Self-Handicapping Scale: Evaluation of Psychometric Properties Among Malaysian and Indonesian University Students Using Rasch Rating Scale Model Analysis

Bambang Sumintono
Faculty of Education, Universitas Islam Internasional Indonesia, Depok, Indonesia
Mei Yui Law
Faculty of Social Science and Humanities,
Tunku Abdul Rahman University of Management and Technology, Kuala Lumpur, Malaysia, and
Novendawati Wahyu Sitasari
Fakultas Psikologi, Universitas Esa Unggul, Jakarta Barat, Indonesia

Abstract
Self-handicapping is a self-defeating behaviour that is commonly adopted by university students prior to performance evaluation for which poor performance might reflect negatively on oneself. This maladaptive coping strategy tends to occur in competitive situations and may lead to academic underachievement. The Self-Handicapping Scale (SHS) is being adopted worldwide to measure students' proclivity in making use of self-handicapping strategies. *Originality:* This instrument was first tested in the United States, but its psychometric properties have yet to be evaluated in Malaysia and Indonesia. *Purpose:* This study aimed to evaluate the psychometric properties of SHS in Malaysian and Indonesian university students' populations using the Rasch Rating Scale Model. *Methodology:* Specifically, the persons' and items' reliability and separation, rating scale's functionality, unidimensionality, item targeting, item quality, and item bias were evaluated using Winstep 4.8.1.0 on a sample of 318 Malaysian and 470 Indonesian university students. *Findings:* Both samples show good unidimensional measures. In terms of certain psychometric attributes, the Indonesian and Malaysian samples have relatively similar qualities. The adoption of SHS in measuring the self-handicapping tendency indicates the scale works well for both Indonesian and Malaysian samples. *Implications:* The findings allow researchers in Malaysia and Indonesia to confidently use the SHS to measure self-handicapping behaviours among university students. This will then enable the design and implementation of a comprehensive intervention programme aimed at reducing self-handicapping and improving the psychological well-being of these future change agents.

Keywords: Self-handicapping, Self-handicapping Scale, Psychometric Properties, Rasch Rating Scale Model, University Students

Introduction

Self-handicapping is a double-edged sword for university students; it serves to protect their self-esteem but also brings adverse consequences that undermine their academic success. This destructive behaviour was first introduced by Jones and Berglas (1978, p. 406), who defined it as "any action or choice of performance setting that enhances the opportunity to externalise failure and to internalise success". Individuals tend to utilise self-handicapping strategies to protect self-esteem and lessen the potential negativity of poor performance, especially when the task is perceived as an index of ability (Thomas & Dinnel, 2007). The

empirical evidence signified that this cognitive strategy is related to self-presentation as it occurs more frequently in the presence of observers (Brown & Kimble, 2009). The pioneer of self-handicapping, Jones and Berglas (1978) emphasised two ways students adopt self-handicapping, namely behavioural and claimed self-handicapping. Behavioural self-handicappers adopt overt destructive actions and blame the obstacles in case of failure (Jensen & Deemer, 2020). Examples include choosing unattainable goals, refusing to practice needed skills, and taking insufficient rest (Schwinger et al., 2022). Meanwhile, claimed self-handicappers make verbal statements concerning the presence of factors that hinder their performance and use the excuses made to justify their failure (Jensen & Deemer, 2020). Examples include claiming illness, insufficient preparation time, and helping others (Coudevylle et al., 2011). Compared to claimed self-handicapping that might reinforce cognitive dissonance, behavioural self-handicapping appears to be more credible but costly as it directly impacts performance negatively.

Nowadays, attention paid to self-handicapping has intensified due to its destructive consequences on academic performance and the daily functioning of university students. The highly competitive academic environment often presents situations where performance and abilities are evaluated, making it an optimal environment for self-handicapping behaviours. Additionally, the narcissistic personality traits, specifically authority, self-sufficiency, and entitlement, predicted self-handicapping (Kalyon et al., 2016). Recent studies indicated that high self-handicapping related to low academic engagement, hardiness, intrinsic motivation, and well-being (Jia et al., 2021; Jumareng & Setiawan, 2021). In addition to its positive association with procrastination, absenteeism, and underachievement (Jia et al., 2021; Schwinger et al., 2022), self-handicapping was also revealed to trigger negative emotions and self-blame (Schwinger et al., 2022). To date, although the impact of self-handicapping in the workplace is yet to be studied much, the various forms of self-handicapping identified in the work environments such as the avoidance of responsibility, poor engagement, poor communication, and poor decision-making might affect work performance negatively (Mitchell et al., 2018).

Since the phenomenon of self-handicapping is complex, common in the academic context, and utilised by both high- and low-achievement students (Boruchovitch et al., 2022), there is an urgent need to measure self-handicapping accurately before effective intervention programmes can be designed and implemented to reduce the self-handicapping of university students. Addressing such deficit behaviour is in line with the United Nations's Sustainable Development Goal (SDG) 3- Good health and well-being. Given that self-handicapping is linked to adverse academic, behavioural, and mental health outcomes, reducing self-handicapping may enhance well-being. Reducing self-handicapping behaviour is vital in promoting well-being because it helps university students deal with academic challenges constructively rather than creating obstacles or excuses that lead to negative outcomes or vicious cycles.

Literature Review

In the past four decades, the most widely used self-reported measure for measuring self-handicapping is the original 25-item Self-Handicapping Scale (SHS) developed by Jones and Rhodewalt (1982). In addition to the original SHS, the 10-item *Self-Handicapping Scale*

(*Strube, 1986),* 14-item Self-Handicapping Scale (Rhodewalt, 1990; Zuckerman et al., 1998), and Academic Self-Handicapping Scale (ASHS) (6-item) (Midgley & Urdan, 1995) are also commonly used for measuring self-handicapping. Except for the ASHS, the remaining instruments stated above are abbreviated versions of the original 25-item SHS.

Following the introduction of the 25-item SHS (Jones & Rhodewalt, 1982), many studies have been conducted to ascertain the factor structure that best represents self-handicapping (Clarke & Maccann, 2016; McCrea et al., 2008; Rhodewalt, 1990; Strube, 1986; Zuckerman et al.,1998). These validation studies revealed both one-factor and multifactorial structures. Strube (1986) conducted a principal component analysis to assess the self-handicapping of university students. The outcome suggested a one-factor, 10-item scale but acknowledged moderate heterogeneity among the items retained. In 1990, Rhodewalt re-analysed the factor structure of the 25-item SHS and noticed that only 14 items loaded significantly on one out of two factors, namely excuse-making and effort. In addition, the exploratory factor analysis performed by Zuckerman et al. (1998) proposed a one-factor 14-item scale. However, McCrea et al. (2008) later suggested a two-factor structure, which is claimed self-handicapping and behavioural self-handicapping. In 2016, Clarke and Maccann published another validation study using data from 484 university students. The findings revealed that only 13 items loaded significantly on one of two factors: internal self-handicapping and external self-handicapping.

Although the original 25-item SHS pioneered self-handicapping tools worldwide, this is not the case in Indonesia and Malaysia. In Indonesia, the SHS validated by Clarke and MacCann (2016) and the ASHS were adapted to measure the self-handicapping of university students (Ariani, 2022; Jumareng & Setiawan, 2021), whereas in Malaysia, the ASHS was adopted to measure the self-handicapping of university students (Cheng & Law, 2015). Given that both of these countries practise a collectivistic culture, university students might engage in self-handicapping to mitigate the impact of failure on their social relationships. The limited research on self-handicapping may be due to the lack of a validated SHS in these countries. It is noteworthy that despite the criticism of SHS, such as the inability to separate between actual self-handicapping and post hoc attributions and the lack of domain specificity (Török et al., 2018), it has remained popular for measuring self-handicapping (Clarke & MacCann, 2016; Török et al., 2018). Hence, this study examined the psychometric properties of the original 25-item SHS. Considering that the validation studies above utilised statistical analysis based on the Classical Test Theory, it would be instructive to evaluate the SHS using the Rasch measurement model (Ratnaningsih et al., 2024).

Rasch rating scale model ensures accuracy when measuring latent traits to assess the items' compliance with the scale (Tennant & Conaghan, 2007). This psychometric analysis is useful for identifying biases based on demographic characteristics such as gender and nationality (Adams et al., 2022). Traditionally, the psychometric properties might not include personal ability and item difficulty, therefore, the accuracy of the measurement could be compromised (Hobart & Cano, 2009). By adopting the Rasch measurement model in assessing SHS, the precision of assessment could be enhanced because the process does not rely on the raw score but applies probabilistic and uses a non-linear function that results in an equal interval scale (Liu & Lim, 2020). In this study, the psychometric properties of the original 25-item SHS

are evaluated using two data sets, namely Malaysian and Indonesian university students. Specifically, the psychometric properties are evaluated at the instrument level (person and item reliability, person separation, floor and ceiling effects, unidimensionality, rating scale functioning, and item targeting) as well as at the item level (item fit statistics, item polarity, and item bias based on gender).

Methodology

This section explains details such as the research design, participants, instrument, procedure, and data analysis.

Research Design

The quantitative method with a cross-sectional design was employed because quantifying items is required to achieve the purpose of scale validation. The study employed the SHS, whereby the statements were measured using a Likert rating scale, which enabled the researchers to quantify the scores. The cross-sectional design was adopted as the self-handicapping tendency of the respondents only needed to be measured once to validate the SHS. Additionally, this design enabled the researchers to collect a large amount of data in a short period of time. In this study, the data were collected using Google Forms. The participants from both countries were able to complete the questionnaire at their convenience. The Google Forms's links were shared face-to-face on campus, through social media platforms, and on the university intranet. The data collection period lasts for 5 months, from April to September 2022.

Participants

The sample gathered through Google Forms comprised 517 Indonesian and 342 Malaysian university students. During data cleaning using Winstep version 4.4.7, although no outliers were detected, the misfit responses (47 and 24 responses from Indonesia and Malaysian data respectively) were excluded for having an Outfit MNSQ index larger than 2.0 (Boone & Staver, 2020). After data cleaning, 470 Indonesian and 318 Malaysian university students were retained, indicating adequate data stability. According to Linacre (1994), a sample size of 250 is sufficient to provide 99% confidence and logit estimation stability of ±0.5 logit, which shows good variation from the population. In this study, undergraduates in Indonesia and Malaysia were recruited regardless of their age, gender, educational level, year of study, and type of university. However, university students who deferred were excluded. Table I presents the demographic profiles of the participants. In both samples, more female university students participated in the study (57% for Indonesian and 63% for Malaysian, respectively). Also, more than half of the participants were in between 20 and 22 (85% for Indonesian and 65% for Malaysian, respectively). For the Indonesian sample, nearly all of them (99%) registered in the Bachelor's degree programme.

Table I. Profile of the Participants

Instrument

Self-Handicapping Scale (SHS) developed by Jones and Rhodewalt (1982) was adopted to measure university students' propensity to adopt self-handicapping strategies. The items were inspected to ascertain that they were designed to measure the self-handicapping tendency of students. The original English version was adopted to measure the self-

handicapping tendency of Malaysian university students. The use of the English version in Malaysia is appropriate because English is an official second language in Malaysia and is widely spoken in the country. To measure the self-handicapping tendency of Indonesian students, the SHS was translated into the national language (*Bahasa Indonesia*) by the subject-matter experts. The language expert from Universitas Esa Unggul translated the SHS from English to Bahasa Indonesia. Another psychology expert then reviewed the translated SHS to ensure the items remained valid. The statements and the item codes for SHS in the English and Bahasa Indonesia versions are available in the appendix. Each statement is measured using a Five-point Likert rating scale, ranging from Very Much Disagree (1), Disagree (2), Somewhat Disagree/Agree (3), Agree (4) to Very Much Agree (5). Eight out of 25 items (Items 3, 5, 6, 10, 13, 20, 22, and 23) are negatively worded items that have been reversed-coded before the analysis. The higher the score, the greater the self-handicapping.

Procedure

The study obtained its ethical clearance with Reference number: TAR UC/EC/2022/02-3 on April 2022 from the Research Development and Management Committee in the university in which the principal investigator works. Upon obtaining ethical approval, the links of the Google Forms were shared with university students in Malaysia and Indonesia. Informed consent was obtained before the data collection. For informed consent, the respondents received a detailed explanation of the purpose of the study, procedures, potential benefits and risks of participation, confidentiality, and their right to withdraw from the study. The respondents were allowed to enquire about the items stated in the Google Forms through email. The responses were automatically saved in spreadsheet in Google Drive.

Data Analysis

The data analysis of this study was performed using Winstep version 4.4.7, in which the Rasch Rating Scale Model was applied to analyse the raw data. This approach begins by counting the raw data, which was ordinal data type, such as frequencies, calculating the odds probability, and then transforming it into equal-interval data through a logarithm function (Boone & Staver, 2020). The outcome of the process was an equal-interval scale at the item level known as logits (an abbreviation for logarithm odds unit), which indicates the item's difficulty level. In this context, higher item logit values indicate items are perceived as more difficult to endorse, while lower values suggest items are more readily agreed upon by respondents. The outcome of this process also informed about the psychometric attributes at the instrument and item level, a unique product of Rasch analysis. This study used the psychometric criteria described in Akhtar and Sumintono (2023) and Liu and Lim (2020) to inform the instrument quality of SHS.

Results

Reliability and Dimensionality

The Winstep software was utilised to analyse these two data sets of SHS independently. Table II presents the instrument and item psychometric attributes regarding item logit, reliability, and scale's dimensionality. As shown in the table, the mean measure (logit) of the items was 0.00 with a small standard error, indicating there were no outliers in the SHS scale and good item precision. Both standard deviation indices are above 0.5, indicating a good spread of item difficulty (0.83 for Indonesian and 0.71 for Malaysian). Additionally, both raw variance indices are above the minimum threshold of 20%; the Eigenvalue in the

Indonesian data is higher than 3, indicating that one item came from another dimension (SH6), but this does not happen in the Malaysian data. Overall, the average outfit mean-square was close to the expected value of 1, signifying a uniform fit to the model with a minute standard error (0.04). A separation index greater than 3 and a reliability index greater than 0.9 suggest satisfactory reliability (Kök & Fisher, 2007). These two indices indicate excellent internal consistency for both data sets.

Table II. Summary of Reliability and Dimensionality of SHS

Rating Scale Analysis

The rating scale analysis of SHS is shown in Table III. The table presents the distribution and performance of each response category for both data sets. The average measure increased monotonically for both samples, from -1.01 logits in category '1' to +0.96 logits in category '5' in the Indonesian sample, whereas in the Malaysian sample, it starts from -0.87 logits to +0.78 logits. This indicated properly functioning categories, with each successive rating category representing a higher latent trait level, as expected in a well-structured rating scale (Liu & Lim, 2020; Ratnaningsih et al., 2024). Additionally, the Outfit Mean Square (MNSQ) values of both data sets were close to the ideal value of 1.00 across all categories. These MNSQ values supported the notion that the rating scale categories operated effectively and that respondents' selections across the scale were coherent with the Rasch model's expectations.

Table III. *Rating Scale Model Analysis Category Statistics for SHS*

The Step values representing the transition points between adjacent categories did not indicate any disordered thresholds for the Indonesian sample. However, for the Malaysian sample, the transition between category '3' (-0.30 logits) and '4' (+0.29 logits) was less distinct, below the cut-off point of 1.4. However, as shown in Figure 1, each category has its peak, which means the scale could differentiate between varying levels of self-handicap. These findings align with Kök and Fisher's (2007) recommendations for an optimally functioning rating scale. The consistent increase in average measures across the categories further supported the scale's capacity to effectively capture the gradations of self-handicapping.

Figure 1. Category probability curves of the SHS from two samples

Item Fit Statistics Indices

The salient quality of Rasch model analysis is that it can provide information about item quality, as shown in Table IV. The logit values, indicative of item difficulty, spanned from -1.66 to +1.83 logits (SD= 0.83) for the Indonesian sample and -1.32 to +1.78 logits (SD=0.71) for the Malaysian sample. These values signify that all items fall within an acceptable range without any outliers, thus ensuring a cohesive scale. In addition, by focusing on the item fit statistics, which typically encompass mean squares (MNSQ), t-statistics (ZStds), and point measure correlations (PT.Meas.Corr.) (Boone & Staver, 2020). For the MNSQ values, outfit values were mostly within the acceptable bounds of 0.5 and 1.5 (Bond & Fox, 2015). Specifically, all except one item (SH20 in the Indonesian sample) were close to the ideal value of 1.0. This indicates that most items fit the model well. The

PT.Meas.Corr. further validated the item quality, with most items showing positive correlations, except for three items in the Indonesian sample (SH10, SH13, and SH22) and one item in the Malaysian sample (SH13). These positive values corroborated the alignment of each item with the underlying latent variable, which is self-handicapping. A consistent PT.Meas.Corr. indicates that the items measure the intended construct effectively.

Table IV. *Item Fit Statistics indices of SHS*

Item Targeting

As mentioned in the Participants section, there was no outlier response in both data sets, the remaining 788 respondents provided acceptable responses on their perception of self-handicapping. Moreover, the respondents' responses were analysed to see how well their personal abilities and item difficulty levels matched; identification of floor and ceiling effects, where respondents' answers consistently skewed towards lower or higher scores, indicating potential targeting issues. The evidence of targeting is shown in Figure 2, which is the Wright map for both data sets. The left column of each Wright map shows the range of self-handicapping scores, from lower scores at the bottom to higher scores at the top, whereas item difficulties appear in the three columns to the right (represented by XX). The SHS instrument demonstrated a moderate level of item difficulty for both samples. Items located to the left of the centre (bottom p=50%), representing lower difficulty levels, had a probability exceeding 50% in the bottom category of the scale; while the one column of items in the far right (top p=50%), representing higher difficulty levels, had a probability below 50% at the top category of the scale. As can be seen, nearly the entire range of person columns from both data sets matched the three item columns, indicating that the item difficulties were well matched with personal abilities.

Figure 2. Wright map of SHS of both samples

Differential Item Functioning (DIF)

The Rasch model can examine the item bias in the SHS based on the demographic variables. In view of DIF analysis being sensitive to sample size (Boone & Staver, 2020), only gender can be examined in this study. This analysis is vital in ensuring test fairness and validity across genders; hence, the test scores reflect actual differences instead of biases in self-handicapping. As shown in Table V, only one item in the Indonesian sample (SHS7) had DIF, whereas none in the Malaysian sample. For SHS7, the DIF Contrast, Rasch-Welch t, and Mantel-Haenszel probability exceeded the acceptable range. This implies that the remaining items functioned well.

Table V. *Differential Item Functioning (DIF) based on Gender*

Discussion

The present study aimed to validate the SHS using the data obtained from Indonesian and Malaysian university students. As mentioned earlier, the SHS has been tested by many researchers (Clarke & Maccann, 2016; McCrea et al., 2008; Rhodewalt, 1990; Zuckerman et al., 1998), but these studies have mostly been limited to Western societies, yielding mixed results and used the classical test theory approach. To date, the 25-item SHS has not been validated in either Indonesia or Malaysia using the Rasch model, as comprehensive evidence

has shown this measurement model is robust in validation (Tennant & Conaghan, 2007) and provides comprehensive information about the psychometric attributes of the instrument (Hobart & Cano, 2009).

The findings of this study indicate that the SHS scale is a valid tool for measuring self-handicapping, in both Indonesian and Malaysian samples. The scale's unidimensionality, determined through principal component analysis of residuals, revealed that a primary dimension captured over the minimum level, 34.5% for the Indonesian sample and 35.6% for the Malaysian sample, with minimal noise levels that did not affect measurement quality (Liu & Lim, 2020). All reliability indices, such as item reliability from both data sets, were excellent, indicating that the data sets from both countries were large enough for the SHS to demonstrate good internal consistency (Ratnaningsih et al., 2024). The item separation index (13.76 and 11.05 for the Indonesian and Malaysian samples, respectively) indicates good item difficulty grouping in the scale. This implies that the SHS resulted in very satisfactory reliability measures and the ability of the scale to distinguish between different levels of self-handicapping.

In the rating scale analysis of the SHS, monotonic assumptions were fulfilled with an increase in measures in both countries' datasets (Van Zile-Tamsen, 2017). However, concerning step calibration, the Malaysian sample appeared to need to collapse into four scales to achieve better rating functionality (Fayers & Machin, 2007). The rating scale in both countries' samples worked well for all categories, indicating that the Rasch analysis is useful in identifying the instrument rating scale's effectiveness (Boone & Staver, 2020).

The unique approach of the Rasch model allowing for a comprehensive investigation at the item level. Interestingly, nearly all items had good psychometric attributes, except for SH20 ("I would rather not take any drug that affected my ability to think clearly") in the Indonesian sample, which requires revision. Regarding polarization, indicated by negative PT.Meas.Corr., it included one item in both countries (SH13) and two items (SH10 and SH22) in the Indonesian sample. All these items are negatively worded items, which means they potentially confuse respondents and thus require further revision (DiStefano & Motl, 2009; Sliter & Zickar, 2014).

Regarding person-item targeting, the findings indicated that the SHS did not exhibit floor and ceiling effects. Hence, the scale functions effectively in measuring the personal abilities of samples. The mean of person measure (-0.01 logits for the Indonesian sample and -0.03 logits for the Malaysian sample) was similar to the item measure (0.00 logits), indicating that the SHS for both country samples tended to have a similar difficulty level compared to personal abilities. The standard deviation of the person logits (0.35 and 0.37 for the Indonesian and Malaysian samples, respectively) was lower than the item logits (0.83 and 0.71), suggesting a wider variation of item difficulty compared to personal abilities, which indicates good targeting of the instrument (Liu & Lim, 2020).

For DIF analysis only gender could be used, as the number of participants in the sub-groups for the other three demographics was not varied enough (Adams et al., 2022). The DIF analysis revealed that one item in the Indonesian sample (SH7: "I feel very anxious before

an exam or performance") was more difficult for male respondents to agree with compared to female respondents. This possibly reflects the situation of the sample, where Indonesian female respondents tend to be more ready than their male counterparts, which means information about the date and reminding it is important to announce regularly. This indicates that the other items were fair, with negligible effects on measurement quality.

Limitations and Recommendations for Future Research

There are a few limitations to discuss in this study. The adoption of convenience sampling resulted in unequal gender distribution, with 37% of males and 63% of females Malaysian university students, and unequal distribution of educational level, with 99% of Indonesian respondents being Bachelor's degree students. However, undeniably, the findings provided valuable insights into the psychometric evaluation of SHS among Malaysian and Indonesian university students. Future studies might employ random sampling with bigger representative sample, which can produce more generalisable results. The other limitation of this quantitative study using self-reported measures is that some of the respondents might provide socially desirable responses on self-handicapping tendency to portray a positive self. This can then reduce the accuracy of the responses. Although this drawback cannot be eliminated, it can be reduced by administering an instrument to assess social desirability, including positive and negative impressions, concurrently with the administration of SHS. Lastly, it would be informative to analyse item bias based on country of origin, given that the samples are from two countries. A bigger sample in each subgroup is needed if researchers intend to perform DIF analysis in other demographics.

Conclusion

Overall, the findings of this study have shown that the SHS is a valid and reliable assessment tool for measuring the self-handicapping behaviours of university undergraduates in both Malaysia and Indonesia. The Rasch analysis demonstrated that the compliance of the scale items to the overall scale is acceptable for both samples. With these findings, researchers or academicians might have higher confidence in adopting this scale as a measurement tool to assess undergraduates' self-handicapping behaviours. This implies that using the scale can provide accurate information about self-handicapping, which is essential for designing effective intervention programmes to reduce self-handicapping among university students in Malaysia and Indonesia. These intervention programmes are then expected to enhance academic performance in the long run. By accurately measuring and reducing self-handicapping among university students, we can shift societal attitudes towards valuing hard work and recognising the importance of mental health in the academic context, rather than focusing solely on innate ability.

References

1. Adams, D., Chuah, K.M., Sumintono, B. and Mohamed, A. (2022), "Students' readiness for e-learning during the COVID-19 pandemic in a South-East Asian university: A Rasch analysis", *Asian Education and Development Studies*, Vol. 11 No. 2, pp. 324-329, https://doi.org/10.1108/AEDS-05-2020-0100.

2. Akhtar, H. and Sumintono, B. (2023), "A Rasch analysis of the International Personality Item Pool Big Five Markers Questionnaire: Is longer better?", *Primenjena Psihologija*, Vol. 16 No. 1, pp. 3-28, https://doi.org/10.19090/pp.v16i1.2401.

3. Ariani, D.W. (2022), "Why externalize failure? Social goals, engagement and self-handicapping in the concern of eastern cultural community", *Cogent Education*, Vol. 9 No. 1, pp. 1-20, https://doi.org/10.1080/2331186X.2022.2140541.

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

5. Boone, W.J. and Staver, J.R. (2020), *Advances in Rasch Analyses in the Human Sciences*, Springer.

6. Boruchovitch, E., Rufini, S.E., Ganda, D.R., Miranda, L.C. and De Almeida, L.S. (2022), "Self-handicapping strategies in educational context: Construction and validation of the Brazilian Self-Handicapping Strategies Scale (EEAPREJ)", *Psicologia: Reflexão e Crítica*, Vol. 35 No. 8, pp. 1-14, https://doi.org/10.1186/s41155-022-00210-6

7. Brown, C.M. and Kimble, C.E. (2009), "Personal, interpersonal, and situational influences on behavioral self-handicapping", *The Journal of Social Psychology*, Vol. 149 No. 6, pp. 609-626.

8. Cheng, S.K. and Law, M.Y. (2015), "Mediating Effect of Self-Esteem in the Predictive Relationship of Personality and Academic Self-Handicapping", *American Journal of Applied Psychology*, Vol. 4 No. 3-1, pp. 51-57, https://doi.org/10.11648/j.ajap.s.2015040301.19

9. Clarke, I.E. and MacCann, C. (2016), "Internal and external aspects of self-handicapping reflect the distinction between motivations and behaviours: Evidence from the Self-handicapping Scale", *Personality and Individual Differences*, Vol. 100, pp. 6-11, https://doi.org/10.1016/j.paid.2016.03.080

10. Coudevylle, G.R., Gernigon, C. and Ginis, K.A.M. (2011), "Self-esteem, self-confidence, anxiety and claimed self-handicapping: A mediational analysis", *Psychology of Sport and Exercise*, Vol. 12 No. 6, pp. 670-675.

11. DiStefano, C. and Motl, R.W. (2009), "Personality correlates of method effects due to negatively worded items on the Rosenberg Self-Esteem scale", *Personality and Individual Differences*, Vol. 46 No. 3, pp. 309-313, http://dx.doi.org/10.1016/j.paid.2008.10.020

12. Fayers, P.M. and Machin, D. (2007), *Quality of Life: The Assessment, Analysis, and Interpretation of Patient-Reported Outcomes*, 2nd ed., John Wiley & Sons.

13. Jumareng, H. and Setiawan, E. (2021), "Self-esteem, adversity quotient and self-handicapping: Which aspects are correlated with achievement goals?", *Cakrawala Pendidikan*, Vol. 40 No. 1, pp. 147-157, https://doi.org/10.21831/cp.v40i1.37685

14. Hobart, J. and Cano, S. (2009), "Improving the evaluation of therapeutic interventions in multiple sclerosis: The role of new psychometric methods", *Health Technology Assessment*, Vol. 13, pp. 1-200. https://doi.org/10.3310/hta13120

15. Jensen, L.E. and Deemer, E.D. (2020), "Attachment style and self-handicapping: The mediating role of the imposter phenomenon", *Social Psychology of Education*, Vol. 23, pp. 1259-1276, https://doi.org/10.1007/s11218-020-09580-0

16. Jia, J., Wang, L.L., Xu, J.B., Lin, X.H., Zhang, B. and Jiang, Q. (2021), "Self-handicapping in Chinese medical students during the COVID-19 pandemic: The role of academic anxiety, procrastination and hardiness", *Frontiers in Psychology*, Vol. 12, pp. 1-12, https://doi.org/10.3389/fpsyg.2021.741821

17. Jones, E.E. and Berglas, S. (1978), "Control of attributions about the self through self-handicapping strategies: The appeal of alcohol and underachievement", *Personality and Social Psychology Bulletin*, Vol. 4, pp. 200-206.

18. Jones, E.E. and Rhodewalt, F. (1982), *The Self-Handicapping Scale*, Princeton, New Jersey: Princeton University.

19. Kalyon, A., Dadandi, I. and Yazici, H. (2016), "The relationships between self-handicapping tendency and narcissistic personality traits, anxiety sensitivity, social support, academic achievement", *Düşünen Adam: Journal of Psychiatry and Neurological Sciences,* Vol. 29 No. 3, pp. 237-246, https://doi.org/10.5350/DAJPN2016290305

20. Kök, G. and Fisher, M.L. (2007), "Demand estimation and assortment optimization under substitution: Methodology and application", *Operations Research*, Vol. 55 No. 6, pp. 1001-1021.

21. Linacre, J.M. (1994), "Sample size and item calibration stability", *Rasch Measurement Transactions*, Vol. 7, p. 328.

22. Liu, V.Y. and Lim, S.M. (2020), "A psychometric evaluation of the brief resilience scale among tertiary students in Singapore", *Asia Pacific Journal of Education*, Vol. 42 No. 3, pp. 464-477, https://doi.org/10.1080/02188791.2020.1845120

23. McCrea, G.L., Miaskowski, C., Stotts, N.A., Macera, L., Hart, S.A. and Varma, M.G. (2008), "Self-report measures to evaluate constipation", *Alimentary Pharmacology & Therapeutics*, Vol. 27, pp. 638-648.

24. Midgley, C. and Urdan, T. (1995), "Predictors of middle school students' use of self-handicapping strategies", *Journal of Early Adolescence*, Vol. 15, pp. 389-411.

25. Mitchell, J., Decker, P., Durand, R., Durand, J. and Moriki, Y. (2018), "Experiencing self-handicapping behavior at work", *Journal of Business Disciplines,* Vol. 2 No. 1, pp. 69-85.

26. Ratnaningsih, I.Z., Prihatsanti, U., Prasetyo, A.R. and Sumintono, B. (2024), "Validation of the Indonesian version of the psychological capital questionnaire (PCQ) in higher education: A Rasch analysis", *Journal of Applied Research in Higher Education*. Vol. ahead-of-print, https://doi.org/10.1108/JARHE-10-2023-0480

27. Rhodewalt, F. (1990), Self-handicappers: Individual differences in the preference for anticipatory, self-protective acts, In: R.L. Higgins, C.R. Snyder and S. Berglas, eds. *Self-handicapping: The paradox that isn't*, Plenum Press, pp. 69-106.

28. Schwinger, M., Trautner, M., Pütz, N., Fabianek, S., Lemmer, G., Lauermann, F. and Wirthwein, L. (2022), "Why do students use strategies that hurt their chances of academic success? A meta-analysis of antecedents of academic self-handicapping", *Journal of Educational Psychology*, Vol. 114 No. 3, pp. 576-596.

29. Sliter, K.A. and Zickar, M.J. (2014), "An IRT Examination of the Psychometric Functioning of Negatively Worded Personality Items", *Educational and Psychological Measurement*, Vol. 74 No. 2, pp. 214-226. https://doi.org/10.1177/0013164413504584

30. Strube, M.J. (1986), "An analysis of the self-handicapping scale", *Basic and Applied Social Psychology*, Vol. 7 No. 3, pp. 211-224.

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

32. Thomas, T. and Dinnel, D.L. (2007), "Is self-worth protection best regarded as intentional self-handicapping behaviour or an outcome of choking under pressure?", *Educational Psychologist*, Vol. 27 No. 4, pp. 509-531, https://doi.org/10.1080/01443410601159910

33. Török, L., Szabó, Z.P. and Tóth, L. (2018), "A critical review of the literature on academic self-handicapping: Theory, manifestations, prevention and measurement", *Social Psychology of Education*, Vol. 21, pp. 1175-1202, https://doi.org/10.1007/s11218-018-9460-z

34. Van Zile-Tamsen, C. (2017), "Using Rasch Analysis to inform rating scale development", *Research in Higher Education*, Vol. 58, pp. 922-933, https://doi.org/10.1007/s11162-017-9448-0

35. Zuckerman, M., Kieffer, S.C. and Knee, C.R. (1998), "Consequences of self-handicapping: Effects on coping, academic performance, and adjustment", *Journal of Personality and Social Psychology*, Vol. 74, pp. 1619-1628.

Table I. Profile of the Participants

Demographic	**Indonesian**	**Malaysian**
	Frequency (%)	
Gender		
Male	205 (43)	117 (37)
Female	265 (57)	201 (63)
Age		
17-19	30 (6)	83 (26)
20-22	399 (85)	206 (65)
Above 22	41 (9)	29 (9)
Educational level		
Diploma	6 (1)	126 (39)
Bachelor degree	464 (99)	192 (61)

Table II. Summary of Reliability and Dimensionality of SHS

	Indonesian Sample	Malaysian Sample
N	470	318
Item logits		
Mean	0.00	0.00
Standard Deviation, SD	0.83	0.71
Standard Error, SE	0.17	0.14
Outfit Mean Square		
Mean	0.99	1.00
SE	0.04	0.04
Separation	13.76	11.05
Reliability	0.99	0.99
Raw variance	34.5%	35.6%
Eigenvalue first contrast	3.7	2.4

Table III. *Rating Scale Model Analysis Category Statistics for SHS*

Category	Indonesian sample (N=470)			Malaysian sample (N=318)		
	Average Measure	Outfit MNSQ	Step	Average Measure	Outfit MNSQ	Step
1	-1.01	1.04	NONE	-0.87	1.04	NONE
2	-0.58	0.94	-2.20	-0.41	0.97	-1.53

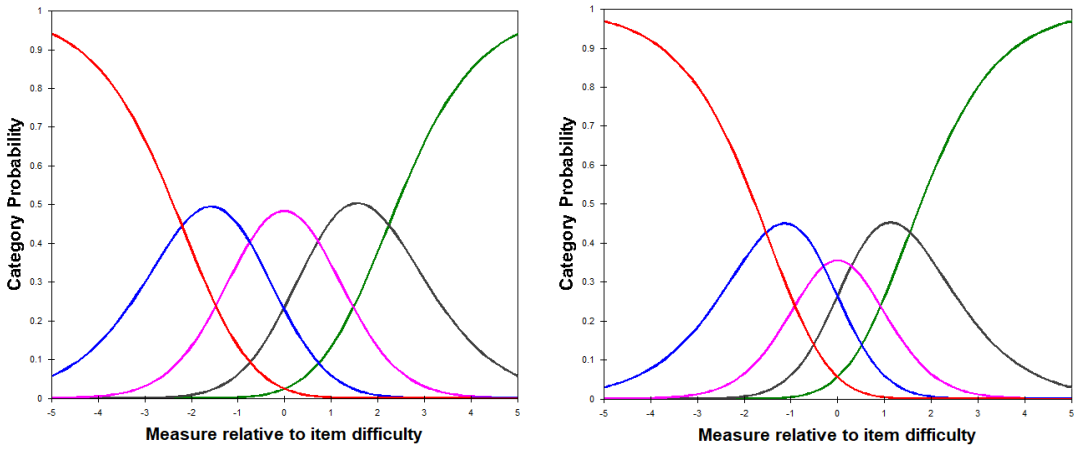
3	-0.03	0.95	-0.75	-0.04	1.02	-0.30
4	+0.58	0.96	+0.72	+0.41	0.97	+0.29
5	+0.96	1.08	+2.23	+0.78	1.02	+1.54

Table IV. *Item Fit Statistics indices of SHS*

Item	Indonesian sample (N=470)			Malaysian sample (N=318)		
	Logit	Outfit MNSQ	PT.Meas. Corr.	Logit	Outfit MNSQ	PT.Meas. Corr.
SH1	0.56	0.75	0.43	0.55	0.74	0.46
SH2	0.79	0.97	0.33	0.16	0.92	0.49
SH3	0.37	1.12	0.03	0.00	1.17	0.13
SH4	0.84	1.21	0.46	0.42	1.14	0.44
SH5	1.83	0.94	0.13	1.09	0.73	0.38
SH6	1.39	0.88	0.14	0.65	0.83	0.23
SH7	-0.95	1.02	0.45	-1.05	1.16	0.36
SH8	-0.63	1.14	0.42	-0.88	1.07	0.38
SH9	0.07	0.96	0.54	-0.28	1.05	0.35
SH10	0.59	1.11	***-0.03***	0.85	0.78	0.08
SH11	-1.66	0.77	0.04	-1.32	1.04	0.08
SH12	-0.63	1.02	0.33	-0.44	0.86	0.24
SH13	0.17	1.27	***-0.30***	0.34	0.90	***-0.03***
SH14	-1.48	0.88	0.22	-0.93	0.61	0.18
SH15	0.05	1.33	0.49	0.15	1.43	0.45
SH16	-0.99	0.73	0.34	-1.08	0.81	0.42
SH17	-0.87	0.78	0.26	0.10	1.30	0.18
SH18	0.10	0.79	0.49	0.23	0.73	0.54
SH19	-0.13	1.16	0.47	-0.10	1.10	0.44
SH20	0.69	***1.66***	0.20	1.78	1.37	0.27
SH21	0.03	0.92	0.27	0.18	1.08	0.45
SH22	0.08	0.87	***-0.11***	0.28	1.38	0.30
SH23	0.67	0.82	0.13	-0.02	0.95	0.42
SH24	-0.46	0.83	0.25	-0.10	0.92	0.07
SH25	-0.45	0.82	0.47	-0.59	0.98	0.51

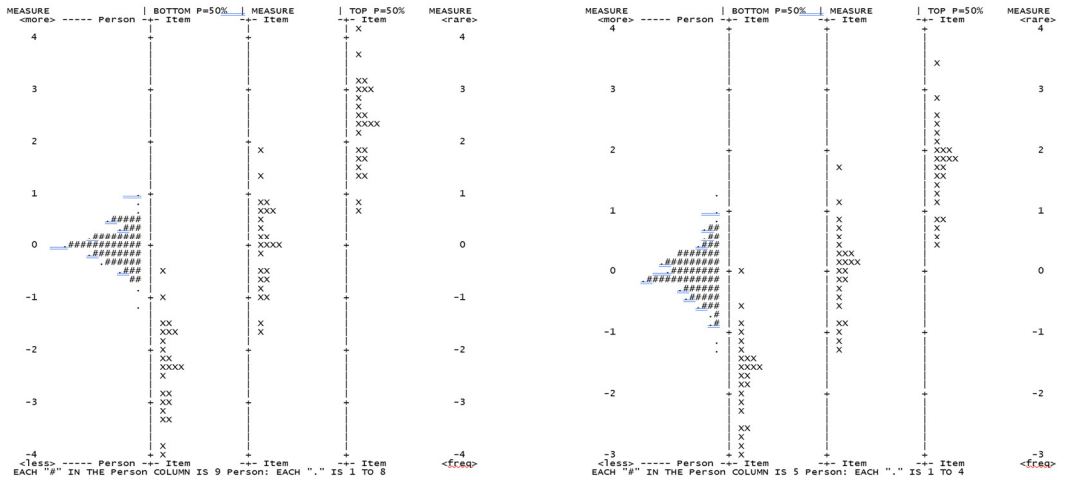
Table V. *Differential Item Functioning (DIF) based on Gender*

Data set & Item		DIF Contrast	Rasch-Welch t	Prob.
Indonesian sample	SHS7	0.77	-6.56	0.000



a. Indonesian sample b. Malaysian sample

Figure 1. Category probability curves of the SHS from two samples



a. Indonesian sample b. Malaysian sample

Figure 2. Wright map of SHS of both samples

Appendix I. SHS Items in English and *Bahasa Indonesia*

Code	Item
SH1	When I do something wrong, my first reaction is to blame the situations. Ketika saya melakukan sesuatu yang salah, reaksi pertama saya adalah menyalahkan situasi.
SH2	I tend to put things off until the last moment. Saya cenderung menunda sesuatu sampai detik terakhir.
SH3**	I tend to overprepare (i.e. prepare more than enough) when I have an exam or performance. Saya cenderung mempersiapkan diri secara berlebihan (yaitu mempersiapkan lebih dari cukup) ketika saya menghadapi ujian atau saat harus perform.
SH4	I think I feel sick more frequent than others. Saya pikir saya merasa lebih sering sakit dari pada yang lain.
SH5**	I always try to do my best. Saya selalu mencoba untuk melakukan yang terbaik.
SH6**	Before I sign up for a course or engage in any important activity, I make sure I have the proper preparation or background. Sebelum saya mendaftar atau terlibat dalam aktivitas penting apa pun, saya memastikan bahwa saya memiliki persiapan yang matang
SH7	I feel very anxious before an exam or performance. Saya merasa cemas sebelum ujian atau saat harus perform.
SH8	I am easily distracted by noises or own thoughts when reading. Saya mudah terganggu dengan suara atau pikiran saya sendiri ketika membaca.
SH9	I seldom participate actively in competitive activities so it won't hurt too much if I lose or do poorly. Saya jarang berpartisipasi secara aktif dalam kompetisi sehingga tidak akan terlalu sakit jika saya kalah atau berprestasi buruk.
SH10**	I prefer to be respected for doing my best than admired for my in-born abilities. Saya lebih suka mendapatkan penghargaan atas prestasi karena saya melakukan yang terbaik daripada dikagumi karena bakat saya.
SH11	I would do a lot better if I tried harder. Saya akan melakukan jauh lebih baik jika saya berusaha lebih keras.
SH12	I prefer small rewards in the present than big rewards in the uncertain future. Saya lebih suka hadiah kecil di masa sekarang dari pada hadiah besar di masa depan yang tidak pasti.
SH13**	I generally hate to be in any condition but "at my best". Saya biasanya benci ketika berada di luar zona aman saya.
SH14	Someday I might be able to organize my thinking appropriately. Suatu hari nanti saya mungkin bisa mengatur pemikiran saya dengan matang.
SH15	I sometimes enjoy being sick for a day or two because it takes off the stress. Saya kadang-kadang menikmati sakit selama satu atau dua hari karena menghilangkan stress.
SH16	I would do much better if I did not let my emotions distract me. Saya akan melakukan jauh lebih baik jika saya tidak teralihkan dengan emosi saya.
SH17	When I perform poorly, I often comfort myself by thinking that I am good at other aspects. Ketika kinerja saya buruk, saya sering menghibur diri dengan berpikir bahwa saya baik dalam aspek lain.
SH18	I am prone to find excuse when I don't meet other's expectations. Saya cenderung mencari alasan ketika saya tidak memenuhi harapan orang lain.
SH19	I often think I am out of luck in sports, games, and other measures of talent.

	Saya sering berpikir saya kurang beruntung dalam olahraga, permainan, dan bakat lainnya.
SH20**	I would rather not take any drug that affected my ability to think clearly. Saya lebih suka tidak mengonsumsi obat apa pun yang mempengaruhi kemampuan saya untuk berpikir jernih.
SH21	I stuff myself in food and drink more often than I should. Saya makan dan minum lebih sering dari yang seharusnya.
SH22**	I try to sleep as much as possible the night before something important is coming up, like an exam. Saya mencoba untuk tidur sebanyak mungkin sebelum melakukan sesuatu yang penting seperti ujian.
SH23**	I never let emotional problems in one part of my life interfere with other things in my life. Saya tidak pernah membiarkan masalah emosi saya mengganggu hal lain dalam hidup saya.
SH24	Usually, when I get anxious about doing well, I end up doing better. Biasanya, saya akan melakukan dengan lebih baik, ketika saya mencemaskan hal tersebut.
SH25	Sometimes I get so depressed that even easy tasks become difficult. Kadang-kadang saya menjadi sangat tertekan sehingga tugas-tugas yang mudah pun menjadi sulit.

** Reverse-coded item