-being are interconnected and influence one another (Seligman, 2011).

The development and validation of a PERMA-based scale using a five-point scale for students is an important step in understanding and measuring psychological well-being among students. By providing a valid and reliable tool to measure the dimensions of PERMA, this research aims to offer a useful instrument for understanding and enhancing students' psychological well-being. Evaluating the PERMA dimensions as indicators of psychological well-being will provide deeper insights into how each aspect contributes to a positive and holistic learning experience. It is hoped that this will encourage the development of better policies and practices to support student well-being in higher education institutions.

METHOD

The research employs a positivist paradigm with a quantitative approach. The research design is a cross-sectional survey, which provides information in a short period, and describes the attitudes, opinions, behaviors, or characteristics of a population (Creswell, 2013).

Participants

The research population consists of students from the education program at Universitas Pendidikan Indonesia, distributed across 9 faculties, with an age range of 18 to 25 years. The sample size is 305 individuals, comprising 224 females and 81 males, selected using random sampling to ensure that each individual has an equal chance of being chosen from the population, thereby ensuring that the sample represents the population. The characteristics of the sample are listed in Table 1.

Table 1. Summary of the sample characteristics (N = 305)

		Total			
		N = 305	**Valid %**	**Male (%)**	**Female (%)**
Sex	**Male**	81	26.56		
	Female	224	73.44		
Faculty	**Faculty of Educational Sciences**	169	55.41	46.91	58.48
	Faculty of Language and Literature Education	4	1.31	1.23	1.34
	Faculty of Economics and Business Education	20	6.56	7.41	6.25
	Faculty of Social Science education	14	4.59	1.23	5.80
	Faculty of Mathematics and Natural Sciences Education	5	1.64	2.47	1.34
	Faculty of Sport and Health Education	22	7.21	13.58	4.91
	Faculty of Art and Design Education	8	2.62	3.70	2.23
	Faculty of Engineering and Industrial Education	24	7.87	16.05	4.91
	Regional Campus	39	12.79	7.41	14.73
Semester	**6**	150	49.18	15.33	84.67
	8	148	48.52	25.68	74.32
	10	7	2.30	71.43	28.57

Materials

Data collection was conducted using a research instrument designed to measure the flourishing of education students. The PERMA model, developed by Martin Seligman, with its five core components, serves as a relevant tool for measuring the flourishing of education students. These five components contribute to a holistic assessment of student well-being and help to understand how they develop and reach their maximum potential.

The flourishing scales. The construction of the flourishing instrument refers to Seligman (2011) theory, which measures flourishing among education students through five dimensions: Positive Emotion, Engagement, Relationships, Meaning, and Accomplishment, abbreviated as PERMA. Each dimension plays a crucial role in assessing the flourishing of education students to understand how they develop and reach their potential. According to the PERMA model, the ability to flourish is influenced by five dimensions: (1) Positive Emotion refers to the experience of feelings of happiness, pleasure, gratitude, and optimism; (2) Engagement represents the experience of students feeling fully absorbed in the activities

they undertake, both in learning and teaching; (3) Relationships pertain to the quality and quantity of positive social connections that students have, including support from friends, family, lecturers, and the educational community; (4) Meaning involves the experience of students feeling that what they do has purpose and contributes meaningfully to their own lives and the lives of others; and (5) Accomplishment refers to the feelings of success and competence in achieving academic and personal goals. The instrument is developed based on the categories of activities undertaken by education students, with 19 activities that represent the conditions of PERMA.

Scale of positive and negative experiences. The flourishing instrument employs a Likert scale model. Each item is rated on a 6-point scale, ranging from 0 (never) to 5 (always) or alternatively from 0 (not at all) to 5 (completely). The scores of three items from each domain are averaged to produce a single domain score that ranges from 0 to 5 (with higher scores indicating better well-being). The use of this scale is based on the development of the PERMA flourishing scale in various countries (Lambert D'raven & Pasha-Zaidi, 2016; Didino et al., 2019; Wammerl et al., 2019; Romano et al., 2020; Mahamid et al., 2023).

Data Analysis

Psychological attributes are determined using Rasch modeling through the Winsteps Version 5.7.3 application, taking into account the rating scale values. Rating Scale Measurement (RSM) in Rasch is used to analyze various types of analyses, such as fit statistics, unidimensional diagnosis, and differential item functioning (Bond & Fox, 2015). However, this study only employs RSM to demonstrate the diagnostic function of the rating scale in the PERMA flourishing instrument. RSM is suitable for assessing the difficulty level of items on a rating scale with two or more ordered response categories (Smith et al., 2003). RSM is implemented in the Winsteps software, which is then used to conduct data analysis in this study. The Winsteps software can provide empirical evidence regarding respondents' abilities to understand and differentiate between response categories, as well as inform the quality of the rating scale (Linacre, 2002).

The effectiveness of the rating scale categories is diagnosed using rating scale functions such as observed count, observed category, outfit MNSQ, and Andrich Threshold, which are analyzed meticulously. To achieve the desired results Linacre (2011) has provided seven criteria as guidelines to assist researchers in conducting their analyses. However, as suggested by Linacre (2002), not all criteria are suitable for the analysis of specific rating scales. Therefore, Smith et al. (2003) recommend five criteria for rating scale conditions: (1) each category must have a minimum of ten observations, (2) the observed average should develop monotonically with the categories, (3) the outfit MNSQ value should be less than 2.0, (4) the Andrich Threshold should develop monotonically with the categories, and (5) the Andrich Threshold should develop with a minimum of 1.4 logits and a maximum of 5.0 logits.

These criteria are used as guidelines to analyze and optimize the functioning of the rating scale categories in the study.

RESULTS AND DISCUSSION

The analysis of the rating scale function is assessed based on the output from Table 3.2+ rating (partial credit) scale generated by Winsteps version 5.7.3. The flourishing instrument demonstrates high reliability, with a person reliability of 0.87 and an item reliability of 0.99. The analysis of the rating scale begins by investigating the category structure of the instrument, as shown in Figure 1.

TABLE 3.2 The PERMA Profile Calon Guru.xls ZOU531WS.TXT Jan 10 2025 09:49
INPUT: 306 PERSON 55 ITEM REPORTED: 305 PERSON 55 ITEM 11 CATS WINSTEPS 5.7.3.0

SUMMARY OF CATEGORY STRUCTURE. Model="R"

CATEGORY LABEL	SCORE	OBSERVED COUNT	%	OBSVD AVRGE	SAMPLE EXPECT	INFIT MNSQ	OUTFIT MNSQ	ANDRICH THRESHOLD	CATEGORY MEASURE	
0	0	598	4	-.17	-.30	1.46	1.81	NONE	(-1.82)	0
1	1	471	3	-.18*	-.23	1.15	1.26	-.03	-.98	1
2	2	675	4	-.13	-.16	1.10	1.13	-.56	-.63	2
3	3	908	5	-.10	-.07	.87	.84	-.41	-.42	3
4	4	867	5	-.04	.01	.80	.76	.01	-.25	4
5	5	1940	12	.06	.10	.84	.78	-.75	-.09	5
6	6	1673	10	.16	.19	.80	.73	.29	.08	6
7	7	2761	16	.28	.29	.86	.79	-.26	.30	7
8	8	3021	18	.41	.39	.90	.90	.25	.62	8
9	9	1904	11	.52	.51	1.06	1.05	.91	1.15	9
10	10	1957	12	.67	.64	1.15	1.06	.55	(2.23)	10

Figure 1. Summary of category structure of flourishing instrument

The results shown in Figure 1 are based on five criteria according to Smith et al. (2003), Linacre (2002), and Elfida et al. (2021), for optimizing the function of the rating scale categories, with the following explanations:

Minimum Observed Count of 10 per Category: The observed count for categories 0-10 ranges from 471 to 3021, with the lowest value in category 1 and the highest in category 8. The frequency of values in the observed count is sufficient to estimate a stable rating scale structure.

Observed Average Develops Monotonically According to Category: The observed average does not increase sequentially and consistently (-0.17 > -0.18* < -0.13 < -0.10 < -0.04 < 0.06 < 0.16 < 0.28 < 0.41 < 0.52 < 0.67). There is a problematic observation in category 1, which is greater than the previous category, but categories 2-10 increase sequentially. This result generally indicates that participants with lower abilities tend to support higher response options, while participants with higher abilities support lower response options. Thus, the second criterion is not met.

Outfit MNSQ Greater than 2.0: The outfit MNSQ values range from 0.73 to 1.81 logits, with the criterion met being less than 2 logits, indicating that the criterion is satisfied as the data set introduces more information with lower unexplained noise.

Andrich Threshold Develops Monotonically According to Category: The Andrich threshold is not met because the values do not develop monotonically according to the categories (NONE < -0.03 > -0.56* < -0.41 < 0.01 > -0.75* > 0.29 > -0.26* < 0.25 < 0.91 > 0.55*). The value for category 2 (-0.56) is greater than that for category 1 (-0.03), category 5 (-0.75) is less than category 4 (0.01), category 7 (-0.26) is less than category 6 (0.29), and category 10 (0.55) is less than category 9 (0.91), indicating that these four categories lead to threshold disturbances. The disturbance thresholds can be graphically observed from the category probability response curve in Figure 2, which shows that the peak of the curve for category 10 is absent, and categories 2, 5, and 7 do not reach their peaks. The x-axis in Figure 2 represents the participants' ability relative to the item difficulty level, while the y-axis represents the observed probability of the categories in logit values.

Andrich Threshold Values Range from 1.4 to 5.0 Logits: The range of Andrich threshold values in categories 0-10 does not meet the criterion as it does not exceed 1.4 logits.

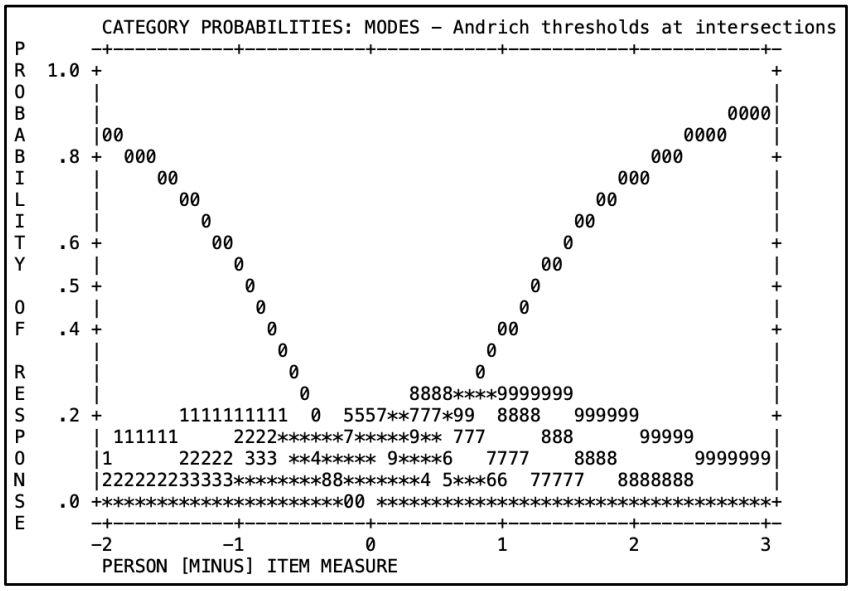

Figure 2. Category probabilities of flourishing instrument

Figure 2 shows the category probabilities of the 0-10 Likert scale in the flourishing instrument, indicating that the detected scale ranges from 0 to 9, with probabilities ranging from 0 to 1, which demonstrates the certainty that a particular category will be chosen. The horizontal axis represents the interaction between students' abilities and item difficulty, with values ranging from -2 to 3, indicating the range of students' abilities relative to item difficulty. Based on the curve, it is observed that a higher curve at certain points indicates that the category is more likely to be chosen at that level of ability or item difficulty, while overlapping curves indicate transitions between response categories.

Based on Figures 1 and 2, there is empirical evidence that the rating scale categories of the flourishing instrument do not function effectively. Therefore, the solution implemented for

the problematic categories is to simplify them to the nearest categories to enhance the meaning and function of the rating scale [25]. From the analysis of the rating scale, it was found that categories 2, 5, 7, and 10 are problematic scales, resulting in the Andrich threshold values not being ordered from negative to positive. Consequently, a recoding was performed, where scale 0 remains 0, scales 1 and 2 become 1, scales 3 and 4 become 2, scales 5 and 6 become 3, scales 7 and 8 become 4, and scales 9 and 10 become 5. Thus, the final decision from the ranking scale analysis is to use the categorization "01122334455." A new scale ranging from 0 to 5 was obtained, and a second analysis was conducted to test the accuracy of the scale. Figure 3 shows the results of the accuracy analysis for the original scale of 0-10 and the new scale of 0-5.

TABLE 3.2 Perma Profile 0-5 ZOU323WS.TXT Aug 25 2024 14:21
INPUT: 305 Person 55 Item REPORTED: 305 Person 55 Item 6 CATS WINSTEPS 4.4.8

SUMMARY OF CATEGORY STRUCTURE. Model="R"

CATEGORY LABEL	SCORE	OBSERVED COUNT	%	OBSVD AVRGE	SAMPLE EXPECT	INFIT MNSQ	OUTFIT MNSQ	ANDRICH THRESHOLD	CATEGORY MEASURE	
0	0	598	4	-.13	-.38	1.41	1.74	NONE	(-2.35)	0
1	1	1146	7	-.06	-.12	1.09	1.15	-.90	-1.10	1
2	2	1775	11	.08	.17	.81	.75	-.41	-.41	2
3	3	3613	22	.42	.51	.83	.74	-.37	.19	3
4	4	5782	34	.89	.87	.89	.86	.21	1.09	4
5	5	3861	23	1.33	1.30	1.11	1.05	1.48	(2.76)	5

Figure 3. Summary of category structure of collapsing category, "01122334455"

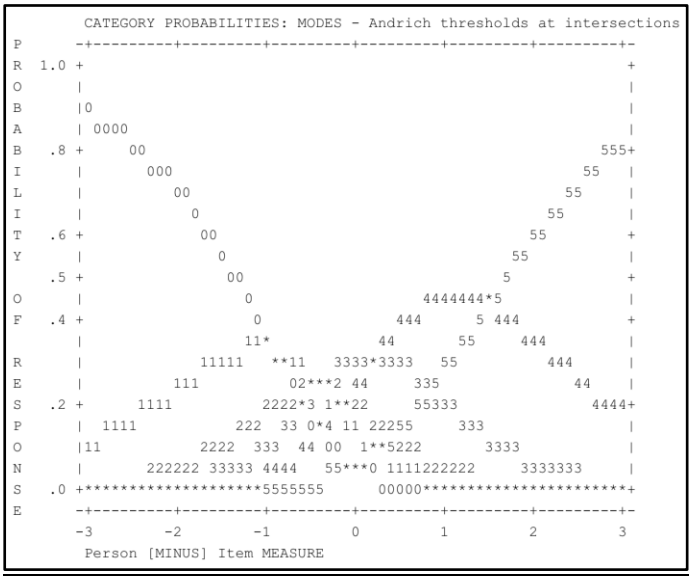

Figure 4. Category probabilities of collapsing category, "01122334455"

The data from the recoded scale categories were reanalyzed using Rating Scale Measurement (RSM) within the Rasch Model. The comparison shows that when the scale is recoded to a 0-5 scale "01122334455," it meets all criteria. Notably, the Andrich threshold

criterion develops monotonically in order according to the categories (NONE < -0.90 < -0.41 < -0.37 < 0.21 < 1.48). However, the recoding of the categories into a five-point scale still does not meet the minimum Andrich threshold range of 1.4 logits. The probability curve generated, as shown in Figure 4, indicates that the peak of the curve for categories 1-5 can be observed, demonstrating that all categories have normal Andrich threshold values. Based on the results of the data analysis from the recoded categories, the final decision to use the 0-5 scale with the code "01122334455" indicates an improvement in the empirical rating scale that meets the Andrich threshold criteria. Therefore, the labels for the new PERMA flourishing scale have been recoded as "very inappropriate" (0), "disagree" (1), "somewhat appropriate" (3), "appropriate" (4), and "very appropriate" (5). This new five-point scale will be used for subsequent data collection and testing.

A well-functioning rating scale in an instrument is crucial for collecting valid and reliable data. A reliable scale produces consistent data over time [28]. A good rating scale allows researchers to accurately measure psychological variables. Research shows that instruments with high validity and reliability can provide more precise and trustworthy data, thereby supporting better decision-making in psychological interventions [29]. Therefore, this study demonstrates a method for analyzing and optimizing the function of the rating scale in the flourishing instrument using Rating Scale Measurement (RSM) within the Rasch Model. The Rating Scale Measurement (RSM) model by Andrich is a variation of the Rasch model used for polytomous data (items with a Likert scale), allowing for the determination of the adequacy of rating scale options to achieve the necessary measurement precision, both in terms of the number of options and their labels (Linacre, 2002; Dubinsky et al., 2024). In the context of the PERMA flourishing instrument, the Likert-type scale used can provide deeper insights into students' well-being.

The PERMA-Profiler is an appropriate instrument to be applied in various contexts (Butler & Kern, 2016), including flourishing. The development of the flourishing instrument based on the PERMA model is grounded in the integration of happiness measurement concepts from both hedonic and eudaimonia perspectives (Elfida et al., 2021). Empirically, the PERMA model significantly contributes to well-being (Huta & Hawley, 2010; Huppert & So, 2013) and has been developed into a flourishing measurement tool (Elfida et al., 2021; Grosvenor et al., 2023). In line with this, research conducted by Diener et al. (2010), Huppert and So (2013), and Rezaei et al. (2015) has validated the PERMA model questionnaire, demonstrating that the model is effective in measuring flourishing. Essentially, the PERMA-Profiler consists of eleven response options with values ranging from 0 to 10, which differs from several long-established happiness instruments widely used in happiness research. Scale measurement is necessary to gain broader insights into students' well-being. Rasch Rating Scale Measurement (partial credit) is a model widely applied to observe ordinal variables that reflect latent variables (Hardigan & Carvajal, 2008; Adams et al., 2017). The

scale assessment in the PERMA model flourishing instrument is conducted based on five criteria recommended by Smith et al. (2003).

The first criterion is met. Each scale category in the PERMA model flourishing instrument has an observed count value of more than ten. This result indicates that the number of observations for each scale is sufficiently stable for estimating the Andrich threshold. This observation is important because the stability of the scale structure contributes to the instrument's validity, ensuring that it can consistently measure the intended construct. If the observed count values are low, the estimation of the Andrich threshold will not be accurate and may be unstable due to inadequate responses (Linacre, 2002).

The second criterion is not met. The observed average does not develop according to the scale categories. Generally, higher measures should indicate higher categories, while lower measures should indicate lower categories [38]. Although the observed average for categories 2 to 10 develops monotonically according to the categories, there is an anomaly in category 1, where the average value is greater than that of the previous category. This finding suggests the presence of response bias or issues in participants' understanding of the instrument, as well as unclear meanings of the rating scale that may lead to controversial or inaccurate measures (Smith et al., 2003; Linacre, 2002).

The third criterion is met. According to the requirements, the outfit MNSQ value must be less than 2.0 logits. The analysis results show that the highest outfit MNSQ value is 1.81 logits. An outfit MNSQ value greater than 1.50 indicates that 50 percent or more of the data is unexplained (Linacre, 2002; Bond & Fox, 2015). Meanwhile, an outfit MNSQ value exceeding 2.0 indicates that the unexplained noise is higher than the explained noise, meaning that the response data contains more falsehoods than useful information, thus not supporting measurements related to the use of scale categories (Elfida et al., 2021). In this context, low noise (criterion met) indicates that the instrument is reliable for measuring students' flourishing. The fourth criterion is not met.

The fourth criterion is not met. Andrich Threshold values must develop according to the scale categories. If this criterion is not met, irregular thresholds will occur (Elfida et al., 2021). The analysis results indicate that the Andrich Threshold values are not met because they do not develop according to the categories, where certain categories have values greater or less than the previous categories, indicating the presence of threshold disturbances. Irregular thresholds indicate low functionality of certain rating steps, caused by the inappropriate use of categories and irregular category frequency patterns (Linacre, 2011). This finding suggests that the instrument may not be fully effective in distinguishing between participants' ability levels, which can affect construct validity. For example, fewer participants are observed in higher categories, while more participants are observed in lower categories. Irregular thresholds do not align with the important conceptual aspects of

the rating scale expected by the Rasch model, where higher measures should be observed in higher categories and vice versa (Linacre, 2011).

The fifth criterion is not met. The Andrich Threshold value range is between 1.4 and 5.0 logits. The analysis results show that the range of Andrich Threshold values is less than 1.4 logits, indicating that the instrument is not sensitive enough to distinguish between different ability levels among participants, which may lead to less accurate conclusions regarding flourishing. The range of Andrich Threshold values indicates that each scale category determines a specific location on the latent variable (Bond & Fox, 2015). A threshold width of less than 1.4 logits signals the presence of category overlap and may indicate that respondents cannot distinguish between scale categories (Bond & Fox, 2015). Therefore, a narrow threshold width, i.e., less than 1.4 logits, is recommended to be combined with adjacent categories or redefined to have more precise meanings (Linacre, 2011).

Based on the results of the scale assessment in Figures 1 and 2, where the Andrich threshold values do not progress from negative to positive, it is recommended to recategorize and recode the scale by merging it with other scale categories. The merging of categories is intended to reduce measurement errors and enhance the functionality of the rating scale (Smith et al., 2003). However, merging categories requires careful consideration, and the meaning or labels of each scale category must be considered (Smith et al., 2003). The combined categories should have a clear logical basis. For example, merging "agrees" and "strongly agree" is reasonable, but merging "agree" and "disagree" is inappropriate because they have different meanings (Elfida et al., 2021). In this study, respondents may not be able to distinguish between categories 2, 5, 7, and 10 from the labels "never" to "always" or from "very inappropriate" to "very appropriate." Based on Linacre (2002) recommendation that simplifying categories can enhance the meaning and function of the rating scale, simplification is carried out by merging the problematic categories into the nearest categories. The decision to simplify the scale is made based on the analysis of the problematic scales, which are less comprehensible to participants when selecting responses. By reducing the number of scale categories, the instrument becomes easier for participants to understand, which can improve the quality of the data obtained. Simplification can help reduce noise in the data, making the instrument more reliable in measuring the intended construct and supporting better validity and reliability of the instrument.

The eleven rating scale categories (0-10) with the label "never" at the initial step of the PERMA model flourishing instrument have been recoded into six scale categories with new labels arranged into options 0-5, with similar labels and adjustments. As shown in Figures 3 and 4, the scale with options 0-5 meets the necessary criteria and demonstrates the functionality of all scale categories. The process of simplifying the problematic scale categories, whether by merging them into lower or higher categories, will enhance the

function and meaning of the rating scale. However, there is no guarantee that the new set of categories will function as expected. Therefore, a diagnosis of the rating scale needs to be conducted at the pilot study stage. The new category set from the pilot data should be retested in subsequent data collection to ensure the functionality of the new rating category set.

A simple and easily understood scale can enhance participant engagement in research. When participants feel that the instrument used is relevant and easy to comprehend, they are more likely to provide honest and accurate responses. The use of rating scales in research generates empirical data that can be used to understand the factors influencing well-being. The PERMA model, which measures various aspects of well-being, allows for a more comprehensive analysis of the interactions among the elements that affect well-being. The use of rating scales in well-being interventions supports evidence-based practices, where decisions are made based on valid data and research outcomes. Thus, a valid and reliable scale can serve as a reference for measuring the effectiveness of well-being programs implemented in various educational settings. However, this research still has limitations related to the sample and the generalization of results. A sample size that is too small is insufficient to represent a broader population.

CONCLUSION

The PERMA model instrument scale for measuring student well-being was analyzed based on five criteria for scale assessment requirements. The analysis results indicate that the simplified scale from 0-10 to 0-5 provides better validity and reliability in measuring the dimensions of PERMA. By using a simpler scale, the instrument not only enhances participants' understanding of the questions posed but also generates more accurate and reliable data. This study demonstrates that the use of the PERMA model as a framework for measuring flourishing can provide valuable insights for the development of student well-being. To strengthen the findings, further research is needed to conduct additional testing in different cultural or population contexts, including social conditions (ethnic background, major, or university) and demographic factors (age and gender), providing a more comprehensive understanding of the validity and reliability of the instrument and its contribution to enhancing well-being in various educational settings. Considering that perceptions of well-being can vary significantly across cultures, it is important to explore how the PERMA model can be applied and validated in different contexts.

REFERENCES

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.

Stallman, H. M. (2010). Psychological distress in university students: A comparison with general population data. *Australian Psychologist*, *45*(4), 249–257. https://doi.org/10.1080/00050067.2010.482109

Duffy, R. D., Bott, E. M., & Allan, B. A. (2016). The role of work in the flourishing of college students. *Journal of College Student Development*, *57*(5), 563–578.

Greenhow, C., & Lewin, C. (2016). Social media and education: Reconceptualizing the boundaries of formal and informal learning. *Learning, Media and Technology*, *41*(1), 6–30.

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

Kern, M. L., Waters, L. E., Adler, A., & White, M. A. (2015). A multidimensional approach to measuring well-being in students: Application of the PERMA framework. *The Journal of Positive Psychology*, *10*(3), 262–271.

Hone, L. C., Jarden, A., & Schofield, G. (2015). The relationship between well-being and academic performance: A study of university students. *Journal of Happiness Studies*, *16*(1), 1–20.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68–78.

Fredrickson, B. L. (2001). The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions. *American Psychologist*, *56*(3), 218–226.

Steger, M. F., Frazier, P., Oishi, S., & Kaler, M. (2006). The meaning in life questionnaire: Assessing the presence and search for meaning in life. *Journal of Counseling Psychology*, *53*(1), 80–93.

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

Butler, J., & Kern, M. L. (2016). The PERMA-Profiler: A brief multidimensional measure of flourishing. *International Journal of Wellbeing*, *6*(3), 1–48.

Khaw, D., & Kern, M. L. (2014). A cross-cultural comparison of the PERMA model of well-being. *The Journal of Positive Psychology*, *9*(1), 62–67.

Diener, E., Wirtz, D., Tov, W., Kim-Prieto, C., Choi, D. W., Oishi, S., & Biswas-Diener, R. (2010). New well-being measures: Short scales to assess flourishing and positive and negative feelings. *Social Indicators Research*, *97*(2), 143–156.

Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Pearson.

Creswell, J. W. (2013). *Research design: Qualitative, quantitative, and mixed method approaches*. SAGE Publications.

Creswell, J. W., & Creswell, J. D. (2023). *Research design: Qualitative, quantitative, and mixed methods approaches* (6th ed.). SAGE Publications.

Lambert D'raven, L., & Pasha-Zaidi, N. (2016). Using the PERMA model in the United Arab Emirates. *Social Indicators Research*, *125*(3), 905–933. https://doi.org/10.1007/s11205-015-0866-0

Didino, D., Taran, E. A., Barysheva, G. A., & Casati, F. (2019). Psychometric evaluation of the Russian version of the flourishing scale in a sample of older adults living in Siberia. *Health and Quality of Life Outcomes*, *17*(1), 1–12. https://doi.org/10.1186/s12955-019-1100-6

Wammerl, M., Jaunig, J., Mairunteregger, T., & Streit, P. (2019). The German version of the PERMA-Profiler: Evidence for construct and convergent validity of the PERMA theory

of well-being in German speaking countries. *Journal of Well-Being Assessment*, *3*(2), 75–96. https://doi.org/10.1007/s41543-019-00021-0

Romano, I., Ferro, M. A., Patte, K. A., Diener, E., & Leatherdale, S. T. (2020). Measurement invariance of the flourishing scale among a large sample of Canadian adolescents. *International Journal of Environmental Research and Public Health*, *17*(21), 1–15. https://doi.org/10.3390/ijerph17217800

Mahamid, F., Veronese, G., & Bdier, D. (2023). The factorial structure and psychometric properties of the PERMA-Profiler Arabic version to measure well-being within a Palestinian adult population. *Middle East Current Psychiatry*, *30*(1), 1–10. https://doi.org/10.1186/s43045-023-00282-9

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

Smith, E. V., Wakely, M. B., de Kruif, R. E. L., & Swartz, C. W. (2003). Optimizing rating scales for self-efficacy (and other) research. *Educational and Psychological Measurement*, *63*(3), 369–391. https://doi.org/10.1177/0013164403063003002

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

Linacre, J. M. (2011). Rasch measures and unidimensionality. *Rasch Measurement Transactions*, *24*(4), 1300–1310.

Elfida, D., Milla, M. N., Mansoer, W. W. D., & Takwin, B. (2021). Adaptasi dan uji properti psikometrik The PERMA-Profiler pada orang Indonesia. *Persona: Jurnal Psikologi Indonesia*, *10*(1), 81–103. https://doi.org/10.30996/persona.v10i1.4986

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53–55.

Dubinsky, M., et al. (2024). The urgency numeric rating scale: Psychometric evaluation in adults with Crohn's disease. *Advances in Therapy*. Advance online publication. https://doi.org/10.1007/s12325-024-03081-8

Butler, J., & Kern, M. L. (2016). The PERMA-Profiler: A brief multidimensional measure of flourishing. *International Journal of Wellbeing*, *6*(3), 1–48. https://doi.org/10.5502/ijw.v6i3.526

Huta, V., & Hawley, L. (2010). Eudaimonia and hedonia: Concepts and findings. *Journal of Happiness Studies*, *11*(2), 113–135.

Diener, E., Wirtz, D., Tov, W., Kim-Prieto, C., Choi, D. W., Oishi, S., & Biswas-Diener, R. (2010). New well-being measures: Short scales to assess flourishing and positive and negative feelings. *Social Indicators Research*, *97*(2), 143–156. https://doi.org/10.1007/s11205-009-9493-y

Grosvenor, L. P., Errichetti, C. L., Holingue, C., Beasley, J. B., & Kalb, L. G. (2023). Self-report measurement of well-being in autistic adults: Psychometric properties of the PERMA Profiler. *Autism in Adulthood*, *5*(4), 401–410. https://doi.org/10.1089/aut.2022.0049

Huppert, F. A., & So, T. (2013). Flourishing across cultures: A review of the literature. *Social Indicators Research*, *110*(3), 837–861. https://doi.org/10.1007/s11205-013-0340-1

Rezaei, M., Soleimani, S., Kianersi, F., & Hojabrian, H. (2015). Development and validation of flourishing questionnaire based on Seligman's model among Iranian university students. *Journal of Health Research*, *3*(1), 3–12.

Hardigan, P. C., & Carvajal, M. J. (2008). Application of Rasch rating scale model to analysis of job satisfaction among practicing pharmacists. *Journal of the American Pharmacists Association*, *48*(4), 522–529. https://doi.org/10.1331/JAPhA.2008.07042

Adams, N., Hamilton, N., Nelson, E. L., & Smith, C. E. (2017). Using telemedicine to identify depressive symptomatology rating scale in a home parenteral nutrition population. *Journal of Technology in Behavioral Science*, *2*(3), 129–139. https://doi.org/10.1007/s41347-017-0025-3

Chong, L., Raina, A., Kotovsky, K., & Goucher-Lambert, K. (2022). The evolution and impact of human confidence in artificial intelligence and in themselves on AI-assisted decision-making in design. *Journal of Mechanical Design*, *145*(3), 1–12. https://doi.org/10.1115/1.4055123