Primenjena psihologija
Vol. 18, No. 2, pp. 227–254, 2025

Research Article

Psychometric evaluation of an Indonesian version of the Brief Self-Control Scale: Item calibration using polytomous Rasch models

Diana Mutiah [1] ✉, Bobby Suwandi [2], and Muhammad Dwirifqi Kharisma Putra [3]

[1] *Faculty of Psychology, Universitas Islam Negeri Syarif Hidayatullah Jakarta, Indonesia*
[2] *Faculty of Psychology, Universitas Pancasila, Indonesia*
[3] *Faculty of Psychology, Universitas Gadjah Mada, Indonesia*

ABSTRACT

This study aimed to evaluate the psychometric properties of the 10-item Indonesian version of the Brief Self-Control Scale (BSCS). It used the polytomous Rasch model, which enables more detailed analysis, including differential item functioning (DIF) analysis. The participants in this study were 1001 Indonesian high school students. We found that the partial credit model (PCM) was a better fit than the rating scale model. Furthermore, the unidimensionality, local independence, and monotonicity assumptions of the PCM were valid for the BSCS. Q5 was the only item that did not fit the PCM. The step parameters of the BSCS functioned well, with values ranging from low to high, as expected, for all items, indicating monotonicity. Person separation reliability was 0.71, indicating that the BSCS has good internal consistency. The DIF analysis showed that item Q5 functioned differently across genders. In general, the remaining nine items of the BSCS have good psychometric properties for measuring self-control.

Keywords: BSCS, calibration, self-control, polytomous Rasch model, validation

UDK: 159.938.3
DOI: 10.19090/pp.v18i2.2559
Received: 11.06.2024.
Revised: 01.10.2024.
Accepted: 19.11.2024.

✉ Corresponding author's email: diana.mutiah@uinjkt.ac.id

Introduction

Self-control is 'the ability to override or change one's inner responses, as well as to interrupt undesired behavioral tendencies and refrain from acting on them' (Tangney et al., 2004). Research on self-control has grown rapidly in the last decade, especially in relation to emerging constructs in psychology, such as the relationship between self-control and self-discipline (Hagger et al., 2021), grit (Fekih-Romdhane et al., 2022), and loneliness (Stavrova et al., 2022). The construct of self-control has attracted substantial attention from psychologists working within a variety of theoretical and methodological frameworks (Duckworth & Kern, 2011; Gillebaart, 2018).

Various studies have used the theoretical basis provided by Tangney et al.'s (2004) definition to propose components of self-control. Maloney et al. (2012) suggested that impulsivity and restraint are aspects of self-control, whereas de Ridder et al. (2011) stated that inhibitory self-control and initiatory self-control are constituents of self-control. However, according to Tangney et al.'s (2004) initial concept, self-control is a unidimensional construct (see Manapat et al., 2021), meaning that it can be expressed as a single continuum, from less or low self-control to high self-control.

In the Indonesian context, a recent study found that self-control is associated with higher pro-environmental behaviour of Indonesians (Zwagery et al. 2023). In a sample of high school students from Indonesia, more than 80% had moderate levels of self-control (Qonita & Herdi, 2023). In addition, high self-control was found to be associated with lower stress in a sample of nurses from Indonesia (Paramitha & Ariani, 2024). These findings indicate that self-control has been widely studied in Indonesia in recent years.

From a methodological standpoint, self-control measures vary in operational definition and procedure, from simple questionnaires to complex scenarios where individuals must choose whether or not to act (Milyavskaya et al., 2019; Pilcher et al., 2023). Among these various procedures and scales, a tool that is widely used to measure self-control is the Brief Self-Control Scale (BSCS; Tangney et al., 2004). The BSCS was developed from the 36-item Self-Control Scale (SCS) that has a five-dimensional factor structure, although it is scored as a unidimensional one-factor model

(Manapat et al., 2021). The BSCS is widely used because of its Likert-type response format, which is easier to apply compared to other formats (i.e., executive function tasks or delay of gratification tasks; for more details, please see Duckworth & Kern, 2011). In addition, shorter scales, such as the BSCS, are particularly useful in real-world settings, where time and resources are often limited (Pechorro et al., 2021).

The BSCS has been adapted into various languages, for example, Arabic (Fekih-Romdhane et al., 2022), Chinese (Fung et al., 2020), Dutch (Kupper et al., 2020), French (Brevers et al., 2017), German (Bertrams & Dickhäuser, 2009), Greek (Papanikolopoulos et al., 2022), Indonesian (Arifin & Milla, 2020; Zwagery et al., 2023), Italian (Chiesi et al., 2020), Japanese (Ozaki et al., 2016), Persian (Asgarian et al., 2020), Portuguese (Pechorro et al., 2021), Spanish (García-Castro et al., 2024), and Turkish (Nebioglu et al., 2012). Given the broad demographic range of application of the BSCS, population-representative studies have found the instrument to show good stability (Cobb-Clark et al., 2023).

However, since its initial development, the factor structure or dimensionality of the BSCS has changed across studies (Manapat et al., 2021; Papanikolopoulos et al., 2022). The original scale, the 36-item SCS, from which the BSCS is derived, consisted of five dimensions but used a single score with a unidimensional factor model (Manapat et al., 2021). The BSCS has been applied as a two-factor model (e.g., de Ridder et al., 2011; Ferrari et al., 2009; Maloney et al., 2012), while the original version was a one-factor model (e.g., Tangney et al., 2004). Consequently, previous studies (e.g., Hagger et al., 2021; Papanikolopoulos et al., 2022) have tested various measurement models named after the researchers who developed them, such as the ‘Maloney model’, ‘de Ridder model’, etc.

Compared to the 13 items of the original version of the BSCS (Tangney et al., 2004), subsequent versions had different numbers of items, e.g. 10 items (de Ridder et al., 2011) and 8 items (Maloney et al., 2012). Arifin and Milla (2020) based the Indonesian version of the BSCS on de Ridder et al.’s (2011) version consisting of 10 items with two dimensions - inhibition and initiation. This instrument was also used by other studies from Indonesia (Paramitha & Ariani, 2024). Conversely, Zwagery et al. (2023) used Ferrari et al.’s (2009) version with 13 items that measure two dimensions (self-

discipline and impulse control). Other Indonesian studies (Qonita & Herdi, 2023) used Tangney et al.'s (2004) 13-item version of BSCS as a unidimensional measure, but they removed three invalid items, leaving 10 items. Thus, four studies from Indonesia used a unidimensional measure with single scores.

To develop the initial version of the BSCS, Tangney et al. (2004) applied a classical test theory (CTT) approach. From a methodological perspective, CTT has many shortcomings, one of which is that the results of item analysis are highly dependent on the sample used (Andrich & Marais, 2019). The Indonesian version of the BSCS (Arifin & Milla, 2020; Paramitha & Ariani, 2024; Qonita & Herdi, 2023; Zwagery et al., 2023) has been validated using CTT and confirmatory factor analysis (CFA); however, as CFA is a congeneric model in CTT, the analysis has limitations such as the estimation of only one standard error of measurement for all respondents (Rusch et al., 2017).

In general, because of the limitations of CTT, various studies have evaluated the psychometric properties of instruments using modern test theory, i.e. item response theory (IRT), the Rasch model, and item factor analysis, which can mathematically overcome the limitations of CTT (Rusch et al., 2017). In the case of the BSCS, IRT (e.g. Manapat et al., 2021) and Rasch models (e.g. Chen et al., 2022) have provided more detailed item-analysis information compared to CTT. However, our literature review did not find any studies from Indonesia taking this approach.

Importantly, the application of the Rasch model to the Indonesian version of the BSCS would greatly assist non-specialist researchers in using this scale because the model provides a raw-score-to-logit conversion table (Saggino et al., 2020). In addition, by applying the Rasch model, measurement invariance or differential item functioning (DIF) in the Indonesian version of the BSCS instrument can be tested, as has been done with BSCS instruments from other countries. The BSCS has been found to be invariant across genders (Papanikolopoulos et al., 2022) and countries (Hagger et al., 2021). However, other studies have focused on testing gender differences in self-control measurement (Gibson et al., 2010; Jo & Bouffard, 2014). Although both studies used instruments other than BSCS, their results

indicate that DIF testing should be performed across genders for Indonesian samples in terms of self-control as a construct.

Therefore, this study aimed to evaluate the psychometric properties of the Indonesian BSCS instrument developed by Arifin and Milla (2020). Psychometric property testing was performed using the Rasch model, including a comparison of the partial credit model (PCM) and rating scale model (RSM) and testing of the assumptions. We believe that our study is the first to test for measurement invariance or DIF of the Indonesian BSCS.

Method

Participants

We used empirical data from 1001 respondents aged 12–19 years (*M* = 15.07 years, *SD* = 1.771 years); 387 females and 614 males participated in this study. A non-probability sampling method (i.e. quota sampling with a target of 1000 respondents) was used; data were collected over 4.5 months until the quota was reached. A Google form was sent to teachers, who then distributed it to their students. All participants received written information about the aim and procedures. They were also informed that participation was strictly voluntary and could be discontinued at any time without explanation. The information was attached to the questionnaire, and the participants gave their consent by completing the questionnaire.

The criteria used to determine the sample size were based on rule-of-thumb (Tennant & Küçükdeveci, 2023), according to which the minimum sample size for Rasch modelling was 250-500 respondents. It should be noted that this study was part of a larger project for determining pornography addiction among high school students; the BSCS was one of several instruments administered in this project.

Instruments

The Indonesian version of the Brief Self-Control Scale

The instrument used in this study is an Indonesian translation and adaptation by Arifin and Milla (2020) of Tangney et al.'s (2004) BSCS. The scale consists of 10 items (de Ridder et al., 2011), in contrast to the 13 original items. The response scale is a Likert-type scale with five response

options, ranging from *strongly disagree* (1) to *strongly agree* (5). In this study, we treated the Likert scale response options as ordered categorical data. Descriptive statistics of the items in our study are provided in Supplementary Materials.

The Rasch model

The Rasch model (Rasch, 1960) is a measurement theory developed by the Danish mathematician Georg Rasch. Mathematically, the Rasch model is simple, but from the perspective of measurement philosophy, it is profound (Mair, 2018). The Rasch model postulates that an individual's opportunity to correctly answer an item is determined by the interaction between two parameters - the item location and the person's trait level (Wu et al., 2016). The Rasch model uses a logit scale to present item difficulty parameters and individual trait level (or ability) parameters (Andrich & Marais, 2019).

Initially, the Rasch model was used only for analysing dichotomous data (e.g., 1 = 'true', 0 = 'false'). However, the Rasch model can also be used to analyse polytomous data, e.g. the Likert scale (Mair, 2018). The family of Rasch models developed to handle polytomous data includes the RSM (Andrich, 1978) and the PCM (Masters, 1982), both called polytomous Rasch models (PRMs; e.g., Andrich, 2013). In PRMs, one of the advantages of PCM parameterisation, compared to RSM parameterisation, is that it allows each item to have a different number of response categories (Andrich & Marais, 2019). Even when the instruments have the same number of response categories, the PCM provides information on the step parameter structure for each item. Then, the 'disordering of step parameters' of one or more items can be identified; on the other hand, it cannot be detected if the RSM is chosen without performing PCM analysis first (Wu et al., 2016).

The basic assumptions of the Rasch model, i.e. the unidimensionality and local independence assumptions, must be met (Andrich & Marais, 2019). The unidimensionality assumption postulates that all items in the measuring instrument measure a single latent variable. The local independence assumption assumes that an individual's response to one item should not influence their answer to another item (Mair, 2018). However, other assumptions, such as monotonicity (i.e., an assumption that

the probability of a positive response to an item [or, in the case of polytomous items, the transition from one response category to the next] should increase with underlying trait [e.g., self-control] levels) should also be tested (Tennant & Küçükdeveci, 2023).

In this study, the unidimensionality assumption was tested using the principal component analysis of residuals (PCAR; Smith, 2002). The local independence assumption was tested using the raw residual correlations between all pairs of items which is called Q3 statistics (Yen, 1984). Monotonicity was assessed by inspecting item threshold (step) patterns, which are expected by the model to monotonically increase from low to high across the continuum with no disordering (Tennant & Küçükdeveci, 2023).

Data analysis strategy

Because PRM has two parameterisations, rating scale parameterisation (RSM) and partial credit parameterisation (PCM), we performed both analyses and compared the global-fit statistics in the first phase. Models with better fit statistics were chosen. Mathematically, the comparisons between PCM and RSM are valid because the models are nested (Linacre, 2021). In the second phase, we tested the unidimensionality and local independence of the BSCS using the chosen model. In the third phase, we examined item fit statistics, including step parameters, to check for monotonicity. In the fourth phase, person-item maps or Wright Maps were reported. In the fifth phase, person reliability and item reliability were reported. In the sixth phase, the test information function was inspected. The seventh and last phase was DIF analysis. All phases were implemented in the Winsteps 5.1.4 program (Linacre, 2021) using unconditional or joint maximum likelihood estimation methods. We used the Winsteps-integrated 'wrightMap' package in R to create graphs (Irribarra & Freund, 2024).

Results

Global-fit statistics and model comparison

In the first phase, the two models, PCM and RSM, were analysed with two separate calibrations and the global-fit statistics of the models were compared. The global-fit statistics used were log-likelihood chi-squared (χ^2),

Akaike information criterion (AIC), and Bayesian information criterion (BIC), with a lower value indicating a better fitting model. Statistics based on residuals, root mean square residuals (RMSR), were also reported, with lower values indicating a better fitting model. Finally, the root mean square error (RMSE) was also compared, with a lower value indicating a better fitting model (Linacre, 2021; Welter et al., 2024). All global fit statistics and other statistics, such as RMSR and RMSE, indicated that the data for the Indonesian BSCS fit the PCM better than the RSM (for the model comparison results, see Table 1). The model comparison statistics AIC and BIC indicated that PCM fits the data better than RSM. Additionally, RMSR and RMSE were lower for PCM. Based on these findings, the model or parameterisation reported next is the PCM.

Table 1

Fit Statistics of the Rating Scale Model (RSM) and the Partial Credit Model (PCM)

Fit statistics	Model	
	RSM	PCM
Log-likelihood χ^2	20049.803	19442.345
AIC	22073.803	21520.345
BIC	29370.667	29011.889
RMSR	0.674	0.661
RMSE	0.520	0.516

Unidimensionality and local independence

In using PCAR, unidimensionality is achieved based on two criteria: first, the raw variance explained by measures should be greater than 40% (Holster & Lake, 2016); second, the unexplained variance in the 1st contrast should not exceed 2.0 (Smith, 2002). The results of the PCAR of the BSCS showed that 40.4% of the raw variance was explained by measures. The unexplained variance in the 1st contrast was 1.59 (Table 2).

Table 2

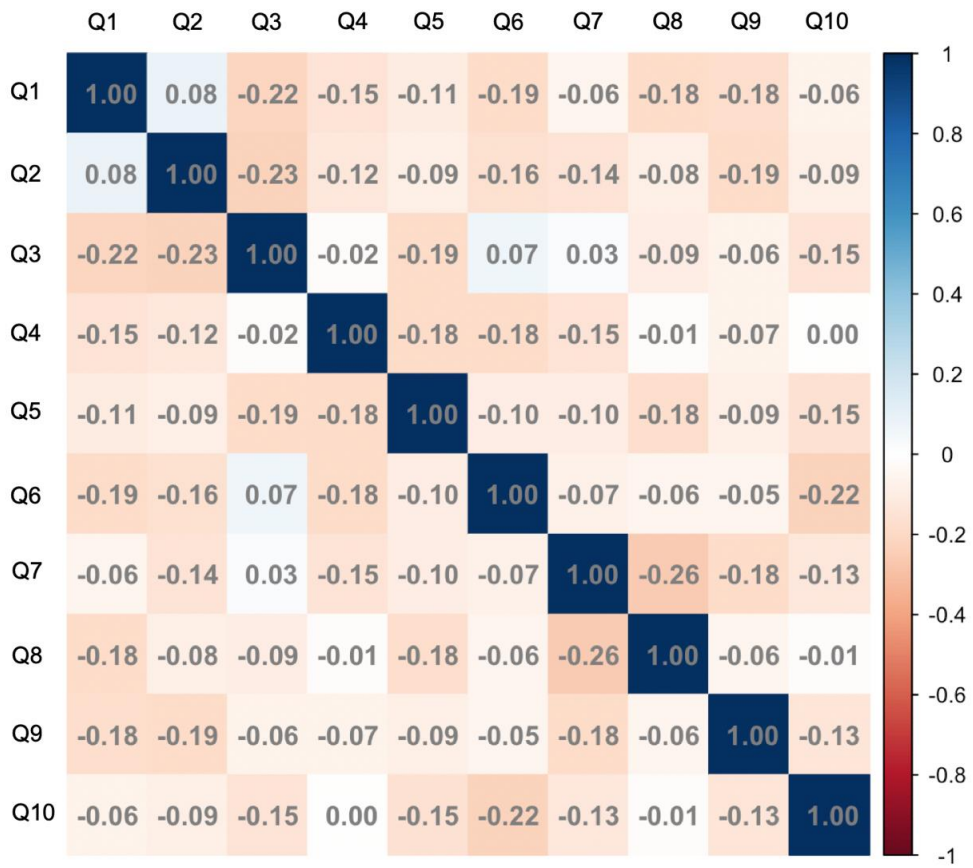
Results of the Principal Component Analysis of Residuals (PCAR)

PCAR statistics	Eigenvalue	Proportion of variance from observed data
Total raw variance	16.791	100%
Raw variance explained by measures	6.791	40.4%
Raw unexplained variance (total)	10.001	59.6%
Unexplained variance in the 1st contrast	1.590	9.5%

These results indicate that the unidimensionality assumption was met for the BSCS. In other words, the BSCS instrument was empirically found to measure one construct: self-control. Furthermore, after PCAR, the residual correlation matrix was inspected to check the assumption of local independence (see Figure 1).

Figure 1

Raw Residual Correlation Matrix (Q3 Statistics)



The results of the Q3 statistical test showed that the majority of items had a negative raw residual correlation (Figure 1). More specifically, there was no positive raw residual correlation with a value greater than 0.25 (DeMars, 2010), indicating that there was no substantial local dependency between the items. In addition, although items Q7 and Q8 had a residual correlation of -0.26, given the direction is negative, this is negligible. In other words, these results suggested that the assumption of local independence of the BSCS was met.

Item measure and fit statistics

After confirming that the assumptions of unidimensionality and local independence were valid for the BSCS, item parameter estimates and fit statistics were inspected (see Table 3). An item was deemed to fit the PCM if Infit and Outfit mean squares (MNSQs) were in the range of 0.5 to 1.5, with a point measure (PTMEA) correlation > 0.30 (Boone, 2020). Table 3 shows the items ordered from the most difficult to the easiest to endorse. The item location rangedbetween -0.971 and 1.603 logits. The more negative the difficulty level, the easier it was to obtain a higher score (a score of 5) on an item and vice versa. The easiest item on the BSCS scale was Q10, '*Saya menolak hal-hal yang buruk untuk diri saya*' [I refuse things that are bad for me], with a difficulty level of -0.971 logit. The most difficult item was Q5, '*Hal yang menyenangkan dan bersenang-senang kadang menahan saya untuk menyelesaikan pekerjaan*' [Pleasure and fun sometimes keep me from getting work done], with a difficulty level of 1.603 logits.

Table 3

Item Parameters, Fit Statistics, and Thresholds for All Items

Item	Measure	Infit MNSQ	Outfit MNSQ	PTMEA Corr.	Step 1	Step 2	Step 3	Step 4
Q5	1.603	1.68	1.67	-0.04	-2.68	0.08	3.12	5.90
Q6	1.322	1.05	1.04	0.47	-1.38	-0.68	2.34	5.01
Q8	0.229	0.78	0.78	0.67	-1.63	-0.94	0.62	2.86
Q9	-0.032	0.88	0.87	0.62	-2.32	-1.14	0.80	2.54
Q7	-0.059	1.05	1.05	0.50	-3.59	-0.83	0.87	3.31
Q3	-0.196	1.05	1.04	0.47	-3.64	-2.35	1.74	3.48
Q4	-0.606	0.87	0.86	0.58	-2.48	-1.76	-0.31	2.13
Q1	-0.636	0.93	0.87	0.57	-1.99	-1.00	-0.63	1.08
Q2	-0.653	0.89	0.84	0.58	-2.17	-1.00	-0.64	1.19
Q10	-0.971	0.85	0.82	0.61	-3.94	-1.37	-0.22	1.64

Table 3 shows that one item did not fit the Rasch PCM. Item 5 (Q5) did not fit the model because the Infit and Outfit MNSQs were outside the range of 0.5–1.5, and the PTMEA correlation was lower than 0.30. Infit and Outfit MNSQs indicated an underfit for item Q5. We believe that random or aberrant

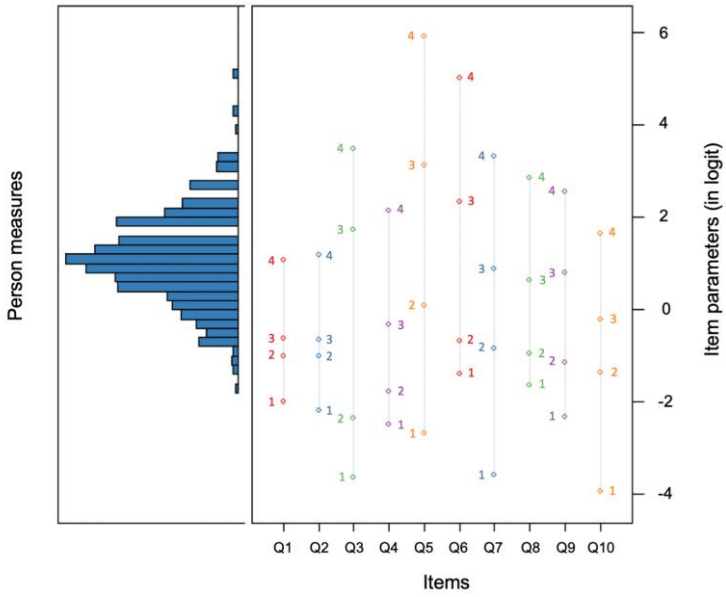
responses caused this, i.e. a person with a low level of self-control had the highest score for item 5. In addition, because of its low and negative PTMEA correlation, item Q5 was not sensitive enough to distinguish between a person with a low level of self-control and a person with a high level of self-control. Another possibility is that the translation from English to Indonesian is suboptimal and that this item is not very specific in measuring self-control in the Indonesian population. Furthermore, by inspecting the step parameter patterns, we found that step parameters for all items monotonically increased from low to high. This finding corroborates the monotonicity assumption of the model.

Wright Map

The Wright Map is one of the most significant innovations resulting from Rasch measurement. Using the Wright Map, the persons and items can be reviewed, and the relationship between persons and items can be inspected. Another aspect of the Wright Map is that persons and items are on the same scale, enabling insight into the respondents' performance on a set of test items (Liu & Boone, 2023). The Wright Map of the Indonesian BSCS is shown in Figure 2.

Figure 2

The Wright Map of the Indonesian BSCS



To complete the information in Figure 2, the mean of item location was 0.000, whereas the mean of person self-control level was 1.062. Thus, there was a 1-logit difference between the mean of the person measure and the mean of the item location. This finding indicates that in our study, person tended to have a higher level of self-control compared to the behavioural content of the items. Furthermore, the step parameters (1, 2, 3, and 4) of item Q5 spread widely across the continuum. However, item Q1 had a narrower step parameter range compared to Q5, Q3, Q6, and Q7. This finding is one of the reasons that PCM was found to be a better fit than the RSM since, in the latter, although the difference in the step parameters of Q5 and Q1 was very large, the threshold range was 'forced' to be mathematically uniform.

Reliability and separation indices

When using the Rasch model, reliability was reported for both item and person. Rasch-based reliability has two aspects: person separation reliability (PSR) and item separation reliability (ISR; Andrich & Marais, 2019). PSR is a measure of how well the measuring instrument differentiates between individuals with high ability and those with low ability and a measure of internal consistency, while ISR is a measure of how reliable the sample size is in classifying items in the hierarchy (Wright & Stone, 1999). A low ISR value indicates the need to increase the sample size so that the item hierarchy can be trusted. The PSR and ISR of the BSCS were 0.71 and 1.00, respectively. The PSR > 0.70 indicated that the BSCS has fairly good (acceptable) internal consistency and the ISR > 0.90 confirmed the item hierarchy (Linacre, 2021).

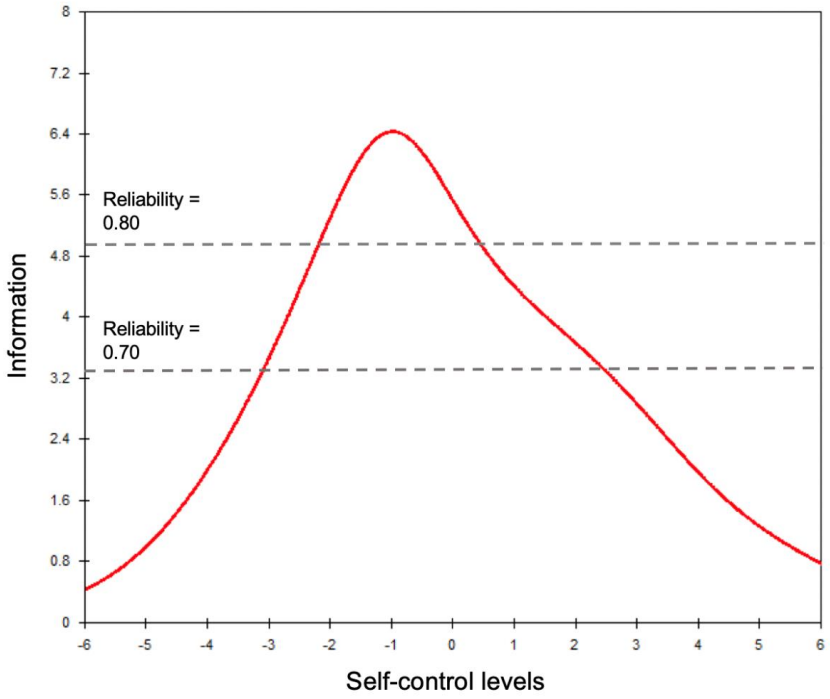
Test information function

The concept of the test information function (TIF) reflects how precisely a scale can measure the underlying trait, i.e. self-control (Wu et al., 2016). More specifically, TIF provides an explanation of the range of abilities for which the test provides the most precise measurement or produces the lowest measurement error when reliability for each ability level, called conditional reliability, can be calculated (DeMars, 2010). In the case of the PCM TIF of the Indonesian BSCS, the estimation results showed that for an

amount of information of 3.34, reliability was 0.70, while for 5.01, reliability was 0.80 (see Figure 3).

Figure 3

Test Information Function Curve of the Indonesian BSCS



Based on the estimation of the conditional reliability of the Indonesian BSCS, self-control was found to range from −3.061 to −2.152 and from 0.432 to 2.446 (71.82% of our study sample was in this range), resulting in a conditional reliability of 0.700–0.799. Furthermore, in the self-control range of −2.143 to 0.423 logits (21.47% of our study sample was within this range), conditional reliability ranged from 0.800 to 0.844. Based on these findings, we concluded that the Indonesian BSCS had optimal measurement precision that covered a fairly wide area (-3.061 to 2.446 logits), with 93.3% of this study's sample being in this range (only 6.7% were outside the range of optimal measurement precision).

DIF Analysis

Differential item functioning exists when different groups of test takers at the same ability level have significantly different chances of answering a test item because the test interacts with off-trait characteristics (Wu et al., 2016). To examine whether the items were of comparable difficulty for different gender groups, we examined the presence of DIF depending on gender. Rasch-based DIF analysis or the Rasch–Welch t-test was used (Smith, 1994). The criterion of at least a 0.400 logit difference (DIF contrast) with a p-value < 0.05 was used for detecting DIF (Linacre, 2013). The results of the DIF analysis (Table 4) indicated that item Q5 ('Fun and having fun sometimes prevent me from finishing my work') showed substantial DIF because the DIF contrast of −0.429 was greater than the predefined criterion. This finding indicates that males have a higher chance of obtaining the highest score (scored 5) on item Q5 compared to females, despite the two groups having the same level of self-control. We suspect that the content of item Q5 (i.e., the phrase 'having fun') had different meanings for males and females, leading to DIF of item Q5. Lastly, item Q9 has significant DIF ($p = 0.028 < 0.050$), but because the DIF contrast is relatively low at 0.200 (< 0.400), this item is categorized as having negligible DIF.

Table 4

Results of the DIF Analysis (Male – Female)

Item	DIF contrast	Joint S.E.	*t*	*df*	*p*
Q1	0.000	0.000	0.01	879	1.000
Q2	0.115	0.094	1.23	872	0.218
Q3	0.053	0.112	0.47	894	0.637
Q4	0.094	0.101	0.93	880	0.350
Q5	-0.425	0.112	-3.80	890	0.000
Q6	-0.029	0.102	-0.28	887	0.776
Q7	-0.021	0.096	-0.22	887	0.826
Q8	0.022	0.090	0.24	883	0.807
Q9	0.200	0.091	2.20	883	0.028
Q10	-0.110	0.094	-1.16	887	0.245

Discussion

The primary goal of this study was to assess the psychometric properties of the Indonesian BSCS using Rasch polytomous models, followed by DIF analysis to assess whether the Indonesian BSCS worked equally well across different gender groups. Based on our initial analysis, we compared the PCM and RSM and found that the PCM parameterisation fit better than the RSM. Such a comparison between the models is in line with previous studies (e.g. Gori et al., 2022; Youngerman et al., 2021). In addition, the Wright Map showed that the step parameter structure of the PCM was not uniform when there were large differences between item Q5 and item Q1. If the RSM were used, the differences in step parameters were forced to be equal. Therefore, the RSM had a poorer fit compared to the PCM. These findings are in line with Wu et al. (2016), who stated that the data rarely fit the RSM due to the assumption of an equal (uniform) step parameter structure for all items.

Based on unidimensionality or internal structure, we confirm that the Indonesian BSCS has a unidimensional factor structure. This factor structure is in line with the first adaptation study (Arifin & Milla, 2020), the original version of the scale (Tangney et al., 2004), and a recent study that also employed the Rasch model for the BSCS in China (Chen et al., 2022). However, this finding is not in line with the results of other studies that applied multidimensional IRT to the BSCS (Manapat et al., 2021) or the previous 10-item version of the BSCS (de Ridder et al., 2011). The 'disagreement' about the BSCS factor structure has been discussed in previous studies (e.g. Manapat et al., 2021).

The Indonesian BSCS did not show local dependence in the local independence test. This finding is in line with Chiesi et al.'s (2020) study that found that almost no BSCS models were modified by freeing the residual correlation in the CFA model, as well as studies that used modern test theory which did not find local dependence problems in the BSCS (Chen et al., 2022; Manapat et al., 2021). In addition, the Indonesian BSCS adaptation study did not modify the model by freeing residual correlation to achieve a model with a good fit (Arifin & Milla, 2020).

We found that all items had step parameters with a monotonically increasing pattern from lower to higher self-control levels. This finding indicates that the monotonicity assumption was met as recommended by previous studies (i.e., Tennant & Küçükdeveci, 2023). Therefore, the category functioning of the Indonesian BSCS does not experience rating scale malfunctioning (Wind, 2023) or disordered thresholds (Andrich, 2013), indicating that all response categories are well-functioning.

However, fit statistics showed that one item, Q5 ('*Hal yang menyenangkan dan bersenang-senang kadang menahan saya untuk menyelesaikan pekerjaan*' [Pleasure and fun sometimes keep me from getting work done]), did not fit the model. This item showed the greatest difficulty in obtaining the highest score. The reason for the poor fit of this item may be associated with random or aberrant responses (e.g., Karabatsos, 2000). We suspect that a number of respondents with a high level of self-control obtained the lowest score (a score of 1) on this item. Conversely, respondents with a low level of self-control may obtain the highest score (a score of 5) on this item. This condition causes fit statistics to experience strain (Outfit MNSQ > 1.5, PTMEA < 0.30) (Karabatsos, 2000). The procedure that can be used to obtain statistical evidence for this is response pattern analysis (Wright & Stone, 1999). In addition, we identified several flagged misfits, i.e. persons whose response patterns were suspected of causing the lack of fit of item Q5 (see Table S2 in Supplementary Materials). However, we did not conduct any follow-up analyses (i.e., analyses after removing misfits) because it was outside the focus of this study.

In the item hierarchy, the easiest item to agree with was Q10, '*Saya menolak hal-hal yang buruk untuk diri saya*' [I refuse things that are bad for me]. The review of the item content showed that the behaviour measured through this item was indeed very normative. However, because the application of the Rasch model to the BSCS is not widely studied, we cannot compare this level of difficulty with that of other studies. Finally, an ISR of 1.000 means that there is no problem with the item hierarchy or item spread (Wright & Stone, 1999).

Furthermore, the DIF analysis showed a DIF of Q5, indicating that this item tended to benefit males. This finding complements those of previous studies that focused on gender differences in self-control (Gibson et al.,

2010; Jo & Bouffard, 2014) but is inconsistent with the findings of other studies that the BSCS is invariant across gender (Chiesi et al., 2020). Although from a Rasch perspective, this method would help non-specialists to use the raw-score-to-logit conversion table (Saggino et al., 2020), when DIF is present, the raw score is no longer sufficient for the Rasch model (Linacre, 1992). Therefore, we did not report the raw-score-to-logit conversion table of the Indonesian BSCS. Further studies with other samples are needed to decide whether the DIF of Q5 item occurs consistently; if so, then dropping the Q5 item could be considered. The Indonesian BSCS had good internal consistency (PSR = 0.71). Our findings align with the results of other studies that applied the Rasch model to the BSCS (e.g. Chen et al., 2022). Based on the TIF, we found that the Indonesian BSCS covered a wide range of self-control levels classified as the optimal measurement precision range (high conditional reliability); 93.3% of our study sample had measures within the good measurement precision range. However, it should be noted that conditional reliability has a different meaning for PSR (single score) and should not be compared. Our finding also reflects the superiority of the Rasch model over CTT or CFA because of the model generates a conditional standard error of measurement for each level of trait levels (e.g. Andrich & Marais, 2019; Rusch et al., 2017).

This study has some limitations, both theoretical and methodological. The main theoretical limitation concerns the chosen BSCS model from a group of models - the 'Maloney model,' 'de Ridder model,' and 'Ferrari model' (e.g. Chen et al., 2022; Chiesi et al., 2020). While testing different models within a single study would be ideal, we only collected data using the 10-item version (i.e., the de Ridder version). Future studies should test various models using the 13-item version of the BSCS. The methodological limitation concerns administering the BSCS simultaneously with the pornography addiction test, which was the main focus of the broader research project. We believe that social desirability or response faking potentially had an impact on the Indonesian BSCS because respondents might have assumed that the BSCS instrument was related to the simultaneously administered pornography addiction instrument. Future studies should focus specifically on measuring self-control with the BSCS so that respondents are not distracted by other (negative) constructs.

Conclusion

In conclusion, this study is the first to validate and evaluate the psychometric properties of the Indonesian BSCS using the Rasch model. It showed that nine of the 10 Indonesian BSCS items were valid for measuring self-control. This study also showed that the basic assumptions of modern test theory - unidimensionality, local independence, and monotonicity - were fulfilled for this instrument. Lastly, our study should be replicated with different samples but applying the same method and same sample characteristics (i.e., high school students) to confirm whether DIF and misfit occur in item Q5.

Conflict of interest

We have no conflicts of interest to disclose.

Data availability statement

Data files are available upon a reasonable request.

References

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43*(4), 561–573. https://doi.org/10.1007/BF02293814

Andrich, D. (2013). An expanded derivation of the threshold structure of the polytomous Rasch model that dispels any "threshold disorder controversy". *Educational and Psychological Measurement, 73*(1), 78–124. https://doi.org/10.1177/0013164412450877

Andrich, D., & Marais, I. (2019). *A course in rasch measurement theory: measuring in the educational, social and health sciences*. Springer Nature Singapore Pte Ltd. https://doi.org/10.1007/978-981-13-7496-8

Arifin, H., & Milla, M. (2020). Adaptasi dan properti psikometrik skala kontrol diri ringkas versi Indonesia. *Jurnal Psikologi Sosial, 18*(2), 179-195. https://doi.org/10.7454/jps.2020.18

Asgarian, F. S., Namdari, M., & Soori, H. (2020). Reliability and validity of Persian version of Brief Self-Control Scale (BSCS) in motorcyclists. *International Journal of Psychiatry in Clinical Practice, 24*(2), 176–182. https://doi.org/10.1080/13651501.2019.1711423

Bertrams, A., & Dickhäuser, O. (2009). High-school students' need for cognition, self-control capacity, and school achievement: Testing a mediation

hypothesis. *Learning and Individual Differences, 19*(1), 135–138. https://doi.org/10.1016/j.lindif.2008.06.005

Boone, W. J. (2020). Rasch basics for the novice. In M. S. Khine (Ed.), *Rasch measurement: Applications in quantitative educational research* (pp. 9-30). Springer. https://doi.org/10.1007/978-981-15-1800-3_2

Brevers, D., Foucart, J., Verbanck, P., & Turel, O. (2017). Examination of the validity and reliability of the French version of the Brief Self-Control Scale. *Canadian Journal of Behavioural Science / Revue canadienne des sciences du comportement, 49*(4), 243–250. https://doi.org/10.1037/cbs0000086

Chen, W., Zhang, G., Tian, X., & Zhao, S. (2022). Factor structure and longitudinal measurement invariance of the Tangney's Brief Self-Control Scale in Chinese adolescents. *Frontiers in Public Health, 10*, Article 802448. https://doi.org/10.3389/fpubh.2022.802448

Chiesi, F., Bonacchi, A., Lau, C., Tosti, A. E., Marra, F., & Saklofske, D. H. (2020). Measuring self-control across gender, age, language, and clinical status: A validation study of the Italian version of the Brief Self-Control Scale (BSCS). *PLoS ONE, 15*(8), e0237729. https://doi.org/10.1371/journal.pone.0237729

Cobb-Clark, D. A., Kong, N., & Schildberg-Hörisch, H. (2023). The stability of self-control in a population-representative study. *Journal of Economic Psychology, 95*, 1–10. https://doi.org/10.1016/j.joep.2022.102599

de Ridder, D. T. D., De Boer, B. J., Lugtig, P., Bakker, A. B., & van Hooft, E. A. J. (2011). Not doing bad things is not equivalent to doing the right thing: Distinguishing between inhibitory and initiatory self-control. *Personality and Individual Differences, 50*(7), 1006-1011. https://doi.org/10.1016/j.paid.2011.01.015

DeMars, C. (2010). *Item response theory*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195377033.001.0001

Duckworth, A. L., & Kern, M. L. (2011). A meta-analysis of the convergent validity of self-control measures. *Journal of Research in Personality, 45*(3), 259–268. https://doi.org/10.1016/j.jrp.2011.02.004

Fekih-Romdhane, F., Postigo, Á., González-Nuevo, C., Dabbous, M., Malaeb, D., Obeid, S., & Hallit, S. (2022). Psychometric properties of the Arabic version of the Oviedo Grit Scale (A-EGO) in non-clinical adults from the general population. *BMC Psychiatry, 22*(1), 792. https://doi.org/10.1186/s12888-022-04466-1

Ferrari, J. R., Stevens, E. B., & Jason, L. A. (2009). The role of self-regulation in abstinence maintenance: Effects of communal living on self-regulation. *Journal of Groups in Addiction & Recovery, 4*(1–2), 32–41. https://doi.org/10.1080/15560350802712371

Fung, S., Kong, C. Y. W., & Huang, Q. (2020). Evaluating the dimensionality and psychometric properties of the brief self-control scale amongst Chinese University Students. *Frontiers in Psychology*, *10*, 2903. https://doi.org/10.3389/fpsyg.2019.02903

García-Castro, F. J., Bendayan, R., & Blanca, M. J. (2024). Validity evidence for the Brief Self-Control Scale in the Spanish adult population: A systematic study. *Journal of Personality Assessment, 106*(2), 208-217. https://doi.org/10.1080/00223891.2023.2220403

Gibson, C. L., Ward, J. T., Wright, J. P., Beaver, K. M., & Delisi, M. (2010). Where does gender fit in the measurement of self-control? *Criminal Justice and Behavior, 37*(8), 883–903. https://doi.org/10.1177/0093854810369082

Gillebaart, M. (2018). The 'operational' definition of self-control. *Frontiers in Psychology, 9*, 1231. https://doi.org/10.3389/fpsyg.2018.01231

Gori, A., Topino, E., Svicher, A., Schuldberg, D., & Di Fabio, A. (2022). Insight orientation scale: A promising tool for organizational outcomes–A psychometric analysis using item response theory. *Frontiers in Psychology, 13*, 987931. https://doi.org/10.3389/fpsyg.2022.987931

Hagger, M., Zhang, C.-Q., Kangro, E.-M., Ries, F., Wang, J. C. K., Heritage, B., & Chan, D. K. C. (2021). Trait self-control and self-discipline: Structure, validity, and invariance across national groups. *Current Psychology, 40*(3), 1015-1030. https://doi.org/10.1007/s12144-018-0021-6

Holster, T. A., & Lake, J. (2016). Guessing and the Rasch model. *Language Assessment Quarterly, 13*(2), 124–141. https://doi.org/10.1080/15434303.2016.1160096

Irribarra, D. T., & Freund, R. (2024). Package 'WrightMap'. Retreived from https://cran.r-project.org/web/packages/WrightMap/WrightMap.pdf

Jo, Y., & Bouffard, L. (2014). Stability of self-control and gender. *Journal of Criminal Justice, 42*(4), 356–365. https://doi.org/10.1016/j.jcrimjus.2014.05.001

Karabatsos, G. (2000). A critique of Rasch residual fit statistics. *Journal of Applied Measurement, 1*(2), 152–176.

Kupper, N., Duijndam, S., & Karreman, A. (2020). Emotion expressivity in the Dutch: Validation of the Dutch translation of the Berkeley Expressivity

Questionnaire. *Psychological Assessment, 32*(5), 28–34. https://doi.org/10.1037/pas0000812

Linacre, J. (1992). Why fuss about statistical sufficiency? *Rasch Measurement Transactions, 6*(3), 230. https://www.rasch.org/rmt/rmt63c.htm

Linacre, J. M. (2013). Differential item functioning (DIF) sample size nomogram. *Rasch Measurement Transactions, 26*(4), 1391. https://www.rasch.org/rmt/rmt264a.htm

Linacre, J. M. (2021). *Winsteps® Rasch measurement computer program user's guide* (Version 5.1.4). Winsteps.com.

Liu, X., & Boone, W. J. (2023). Introduction to advances in applications of Rasch measurement in science education. In X. Liu, & W. J. Boone (Eds.), *Advances in applications of Rasch measurement in science education* (pp. 1-18). Springer. https://doi.org/10.1007/978-3-031-28776-3_1

Mair, P. (2018). *Modern psychometrics with R*. Springer International Publishing AG. https://doi.org/10.1007/978-3-319-93177-7

Maloney, P. W., Grawitch, M. J., & Barber, L. K. (2012). The multi-factor structure of the Brief Self-Control Scale: Discriminant validity of restraint and impulsivity. *Journal of Research in Personality, 46*(1), 111–115. https://doi.org/10.1016/j.jrp.2011.10.001

Manapat, P. D., Edwards, M. C., MacKinnon, D. P., Poldrack, R. A., & Marsch, L. A. (2021). A psychometric analysis of the Brief Self-Control Scale. *Assessment, 28*(2), 395-412. https://doi.org/10.1177/1073191119890021

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*(2), 149–174. https://doi.org/10.1007/BF02296272

Milyavskaya, M., Berkman, E. T., & de Ridder, D. T. D. (2019). The many faces of self-control: Tacit assumptions and recommendations to deal with them. *Motivation Science, 5*(1), 79–85. https://doi.org/10.1037/mot0000108

Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling, 9*(4), 599–620. https://doi.org/10.1207/S15328007SEM0904_8

Nebioglu, M., Konuk, N., & Eroglu, Y. (2012). The investigation of validity and reliability of the Turkish version of the Brief Self-Control Scale. *Klinik Psikofarmakoloji Bülteni-Bulletin of Clinical Psychopharmacology, 22*(4), 340-351. https://doi.org/10.5455/bcp.20120911042732

Ozaki, Y., Goto, T., Kobayashi, M., & Kutsuzawa, G. (2016). Reliability and validity of the Japanese translation of Brief Self-Control Scale (BSCS-J). *Shinrigaku Kenkyu: The Japanese Journal of Psychology, 87*(2), 144-154. https://doi.org/10.4992/jjpsy.87.14222

Papanikolopoulos, P., Mastrotheodoros, S., Papanikolopoulou, S. Eleni Parlapani, E., & Kaprinis, S. G. (2022) The brief self-control scale: Dimensionality and psychometric properties in Greek young adults. *European Journal of Developmental Psychology, 19*(6), 925-937. https://doi.org/10.1080/17405629.2021.1964070

Paramitha, D. S., & Ariani, L. (2024). Hubungan kontrol diri dengan stress kerja pada bidan di rumah sakit Banjarmasin. *MANUJU: Malahayati Nursing Journal, 6*(6), 2415-2425. https://doi.org/10.33024/mnj.v6i6.11124

Pechorro, P., DeLisi, M., Gonçalves, R. A., Quintas, J., & Hugo Palma, V. (2021). The brief self-control scale and its refined version among incarcerated and community youths: Psychometrics and measurement invariance. *Deviant Behavior, 42*(3), 425–442. https://doi.org/10.1080/01639625.2019.1684942

Pilcher, J. J., Morris, D. M., & Erikson, D. N. (2023). Self-control measurement methodologies: An integrative approach. *Psychological Reports, 126*(3), 1108–1129. https://doi.org/10.1177/00332941211067969

Qonita, M., & Herdi, H. (2023). Asesmen kontrol diri peserta didik dan implikasinya terhadap program bimbingan dan konseling di SMA. *Jurnal Mahasiswa Bimbingan dan Konseling, 9*(2), 331-341. http://dx.doi.org/10.31602/jmbkan.v9i2.11257

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Danish Institute for Educational Research.

Rusch, T., Lowry, P. B., Mair, P., & Treiblmaier, H. (2017). Breaking free from the limitations of classical test theory: Developing and measuring information systems scales using item response theory. *Information & Management, 54*(2), 189-203. https://doi.org/10.1016/j.im.2016.06.005

Saggino, A., Molinengo, G., Rogier, G., Garofalo, G., Loera, B., Tommasi, M., & Velotti, P. (2020). Improving the psychometric properties of the dissociative experiences scale (DES-II): A Rasch validation study. *BMC Psychiatry, 20*, 8. https://doi.org/10.1186/s12888-019-2417-8

Smith, E. (2002). Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *Journal of Applied Measurement, 3*(2), 205–231.

Smith, R. M. (1994). Detecting item bias in the Rasch Rating Scale Model. *Educational and Psychological Measurement, 54*(4), 886-896. https://doi.org/10.1177/0013164494054004004

Stavrova, O., Ren, D., & Pronk, T. (2022). Low self-control: A hidden cause of loneliness? *Personality and Social Psychology Bulletin, 48*(3), 347-362. https://doi.org/10.1177/01461672211007228

Tangney, J. P., Baumeister, R. F., & Boone, A. L. (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *Journal of Personality, 72*(2), 271–322. https://doi.org/10.1111/j.0022-3506.2004.00263.x

Tennant, A., & Küçükdeveci, A. A. (2023). Application of the Rasch measurement model in rehabilitation research and practice: early developments, current practice, and future challenges. *Frontiers in Rehabilitation Sciences, 4*, Article 1208670. https://doi.org/10.3389/fresc.2023.1208670

Welter, V. D. E., Dawborn-Gundlach, M., Großmann, L., & Krell, M. (2024). Adapting a self-efficacy scale to the task of teaching scientific reasoning: Collecting evidence for its psychometric quality using Rasch measurement. *Frontiers in Psychology, 15*, Article 1339615. https://doi.org/10.3389/fpsyg.2024.1339615

Wind, S. A. (2023). Detecting rating scale malfunctioning with the Partial Credit Model and Generalized Partial Credit Model. *Educational and Psychological Measurement, 83*(5), 953-983. https://doi.org/10.1177/00131644221116292

Wright, B. D., & Stone, M. (1999). *Measurement essentials* (2nd ed.). Wide Range, Inc.

Wu, M., Tam, H. P., & Jen, T.-H. (2016). *Educational measurement for applied researchers: theory into practice*. Springer Nature Singapore Pte Ltd.

Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement, 8*(2), 125-145. https://doi.org/10.1177/014662168400800201

Youngerman, E., Dahl, L. S., & Mayhew, M. J. (2021). Examining the psychometric properties of a New Integrative Learning Scale. *Research in Higher Education, 62*(6), 829-854. https://doi.org/10.1007/s11162-021-09623-1

Zwagery, R. V., Erlyani, N., Mayangsari, M. D., & Sugma, K. A. A. (2023). Kontrol diri dan perilaku pro-lingkungan. *Psibernetika, 16*(2), 63-69. http://dx.doi.org/10.30813/psibernetika.v16i2.3654

Supplementary Materials

Table S1

Descriptive Statistics for all Indonesian BSCS Items

Items	*M*	*SD*	Skewness	Minimum	Maximum
Q1	4.229	0.849	-1.103	1.000	5.000
Q2	4.206	0.841	-1.040	1.000	5.000
Q3	3.400	0.663	0.599	1.000	5.000
Q4	3.996	0.756	-0.535	1.000	5.000
Q5	2.820	0.666	0.017	1.000	5.000
Q6	3.033	0.740	-0.319	1.000	5.000
Q7	3.502	0.806	-0.177	1.000	5.000
Q8	3.594	0.876	-0.455	1.000	5.000
Q9	3.631	0.863	-0.213	1.000	5.000
Q10	4.061	0.817	-0.587	1.000	5.000

Figure S1

Plot of proportion of respondents for all response categories in all items

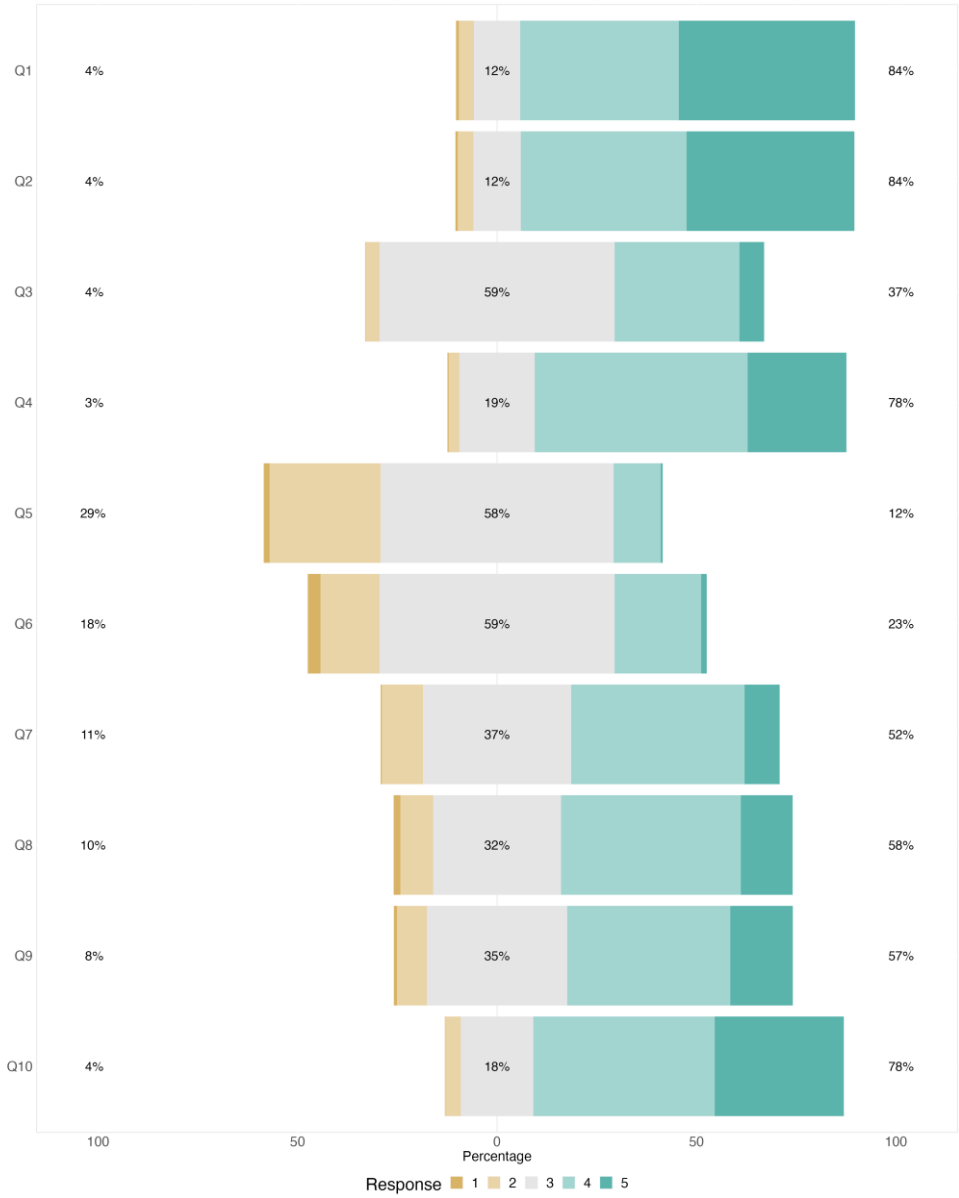


Table S2

Response pattern analysis for explaining misfit of Q5 (selected misfit persons)

Person ID	Theta	Response pattern										Outfit MNSQ	Outfit ZSTD
		Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10		
P0654	2.385	5	5	3	5	1	5	4	5	4	5	3.22	4.1
P0293	1.828	5	3	4	4	1	5	4	4	5	5	2.88	2.7
P0832	2.099	5	5	3	5	1	4	5	4	4	5	3.05	2.9

Note. Table S1 shows three examples of our respondents who did not fit the partial credit model. These respondents had high levels of self-control, but instead responded with the lowest scores on Item Q5. This response pattern is what mathematically causes: (1) Outfit MNSQ of Q5 > 1.50 (underfit); (2) PTMEA low and negative.