Construct Validity of the Kessler Psychological Distress Scale (K-10) Instrument in the Elderly with Disabilities: Rasch Model Analysis

Lukman Hilfi[1], Mulya Nurmansyah Ardisasmita[2], Fauzura Nuril Mahra[3]

[1,2,3] Universitas Padjadjaran
Indonesia
lukman.hilfi@unpad.ac.id*

Check for updates

Abstract

Aging contributes to functional decline and heightened psychological distress that affects the older adults' quality of life. The World Health Organization (WHO) recommends the 10-item Kessler Psychological Distress Scale (K-10) for assessing psychological distress, however the Classical Test Theory (CTT)-based evaluation method has limitations. Item Response Theory (IRT), specifically the Rasch model, offers a more accurate method for measuring the instrument's validity and reliability. This study examined the construct validity of the K-10 Indonesian version in older adults with disabilities. A cross-sectional method and secondary data from 321 elderly individuals (≥60 years) with disabilities in Bandung City, Indonesia, were used between August and February 2025. The instrument's construct validity was analyzed by performing the Rasch model. A total of 315 responses were eligible for analysis. The K-10 showed an overall fit to the Rasch model (mean outfit MnSq = 1.03±0.27). Item C10 (—feel worthless‖) was identified as misfitting (outfit MnSq >1.4; ZSTD >2), although no factor bias was detected. The explained variance reached 54.10%. Reliability indicators were strong, with a Cronbach's alpha of 0.89 and a person separation index above 2. However, category functioning and targeting did not fully meet Rasch expectations. The correlation between K-10 and WHODAS 2.0 showed a moderate positive relationship, though it did not reach the criterion for convergent validity (r >0.6). The Indonesian version of the K-10 has acceptable construct validity in a sample of older adults with disabilities in Bandung City, Indonesia.

A. Introduction

Humans, as living beings, undergo a life cycle process of growth and development that runs parallel to the aging process, at the end of which is the stage known as old age or senior citizenship (Gu, 2022). Individuals who undergo the natural aging process will experience physical, emotional, physiological, and cognitive changes, as well as a gradual decline in functional abilities in daily living skills, which is defined as impairment and can lead to the onset of disability, thereby affecting an individual's quality of life (QOL) and health, such as by triggering psychological distress and despair (Pilar Matud & Concepción García, 2019).

In 2023, the World Health Organization (WHO) reported 14% of older adults experience psychological distress, and according to the Global Health Estimates (GHE) in 2019, this condition accounts for 10.6% of total disabilities among older adults, with the most common categories being depression and anxiety.

Furthermore, one-quarter of global deaths caused by suicide (27.2%) occurred among those aged 60 and above in 2019 (World Health Organization, 2023). To address this issue, the WHO has established a measurement tool to assess the severity of various psychological distress conditions over the past four weeks using the Kessler Psychological Distress Scale, available in both six-item (K-6) and ten-item (K-10) versions, which are valid and reliable (Lins et al., 2021). The results of the instrument evaluation can be considered valid and of high quality if they meet the criteria for validity and reliability testing using a statistical analysis method (Maulana, 2023).

The Classical Test Theory (CTT) or Traditional Theory method, that is commonly used, has limitations because it focuses on total scores, uses an ordinal scale without fixed intervals, and does not account for differences in item preference levels, which can lead to less accurate results (Fernanda & Hidayah, 2020). As an alternative, Item Response Theory (IRT), known as Rasch model analysis, offers a more personalized and linear approach, meeting the criteria for objective measurement, particularly in clinical health research (Bond et al., 2020; Sumintono, 2016). Studies in Australia, Bangladesh, and China indicate that while the K-10 is valid and reliable according to CTT, IRT results suggest poor reliability, necessitating modifications to some items (Calkin et al., 2023; Uddin et al., 2018; Zhu et al., 2021). After revisions, the Rasch model demonstrated unidimensionality and improved reliability. In Indonesia, the K-10 has been validated using CTT, but no IRT analysis has been conducted (Paramartadewi et al., 2023).

To date, no study has validated the Indonesian version of the Kessler Psychological Distress Scale (K-10) using the Rasch model among older adults with disabilities. Previous validations have relied on Classical Test Theory, leaving a methodological gap in evaluating its item-level performance and measurement invariance in this specific population. Thus, the objective of this study is to assess the validity and reliability of the Indonesian K-10 by employing the Rasch model, taking into consideration the sociodemographic factors of the participants.

B. Methods

The study applies a quantitative analytical method with a cross-sectional design, based on secondary data obtained from the *Riset Percepatan Lektor Kepala* (RPLK) research dataset under the name of Lukman Hilfi with Contract Number 1549/UN6.3.1/PT,00/2023 dated March 27, 2023.

The population in this study consists of data on older adults (aged ≥60 years) with disabilities in the city of Bandung in 2023, totaling 321 persons, with a sampling technique using total sampling. The disabilities among the elderly population were caused by impairments, including amputation, limb paralysis, cataracts, blindness, and speech impairment. Additionally, six cases with incomplete responses greater than 30% were removed prior to conducting the Rasch model analysis (World Health Organization, 2010). The final sample we used for analysis was 315 persons.

This study was conducted in the city of Bandung from August 2024 to February 2025 using secondary data from a study conducted from July to September 2023. Ethical approval for this study was granted by the Ethics Committee of Universitas Padjadjaran (No. 1131/UN6.KEP/EC/2024).

1. Demographic data, including sociodemographic information such as age, gender, marital status, educational history, employment history, health insurance status, and type of difficulty/impairment/disability (Bonifay et al., 2015).
2. 10-item Kessler Psychological Distress Scale (K-10) instrument.

The K-10 consists of 10 questions to measure levels of anxiety, depression, stress, and various other psychological distress conditions over the past four weeks (Lins et al., 2021). The instrument is administered using a method tailored to the conditions of each individual in the population, namely interviewer-administered or self-administered under our supervision. The Indonesian version has been translated by the Language Center of the Faculty of Cultural Sciences at Universitas Padjadjaran. The instrument employed a five-point Likert scale, with response options ranging from 1 (none of the timel) through 5 (all of the time). Total scores were subsequently categorized into four levels of severity: subclinical (10–19), mild (20–24), moderate (25–29), and severe (30–50) (Andrews & Slade, 2001).

3. 12-item self-administered WHODAS 2.0

The World Health Organization Disability Assessment Schedule (WHODAS 2.0) is a measurement method developed by the WHO to assess overall health and disability levels over a 30-days (World Health Organization, 2010). To capture functional health, the instrument employs 12 items spanning six

domains: cognitive function, physical mobility, self-maintenance, social relations, capacity for daily tasks, and engagement in society. Scores extend from 0 (no disability) to 100 (maximum disability). Based on the total score, disability is interpreted as mild (5–24), moderate (25–49), severe (50–95), or extreme (96–100). We used the 12-item self-administered WHODAS 2.0 Indonesian version, which had been previously tested and retranslated into English, and used in various previous studies (Hilfi et al., 2021). WHODAS 2.0 employs a Likert-type scoring system, assigning values from 0 (absence of disability) through 4 (extreme disability or complete inability to function). In this study, the interval logit scale derived from WHODAS 2.0 responses served as the basis for assessing convergent validity with the K-10 (World Health Organization, 2010).

Rasch model analysis was performed using Winsteps version 5.5.3 (Linacre, 2024) and was applied to assess the construct validity of the K-10, assuming that each item scale has a different preference level. The Rasch model converts the ordinal scale on the raw score into a specific measurement unit called logit (log-odds unit) (Chong et al., 2022).

Construct validity is defined as a result that is consistent with the interpretation or meaning of the values in a test and can explain the underlying process of a test's performance and its relationship to other phenomena (Piedmont, 2023). This study examined the construct validity of the K-10 through an evaluation of seven indicators (Pitts & Naumenko, 2016).

1. Fit Statistics

There are three assessment criteria: data fit with the model (overall fit), item quality in the instrument (item fit), and person fit with the model (person fit). Item and person misfit was identified when at least two criteria were violated: outfit MnSq values outside 0.5–1.5 logits, outfit ZSTD outside –2.0 to 2.0 logits, or Pt Measure Corr beyond 0.40–0.85 (Sumintono, 2016).

2. Unidimensionality

An instrument must capture a single latent trait at any given time with a variance measure ≥50% (Bonifay et al., 2015).

3. Response Category Function

Assessing how well the response categories (response options) on each item perform by evaluating four indicators (Bond et al., 2020): (1) the mean and Andrich threshold for each category must increase monotonically, (2) Acceptable Andrich threshold differences lie between 1.4 and 5 logits, (3) the outfit mean square for all categories should be less than 1.5, and (4) each response category must demonstrate a probability curve peaking as the most likely response across several points of the measurement continuum.

4. Reliability and Separation

    a. Reliability

Assessing how well the instrument used can produce stable and consistent results using reliability tests (Cronbach's alpha), person reliability, and item reliability with results >0.60 (Kholil & Faradiba, 2024). The analysis employed SPSS version 28.0 for preliminary statistics, and Cronbach's alpha (α = 0.89) was confirmed based on the Winsteps reliability output.

    b. Separation Index

Assessing how well the instrument used can stratify persons or items with a separation index value must be ≥2 (Boone, 2016).

5. Targeting

In psychometric terms, targeting refers to the extent to which the selected items are appropriately aligned with the sample's response tendencies. An instrument will be categorized as good if it meets the following criteria: (1) the average person-item difference (M) <1 logit (Cantó-Cerdán et al., 2021), and (2) the higher the targeting area, the better the targeting (Arias González et al., 2015).

6. Differential Item Functioning (DIF) of Misfitting Items

Detecting possible causes of misfitting items through bias factors such as gender, age and marital status, educational history, employment status, and health insurance (Mardiana et al., 2022; Sampasa-Kanyinga et al., 2018) with a DIF contrast exceeding 0.5 with statistical significance at $p < 0.05$ (Bond et al., 2020).

7. Convergent Validity

Investigating correlations with other validated instruments with a correlation coefficient (r) >0.6 (Akoglu, 2018).

C. Results and Discussion

From the initial pool of 321 participants, six were excluded because their responses were more than 30 percent missing. The remaining 315 valid cases formed the analytic dataset used for all Rasch model computations and subsequent interpretations. The following sections describe the key psychometric findings, including fit statistics, unidimensionality, response category functioning, reliability and separations index, targeting and Differential Item Functioning (DIF), followed by a discussion of their practical implications for psychological distress screening among older adults with disabilities. Tables 1 and 2 show the analysis results of the K-10 instrument and each item.

1. Fit Statistics

Fit statistics for the K-10 were examined. The scale displayed satisfactory overall fit to the Rasch model (average outfit MnSq = 1.03). One item C10, *Dalam 4 minggu terakhir, seberapa sering anda merasa tidak berharga?* was identified as misfitting item, with an outfit MnSq = 1.69, an outfit ZSTD = 5.8, and a point-measure correlation of 0.60. This pattern suggests that item C10 merits further investigation (for example, cognitive interviewing or re-translation).

Table 1. Construct validity assessment of the K-10 using the Rasch model

Indicators	K-10	Expectation
Sample (N)	315	N/A
Overall average outfit MnSq (SD)	1.03 (0.27)	0.5-1.5
Number of misfitting item (label)	1 item (C10)	None or Minimum
Raw variance explained by measure	54.10%	$\geq 50\%$
Unexplained variance by the first contrast (eigenvalue)	9.2% (2)	< 15% (≤ 2)
Measurement reliability (Cr- α)	0.89	0.68
Person-reliability	0.85	0.68
Item-reliability	0.98	0.68
Person-separation	2.40	> 2
Item-separation	6.36	> 2
WHODAS 2.0 Correlation	0.442	> 0.4

SD: Standard deviation; Cr-α: Cronbach's alpha

Table 2. Mean Square (MnSq) of the 10-item Kessler Scale (K-10)

Kessler Psychological Distress Scale (K-10)	N=315	
	Infit MnSq	Outfit MnSq
Dalam 4 minggu terakhir, seberapa sering anda merasa kelelahan tanpa sebab?	1.20	1.17
Dalam 4 minggu terakhir, seberapa sering anda merasa gugup?	0.95	0.94
Dalam 4 minggu terakhir, seberapa sering anda merasa sangat gugup sampai tidak ada yang bisa menenangkan anda?	0.84	0.84
Dalam 4 minggu terakhir, seberapa sering anda merasa tidak berdaya?	1.17	1.13
Dalam 4 minggu terakhir, seberapa sering anda merasa gelisah atau cemas?	0.81	0.78
Dalam 4 minggu terakhir, seberapa sering anda merasa sangat gelisah sampai anda tidak bisa duduk tenang?	1.25	1.21
Dalam 4 minggu terakhir, seberapa sering anda merasa tertekan?	0.93	0.96

|		
Dalam 4 minggu terakhir, seberapa sering anda merasa bahwa semuanya adalah usaha sia-sia?	0.72	0.72
Dalam 4 minggu terakhir, seberapa sering anda merasa sangat sedih sampai tidak ada yang bisa menghibur anda?	0.85	0.83
Dalam 4 minggu terakhir, seberapa sering anda merasa tidak berharga?	1.26	1.69

2. Unidimensionality

The explained variance value, unexplained variance in the first contrast value, and the eigenvalue units were obtained as stated in table 1, indicating unidimensionality, model fit, and good precision for the sample (Bonifay et al., 2015). The eigenvalue units of two indicates that there are two additional variances outside the standard measurement dimension, equivalent to two items (loading >0.3), namely C2 (loading = 0.84) and C3 (loading = 0.83). Meanwhile, C1 (loading = 0.30) is still within the range of suspicion of being related to other additional dimensions (Bond et al., 2020).

3. Response Category Function

Evaluation of the response category function values from K-10 shows that:

a. The values of the observed average (–1.81 < –1.02 < –0.31 < 0.36 < 1.13) and the Andrich thresholds ($\tau1$ –1.25 < $\tau2$ –0.77 < $\tau3$ 1.13 < $\tau4$ 1.89) consisting of 5 item choice categories (never, rarely, sometimes, almost always, and always) show a consistent increase and no category disrupts the calculation.

b. The difference in Andrich thresholds between category 2 *jarang*' and category 3 *kadang- kadang*', as well as the difference between category 3 *kadang-kadang*' and category 4 *hampir setiap saat*', shows a value less than 1.4 logits, indicating that these categories tend to misfit the model.

c. The overall outfit MnSq value for all categories is less than 1.5 logits, indicating that they fit the model.

d. The observed peak probabilities for the five response categories did not conform to the expected ordering of category functioning under the Rasch model. In particular, the intermediate categories (*'jarang'*, *'kadang-kadang'*, and *'hampir setiap saat'*) demonstrated depressed peak values relative to the more extreme categories (*'tidak pernah'* and *'setiap saat'*).

Figure 1 displays the probability curves of the five categories. The latent construct assessed, psychological distress, is mapped on the x-axis, while the probability of category endorsement (0–1) is shown on the y-axis.
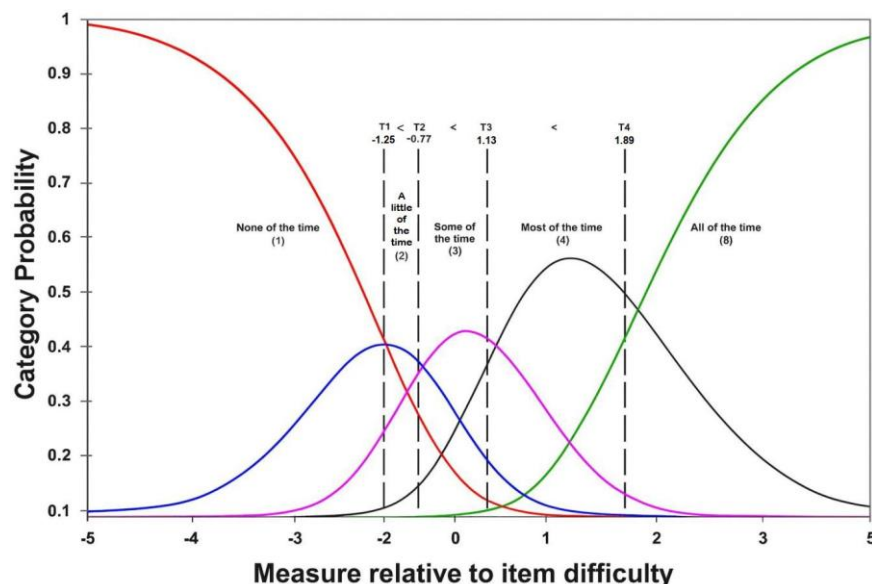
Figure 1. Category probability curve from the Kessler Scale (K-10)

4. Reliability and Separation
   a. Reliability
      The person reliability, item reliability, and Cronbach's alpha of the K-10 indicated excellent internal consistency as shown in the Table 1. This suggests that the items within the instrument measured the same underlying construct consistently. High reliability coefficients demonstrate that the K-10 produces precise and replicable measurements of psychological distress among older adults with disabilities.
   b. Separation Index
      Table 1 shows that the person and item separation indexes met the requirements, since the values are greater than two, enabling stratification of persons and items based on psychological distress levels using the K-10.

5. Targeting

The average person measure was –0.65 logits (SD = 1.26), compared to 0.00 logits (SD = 0.45) for items, producing a difference of 0.65 logits. As this falls below the 1-logit criterion, targeting is considered adequate, with 137 participants (43.5%) situated within the overlap of item thresholds and person distress levels. Meanwhile, 144 participants (45.7%) have values below the lowest logit of the item preference level, and 34 participants (10.8%) have values above the highest logit of the item preference level. Figure 2 shows that item distribution adequately targets the sample, though further calibration may improve interpretability for community-based screening contexts.
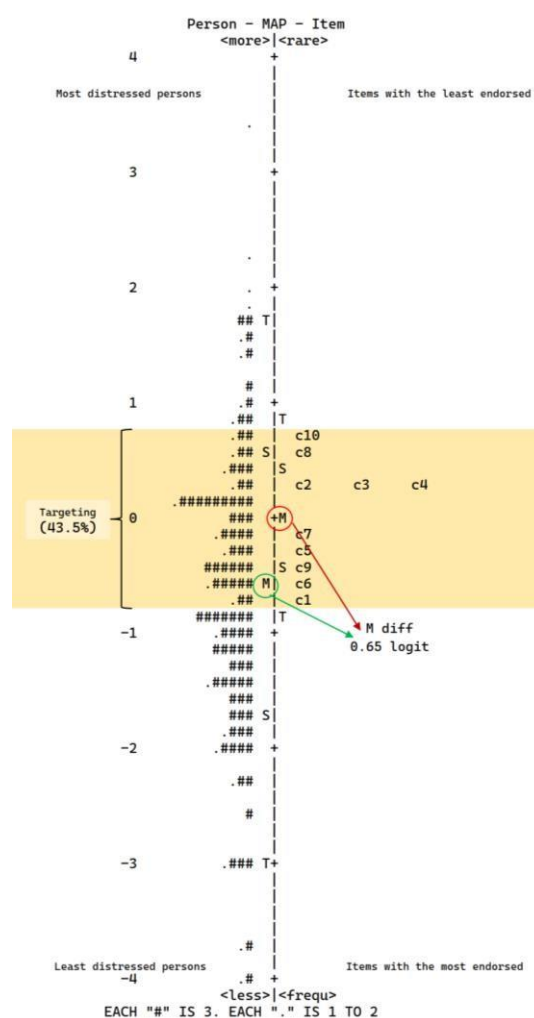
Figure 2. Wright Person-Item Distribution Map. Each represents 3 subjects, and each corresponds to one person (M = mean; S = 1 SD; T = 2 SD).

6. Differential Item Functioning (DIF) of Misfitting Item

The Differential Item Functioning (DIF) contrast and probabilities of the misfitting item C10 were evaluated based on age, gender, marital status, highest level of education, employment status, and health insurance coverage. The results showed no significant DIF (DIF contrast >0.5 and p <0.05) in the groups with these characteristics that might affect item C10 (Bond et al., 2020).

7. Convergent Validity

The K-10 has a statistically significant positive correlation with the WHODAS 2.0, with a moderate correlation category, with a value of r = 0.442 and p-value = 0.00 (<0.05) (Akoglu, 2018). However, this correlation does not meet the criteria for convergent validity (r >0.6) for the K-10.
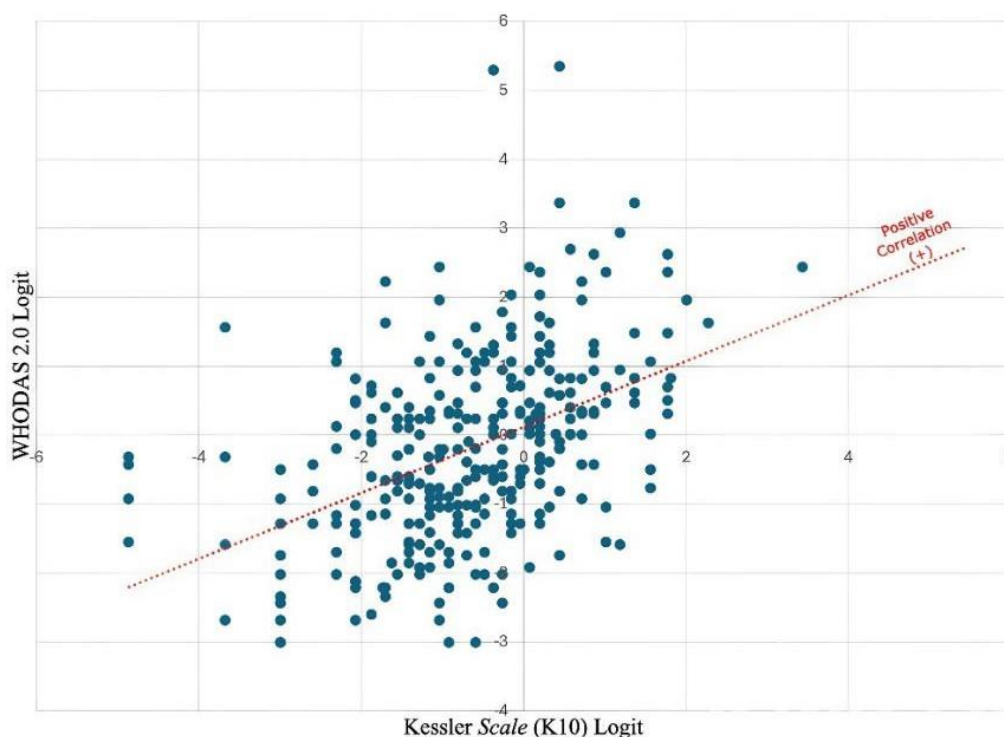
Figure 3. Distribution Map of the Kessler Scale (K-10)–WHODAS 2.0

The results of the K-10 analysis have met the Rasch model expectations, which supports construct validity so that the instrument can be used effectively. To our knowledge, this is the initial study of the construct validity of the Indonesian adaptation of the K-10 in older adults with disabilities using Rasch model. Given the substantial burden of psychological distress among older adults, especially those with disabilities, reliable measurement tools are indispensable for guiding community mental health and well-being programs.

In this study, we used 315 data points for other analyses, and the results indicated that the K-10 has good reliability and unidimensionality, with one misfitting item (C10) identified: *Dalam 4 minggu terakhir, seberapa sering anda merasa tidak berharga?*

The average outfit MnSq was 1.03 logit with one misfitting item, C10, *Dalam 4 minggu terakhir, seberapa sering anda merasa tidak berharga?* This indicates that the performance of this item is unpredictable and does not represent the same construct. Consistent with the present study, prior study from Bangladesh indicated that C10 functioned as a misfitting item, largely attributable to its high positive residuals and their potential impact on the model's overall misfit (Uddin et al., 2018). Misfits were also found in some participants due to lucky guesses, carelessness, incorrect answer selection, misunderstanding, or lack of knowledge about the question item, thereby affecting the quality of the item. In this study, misfit may be caused by differing interpretations among participants, particularly for those with differing perspectives on the concept of tidak berharga‖ in item C10. During the adaptation process, it was found that translating words like worthless‖ literally as tidak berharga‖ may not always be

appropriate, as the meaning can vary across cultural and psychological contexts. In some cultures, feelings of worthlessness‖ may also be associated with spiritual aspects, social status, or economic conditions, which may not align with the psychological distress aspects intended in this instrument scale. Therefore, it may be considered to modify the sentence options to be more easily interpreted and have similar meanings, such as useless/tidak berguna‖ or meaningless/tidak berarti (Febriana & Sari, 2024; Kalb, 2025).

The unidimensionality evaluation showed a measurement variance of 54.1% (≥50% = good) which means that most of the measurement variance comes from the same type of construct, namely the level of psychological distress of a person. Meanwhile, the eigenvalue units of 2.0 (≤2) indicate that there are 2 additional variances equivalent to 2 items (loading> 0.3), namely C2 (loading = 0.84) and C3 (loading = 0.83) (Bonifay et al., 2015). Meanwhile, C1 (loading = 0.30) is still within the limits of suspicion of having a relationship with other additional dimensions (Bond et al., 2020). This may indicate that other psychological aspects may contribute to the measurement of K-10, such as individual perceptions of their health conditions or social aspects that affect psychological distress.

The DIF evaluation did not show any bias in any characteristics, including the different procedures for filling out the instrument through interviewer-administered or self-administered, which indicates that K-10 is not fixated on only one side of the perspective in calculating the measurement scale. This finding is consistent with evidence reported in studies conducted in China (Zhu et al., 2021) and Bangladesh (Uddin et al., 2018) where no bias was found in various characteristics, especially in age and gender. In addition, it is also in line with research on Dutch, Moroccan, and Turkish participants where no bias was detected in all characteristics such as gender, age and marital status, educational history, employment status, and health insurance (Mardiana et al., 2022; Sampasa-Kanyinga et al., 2018). In other words, the K-10 can be used fairly and does not show systematic differences in the way each participant answers the items on this scale, so that the validity and reliability of the instrument are maintained.

Comparable reliability values have been documented in prior research, such as 0.884 in Indonesia (Paramartadewi et al., 2023), and 0.916 in China (Zhu et al., 2021), along with similar findings from other countries (Calkin et al., 2023). The separation index of the instrument has also met the requirements (≥2) in the sense that it can distinguish items and participants based on their level of psychological distress.

The K-10 response category results scale showed shortcomings in meeting the Rasch model expectations of, which can be seen based on the difference in peak points on the probability curve and the Andrich threshold difference of less than 1.4 logits between category 2 _jarang' and category 3 _kadang-kadang', as well as the difference between category 3 _kadang-kadang' and category 4 _hampir setiap saat'. Andrich's threshold difference of fewer than 1.4 logits indicates a lack of clear distinction and is not sufficiently informative in the response categories on the measurement scale, which tends to confuse participants' choices and may lead to item misfitting (Calkin et al., 2023).

Response categories that are less in line with the indicators can be rescored by combining several categories to better match the required indicators (Bond et al., 2020). By combining these categories, will reduce the number of response categories from 5 to 4 response categories on the Indonesian version of the K-10 instrument which is in line with research in Australia (Calkin et al., 2023), China (Zhu et al., 2021), and Bangladesh (Uddin et al., 2018), which found that the irregularity of the threshold refers to a linear increase in pressure that is not progressively reflected by the response category options (Bond et al., 2020).

In this study, the response categories were not modified because they still showed adequate criteria for unidimensionality, and were therefore still able to differentiate between the participants' measured levels of psychological distress. To optimize the psychometric properties of the Indonesian version of the K-10, it may be necessary to merge certain response categories, thereby producing a scale that is more informative and easier for participants to interpret (Bond et al., 2020).

Although the target area of the K-10 was 43.5%, some participants were judged to have psychological distress levels that were too high for other participants to answer the items and therefore not relevant to the target sample and vice versa, some participants had psychological distress levels that were too low for other participants to answer the items. This is consistent with findings reported in studies conducted in China (Zhu et al., 2021) and Bangladesh (Uddin et al., 2018) which showed that their version of K-10 struggled to distinguish people with low levels of psychological distress, which may be caused by participants who are already satisfied and in good condition with their current level of distress lower than

the expected level of psychological distress, and vice versa. However, further research is needed involving participants with more equal levels of psychological distress overall to achieve good targeting.

Convergent validity for the K-10 could not be confirmed as it did not meet the measurement standard (r>0.6). However, the results of the correlation test conducted with WHODAS 2.0 show a positive correlation coefficient classified as moderate strength or moderate from the observation between the K-10 score and WHODAS 2.0 which indicates that the condition of disability has an influence on a person's psychological distress. This finding is consistent with previous research, which found a positive correlation with WHODAS 2.0 (Sampasa-Kanyinga et al., 2018). This result suggests that the Indonesian version of the K-10 provides an accurate and valid assessment of psychological distress in this population, with measurement expressed on a logit scale.

Overall, the findings of this study demonstrate that the Indonesian version of the K-10 holds strong practical value for clinical and community settings. Its application can assist health professionals in the primary care services, such as Community Health Centre *(Puskesmas)* and Private Health Clinic, in recognizing early signs of psychological distress among older adults with disabilities and in designing appropriate psychosocial responses within primary care services.

The Indonesian version of the K-10 is suitable for integration into community outreach programs (*pengabdian masyarakat*) and public health nursing services. Its simplicity allows use in screening and counseling activities conducted by *Puskesmas* staff or community health workers for early detection of psychological problems among older adults.

Several limitations should be acknowledged. The distribution of psychological distress among participants did not fully correspond with the targeted range, as many reported levels that were either above or below the ideal targeting zone. Additionally, recruitment was limited to participants with disabilities residing in Bandung, which may constrain the generalizability of the findings across other Indonesian regions. Nevertheless, these results are expected to enhance current understanding and contribute to the evidence base on psychological distress in Indonesia.

D. Conclusion

The findings indicate that, despite the presence of misfitting item, inadequate category functioning, and incomplete targeting of participants, the Indonesian version of the K-10 met key psychometric criteria under the Rasch model, showing unidimensionality and satisfactory internal consistency. These results suggest that the Indonesian version of the K-10 has acceptable construct validity for assessing levels of psychological distress in a sample of older adults with disabilities in Bandung, Indonesia.

The Indonesian version of the K-10 is feasible for use in community health programs and nursing practice for early screening of psychological distress among older adults with disabilities. Its application can strengthen mental-health components in elderly care at Puskesmas and community-based services.

E. Acknowledgment

F. Author Contribution Statement

Lukman Hilfi was responsible for the study design, Rasch model analysis, interpretation of findings, and preparation of the manuscript draft. Mulya Nurmansyah Ardisasmita and Fauzura Nuril Mahra contributed to the data collection, critical feedback and revisions on the final draft. All authors reviewed and approved the final manuscript.

References

Akoglu, H. (2018). User's guide to correlation coefficients. *Turkish Journal of Emergency Medicine*, *18*(3), 91–93. https://doi.org/10.1016/j.tjem.2018.08.001

Andrews, G., & Slade, T. (2001). Interpreting scores on the Kessler Psychological Distress Scale (K10). *Australian and New Zealand Journal of Public Health*, *25*(6), 494–497. https://doi.org/10.1111/j.1467-842X.2001.tb00310.x

Arias González, V. B., Crespo Sierra, M. T., Arias Martínez, B., Martínez-Molina, A., & Ponce, F. P. (2015). An in-depth psychometric analysis of the Connor-Davidson Resilience Scale: calibration with Rasch-Andrich model. *Health and Quality of Life Outcomes*, *13*(1), 154. https://doi.org/10.1186/s12955-015-0345-y

Bond, T. G., Yan, Z., & Heene, M. (2020). Applying the rasch model: Fundamental measurement in the human sciences. In *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. Taylor and Francis. https://doi.org/10.4324/9780429030499

Bonifay, W. E., Reise, S. P., Scheines, R., & Meijer, R. R. (2015). When Are Multidimensional Data Unidimensional Enough for Structural Equation Modeling? An Evaluation of the DETECT Multidimensionality Index. *Structural Equation Modeling: A Multidisciplinary Journal*, *22*(4), 504–516. https://doi.org/10.1080/10705511.2014.938596

Boone, W. J. (2016). Rasch analysis for instrument development: Why, when, and how? *CBE Life Sciences Education*, *15*(4). https://doi.org/10.1187/cbe.16-04-0148

Calkin, C. J., Numbers, K., Brodaty, H., Sachdev, P. S., & Medvedev, O. N. (2023). Measuring distress in older population: Rasch analysis of the Kessler Psychological Distress Scale. *Journal of Affective Disorders*, *330*, 117–124. https://doi.org/10.1016/j.jad.2023.02.116

Cantó-Cerdán, M., Cacho-Martínez, P., Lara-Lacárcel, F., & García-Muñoz, Á. (2021). Rasch analysis for development and reduction of Symptom Questionnaire for Visual Dysfunctions (SQVD). *Scientific Reports*, *11*(1), 14855. https://doi.org/10.1038/s41598-021-94166-9

Chong, J., Mokshein, S. E., & Mustapha, R. (2022). Applying the Rasch Rating Scale Model (RSM) to investigate the rating scales function in survey research instrument. *Jurnal Cakrawala Pendidikan*, *41*(1). https://doi.org/10.21831/cp.v41i1.39130

Febriana, M. O., & Sari, J. D. E. (2024). Self-Assessment of Self-Worth and Its Relationship with Depression Level: A Case Study of Female Students in Banyuwangi, Indonesia. *Health Dynamics*, *1*(9), 340–348. https://doi.org/10.33846/hd10905

Fernanda, J. W., & Hidayah, N. (2020). Analisis Kualitas Soal Ujian Statistika Menggunakan Classical Test Theory dan Rasch Model. *Square : Journal of Mathematics and Mathematics Education*, *2*(1), 49. https://doi.org/10.21580/square.2020.2.1.5363

Gu, W. (2022). Healthy Long-Lived Human Beings—Working on Life Stages to Break the Limitation of Human Lifespans. *Biology*, *11*(5). https://doi.org/10.3390/biology11050656

Hilfi, L., Atik, N., Raksanagara, A. S., Sunjaya, D. K., Paramita, S. A., Yamazaki, C., Koyama, H., & Hamazaki, K. (2021). Rasch Model Analysis of the Indonesian Version of World Health Organization Disability Assessment Schedule (WHODAS 2.0). *The Kitakanto Medical Journal*, *71*(4), 275–288. https://doi.org/10.2974/kmj.71.275

Kalb, D. (2025). *Value and Worthlessness*. Berghahn Books. https://doi.org/10.1515/9781805398233

Kholil, M., & Faradiba, S. S. (2024). Analysis Of Mathematics Summative Examination Problems In Class Vii Of Smp Negeri 2 Ngantang Using Anates. *JME (Journal of Mathematics Education)*, *9*(1), 119–126. https://doi.org/10.31327/jme.v9i1.2021

Lins, G. O. de A., Lima, N. A. da S., de Sousa, G. S., Guimarães, F. J., Frazão, I. da S., & Perrelli, J. G. A. (2021). Validity and reliability of Kessler Psychological Distress Scale for Brazilian elderly: a cross-sectional study. *Revista Brasileira de Enfermagem*, *74*. https://doi.org/10.1590/0034-7167-2020-0365

Mardiana, L., Astuti, P. A. S., Suariyani, N. L. P., & Wirawan, D. N. (2022). Disability and Lower Social Gradient Increased Risk of Depression Among Pre-Elderly in Indonesia: Finding From a Nationwide Health and Socio-Economic Survey. *Asia-Pacific Journal of Public Health*, *34*(5), 501–509. https://doi.org/10.1177/10105395221090047

Maulana, H. (2023). Analisis Kualitas Instrumen Evaluasi Pembelajaran Menggunakan Media Digitalisasi Untuk Memotivasi Hasil Belajar Peserta Didik. *Jurnal Pendidikan Bhinneka Tunggal Ika*, *1*(4). https://doi.org/10.51903/bersatu.v1i4.255

Paramartadewi, P., Bashori, K., & Hidayah, N. (2023). Psychological distress scale among caregivers in delivering care of patients with post-stroke rehabilitation: A validity and reliability testing study. *Journal of Holistic Nursing Science*, *10*(2), 103–110. https://doi.org/10.31603/nursing.v0i0.8854

Piedmont, R. L. (2023). Construct Validity. In *Encyclopedia of Quality of Life and Well-Being Research* (pp. 1332–1332). Springer International Publishing. https://doi.org/10.1007/978-3-031-17299-1_539

Pilar Matud, M., & Concepción García, M. (2019). Psychological distress and social functioning in elderly spanish people: A gender analysis. *International Journal of Environmental Research and Public Health*, *16*(3). https://doi.org/10.3390/ijerph16030341

Pitts, R. T., & Naumenko, O. (2016). *The 2014 Standards for Educational and Psychological Testing: What Teachers Initially Need to Know* (Vol. 2, Issue 1).

Sampasa-Kanyinga, H., Zamorski, M. A., & Colman, I. (2018). The psychometric properties of the 10-item Kessler Psychological Distress Scale (K10) in Canadian military personnel. *PLOS ONE*, *13*(4), e0196562. https://doi.org/10.1371/journal.pone.0196562

Sumintono, B. (2016). *Aplikasi Pemodelan Rasch pada Asesmen Pendidikan: Implementasi Penilaian Formatif (assessment for learning)*. Google Schoolar

Uddin, M. N., Islam, F. M. A., & Al Mahmud, A. (2018). Psychometric evaluation of an interview-administered version of the Kessler 10-item questionnaire (K10) for measuring psychological distress in rural Bangladesh. *BMJ Open*, *8*(6), e022967. https://doi.org/10.1136/bmjopen-2018-022967

World Health Organization. (2010). *Measuring Health and Disability: Manual for WHO Disability Assessment Schedule WHODAS 2.0*. World Health Organization.

World Health Organization. (2023). *Mental health of older adults*.

Zhu, G., Wang, Z., Guan, P., Zhu, Y., Li, J., Liu, Y., Shi, F., & Wang, S. (2021). *A Rasch Analysis of the 10-item Kessler Psychological Distress Scale (K10) in the Urban-rural fringe of China*. https://doi.org/10.21203/rs.3.rs-441385/v1