Jurnal Penelitian dan Evaluasi Pendidikan
Volume 29, No. 1, June 2025 (85-97)
Online: http://journal.uny.ac.id/index.php/jpep

Psychometric properties of the general conspiracy belief scale using item response theory

Sumin Sumin[1]*; Khairawati Khairawati[1]; Mohd. Shahrul Kamaruddin[2]
[1]Institut Agama Islam Negeri Pontianak, Indonesia
[2]Universiti Malaysia Sarawak, Malaysia
*Corresponding Author. E-mail: amien.ptk@gmail.com

ARTICLE INFO

Article History
Submitted:
23 April 2025
Revised:
24 April 2025
Accepted:
25 April 2025

Keywords
general conspiracy belief; item response theory; conspiracy; psychometrics; scales

Scan Me:

ABSTRACT

Evaluation of the psychometric properties of conspiracy theory belief instruments has been dominated by classical approaches with limitations, especially in dependence on sample size and inaccuracies in item-level analysis. This study aims to fill this gap by applying a polytomous Item Response Theory (IRT) approach to reanalyze the General Conspiracy Belief Scale (GCBS). This study aims to re-examine the psychometric properties of the GCBS with an IRT approach to produce measurements that are more precise and independent of sample characteristics. The research design used was a quantitative replication utilizing secondary data from 2,495 students at the college level. The instrument used consisted of 15 items on a five-category Likert scale. The analysis was conducted using three polynomial IRT models, namely the Graded Response Model (GRM), Partial Credit Model (PCM), and Generalized Partial Credit Model (GPCM), with the help of R software. The results showed that the GRM model was the model that best fit the data, with most items showing high distinctiveness and providing maximum information on respondents with low to moderate levels of conspiratorial belief. Empirical marginal reliability coefficients were high, indicating that the instrument's internal consistency was perfect. This study contributes to the field by offering a more robust and nuanced psychometric evaluation of the GCBS through IRT, providing researchers with a validated framework for assessing conspiracy beliefs with higher accuracy and scale precision. However, the limitation of this study lies in the use of secondary data sourced from one particular population group, so the generalizability of the findings still needs to be further examined in a more diverse context.

To cite this article (in APA style):
Sumin, S., Khairawati, K., & Kamaruddin, M. S. (2025). Psychometric properties of the general conspiracy belief scale using item response theory. *Jurnal Penelitian dan Evaluasi Pendidikan, 29(1)*, 85-97 doi: https://doi.org/10.21831/pep.v29i1.84569

INTRODUCTION

Psychometric measuring instruments can be divided into two groups based on the aspects measured: test and non-test instruments. Test instruments are used to measure cognitive aspects such as knowledge or skills, while non-test instruments are used to measure aspects of attitude and personality. A set of measuring instruments in tests and non-tests is considered quality if it meets several requirements or criteria, including valid, reliable, and bias-free (Kaplan & Saccuzzo, 2017). Item parameters, such as item difficulty, item discrimination, and false guesses, can influence validity and reliability in test instruments. Meanwhile, the psychometric properties of items on non-test instruments can be influenced by several factors, such as one's ability to reduce concepts, constructs, and indicators into measurement items. However, the item analysis tool is no less important in instrument development.

Two theories underlying psychometric item analysis exist: classical test theory (CTT) and IRT. Experts agree that CTT offers various conveniences, but CTT also has several weaknesses. To overcome the weaknesses of CTT, experts have developed IRT, which has several

10.21831/pep.v29i1.84659

advantages over CTT (Embretson & Reise, 2000). As an example of the application of this theory, we can look at the speculative issues that are currently rife. Amid the rise of speculative issues, including conspiracy theories often associated with various global phenomena, there is a profound need to ensure that instruments measuring belief in conspiracy theories have reliable psychometric properties. Brotherton et al. (2013) have developed a general conspiracy belief scale (GCBS) and validated it using a CTT approach. The study (Brotherton et al., 2013) was verified by Drinkwater et al. (2020); both studies successfully proved the validity and estimated the reliability of the general conspiracy theory belief scale using the CTT approach. However, given the limitations of CTT, especially on large samples, it is important to verify the psychometric properties of GCBS with more sophisticated approaches, such as IRT. Brotherton et al. (2013) explained that belief in conspiracy theories is now getting more attention from psychologists. Recent facts and phenomena, such as conspiracy theories around COVID-19 and the role of global elites, emphasize the importance of re-examining existing conspiracy theory measurement tools. Based on this background, we see the need to analyze the psychometric properties of the GCBS using the IRT approach, given that previous studies have only used the CTT approach.

Measuring conspiracy theory adoption has become an important focus in social psychology to understand the factors influencing the spread and acceptance of conspiracy narratives in society. One of the most recognized measurement tools in this field is the GCBS, developed by Brotherton et al. (2013). The scale is designed to measure general beliefs in conspiracy theories through five primary dimensions: alleged government malfeasance, belief in extraterrestrial cover-ups, perceived malevolent global conspiracies, perceived threats to personal well-being, and belief in control of information. This scale's construct validity and reliability have been confirmed through confirmatory factor testing, showing that the five dimensions represent a stable conspiracy belief structure psychometrically.

Previous studies have developed and adapted the GCBS to address research needs in diverse social and cultural contexts. Kaplan (2024) developed an additional dimension that integrates grandiose and vulnerable personality factors into the measurement of conspiracy theories, suggesting that individual psychological characteristics have an important role in driving conspiratorial thinking tendencies. Other studies have also highlighted the relevance of GCBS in more specific social contexts, such as the COVID-19 pandemic. Van Prooijen and Douglas (2017) found that individuals with low trust in the government were likelier to believe conspiracy theories related to the pandemic. This study supports the findings of Van Prooijen and Douglas (2017), which show that conspiracy theories tend to flourish in situations of crisis or social uncertainty. In addition, the relevance of GCBS in a digital context has also been studied. Pennycook et al. (2018) showed that this scale can measure the relationship between exposure to misinformation on social media and the level of belief in conspiracy narratives. This research underscores the importance of social media algorithms in amplifying sensationalized content that often supports conspiracy theories. Validation of GCBS was conducted by Lantian et al. (2016), who modified and expanded the items on the GCBS to improve its reliability and validity in the context of contemporary conspiracy theories. However, studies of the GCBS's psychometric properties through modern approaches such as IRT are still rare, especially when utilizing the original data from the scale developers. This study's novelty lies in the comprehensive application of polytomous IRT models such as GRM, PCM, and GPCM to re-evaluate the GCBS based on the same raw data as the original study. This approach allows for a more objective comparison of item quality and measurement accuracy while providing new insights into model fit in the context of conspiracy theory beliefs.

Assessment of the psychometric properties of the GCBS is still lacking in-depth studies, primarily through modern approaches such as IRT. Previous studies, such as those conducted by Brotherton et al. and Drinkwater et al., relied on the CTT approach to evaluate validity and

10.21831/pep.v29i1.84659

reliability. However, this method has significant limitations. CTT's dependence on sample size often leads to unstable parameter estimates, especially in large samples. The IRT approach offers a more accurate solution independent of sample size, making it important to improve the understanding of the psychometric properties of the GCBS in various populations and contexts.

According to Retnawati (2014), typical assumptions made about data can be used to classify polyatomic IRT as a model with nominal or ordinal degrees of measurement. The nominal item response model applies to items that contain multiple answer choices, which are not arranged in any particular order, and which assess different levels of ability. At the same time, the ordinal response model allows items to be evaluated based on a specific number of categories, for example, using a Likert Scale. Likert scales are scored using ordinal scores, which are determined by following the scoring parameters of the ordered answer categories. Item Response Theory (IRT) is the statistical approach underlying the analysis of the relationship between an individual's latent abilities and responses to test items (Hambleton et al., 1991). IRT has become the basis for developing measurement instruments, especially in data with more than two response categories (polytomous). When data are polytomous, the selection of an appropriate IRT model becomes important to analyze the psychometric properties of the Item. The GRM, PCM, and GPCM are the three main models in polytomous IRT that are often used by researchers (Hambleton et al., 1991; Ostini & Nering, 2006; Sözer & Kahraman, 2021). Each model has unique characteristics that determine its use according to the type of scale and data being analyzed.

IRT works optimally with some basic assumptions that must be met. First, the assumption of Unidimensionality states that each set of items only measures one dimension of latent ability. Second, local independence ensures that responses to other items do not influence responses to an item once latent abilities are controlled for. Third, item parameter invariance ensures that item parameters (difficulty, discrimination, and guessing) are independent of a particular population, resulting in a consistent measurement scale (Kim, 2018; Ostini & Nering, 2006; Sözer & Kahraman, 2021). The sample size required in IRT analysis depends on the complexity of the model and the number of parameters being estimated. The minimum recommended sample size for simple models such as PCM is 200-300 respondents. For more complex models such as GPCM, the optimal sample size ranges from 500-1,000 respondents (Dai et al., 2021). In addition, the more response categories and items analyzed, the larger the sample size needed to ensure the stability of parameter estimates. IRT offers greater flexibility and accuracy in evaluating item quality and individual ability than CTT. One of the main advantages is its ability to generate item parameters independent of the respondent population. This study aims to fill this gap by applying a polytomous IRT approach to evaluate GCBS in more depth. This study's state of the art lies in the comparative approach of three IRT models; GRM, PCM, and GPCM, to determine the model that best fits the ordinal scale data structure. With this analysis, the study shows the statistically best model and explains how each Item contributes to measurement accuracy. An important contribution of this study to the development of measurement theory is that it extends the applicability of IRT to non-cognitive instruments of a socio-psychological nature and shows that item-level analysis can improve overall measurement quality. This study also provides a methodological basis for future researchers who wish to develop or evaluate similar scales in a broader context.

RESEARCH METHOD

We used a replication research method with a quantitative approach. Replication research aims to repeat or replicate a study that has been conducted previously to test the truth, reliability, and validity of previous research results. Replication research can be conducted using the same or similar research design as the original research (Creswell & Creswell, 2017). The primary

10.21831/pep.v29i1.84659

purpose of replication research is to extend the generalizability of previous research results and ensure that the research results are reliable and consistent.

The sample from Brotherton et al. (2013) was randomly drawn using a lottery method on undergraduate students from several universities in the UK and Ireland. The sample size used in the Brotherton et al. (2013)study was 2,495 respondents was 2,495 respondents. Most participants were female (77.9%) in the UK and Ireland (75.7%), aged between 18 and 59 years. The data can be downloaded on the open-access psychometrics website via the following URL address: http://openpsychometrics.org/_rawdata/GCBS.zip. The instrument of this study is the GCBS, which was explored through the research of Brotherton et al. (2013). The GCBS is a unidimensional measurement consisting of 5 aspects, namely, government malfeasance (GM), extraterrestrial cover-up (ET), evil global conspiracy (MG), personal well-being (PW), and information control (CI). Each aspect is only measured with 3 statement items. The GCB scale was designed using a Likert scale with five categories; each response was given a qualitative label: 1: not accurate, 2: probably not true, 3: undecided, 4: probably accurate, 5: true, while respondents who did not respond were given a score of 0 (Brotherton et al., 2013).

According to the study of Brotherton et al. (2013), and Drinkwater et al. (2020), the GCBS has met content validity, factor validity, construct validity (discriminant validity and convergent validity), internal consistency, and test-retest reliability (Drinkwater et al., 2020; Shapiro et al., 2016). A full description of the general conspiracy theory belief scale can be seen on the open-access psychometrics website at the following URL address: http://openpsychometrics.org/tests/GCBS/. This study analyzed the psychometric properties of the GCBS using a polytomous item response theory approach. We analyzed the data using the "mirt" package with R Studio version 4.1.3 (Team, 2013). The syntax of the R program in this study can be seen in the appendix. We used three types of Polytomous IRT to find the best model, namely, the Partial Credit Model (Embretson & Reise, 2000; Masters, 2016), the Graded Response Model (Samejima, 1997), and the Generalized Partial Credit Model (Muraki, 1992). We did not check the instrument's dimensionality, validity, and reliability because it was done in the study of Brotherton et al. The model fit criteria used were the smallest AIC, Chi-Square Probability >0.05, RMSEA <0.08, and TLI and CFI >0.9 (Cangur & Ercan, 2015; Hu & Bentler, 1999) the best model will be selected when more than one model fit criteria have been met.

FINDINGS AND DISCUSSION

Findings

Assumption of Unidimensionality

Assumption of Unidimensionality Brotherton et al. (2013) successfully proved that the GCBS scale used in this study is a unidimensional measurement with five factors. Referring to the study, in this study, we did not check the assumption of Unidimensionality because we did not create or develop a new instrument, but through this study, we intend to re-analyze the psychometric properties using the Polytomic IRT approach on the same scale and data as the studies of Brotherton et al. (2013) and Drinkwater et al. (2020).

Assumption of Local Independence

Local independence is a constant ability that affects test performance or can be seen as an individual's response to a particular question independent of others (Sözer & Kahraman, 2021). Several statistical approaches, including Yen's Q3 (Chen & Thissen, 1997) and Jack-knife statistics, can be used to evaluate the assumption of local independence. The Q3 statistic developed by Yen (1984) considers the relationship between pairs of items. In the first stage,

10.21831/pep.v29i1.84659

Item and individual attributes are estimated using an IRT model that fits the data; the next stage forms a residual matrix from the residuals of each Item. Finally, the relationship between items is assessed using Yen's Q3 criterion; items do not meet the local independence assumption if the correlation is >0.2. According to some studies, local independence is automatically met if the assumption of Unidimensionality is met (Embretson & Reise, 2000; Hambleton & Swaminathan, 1985). The factor analysis results on the GCB scale used in this study indicate that the items included have a unidimensional structure. Since the criterion of Unidimensionality has been met, the premise of local independence has also been met.

Fit Model Selection

The best unidimensional polytomous IRT model was selected from the three models proposed in this study. The goodness of fit model criteria or model fit indices not only select the best model (AIC and BIC) but also test the fit of the empirical data to the conceptual model (i.e., Chi-Square probability and RMSEA). Using the R program syntax for unidimensional polynomial IRT developed by Desjardins & Bulut (2018), the statistical values for the five model fit criteria are shown in Table 1.

Table 1. The Goodness Of Fit Model

Model	AIC	BIC	p.M2	RMSEA	TLI	CFI
GPCM	94364.090	94882.250	0.000	0.061	0.785	0.855
GRM	93600.630	94118.790	0.000	0.051	0.850	0.899
PCM	94741.560	95178.210	0.000	0.087	0.561	0.570

Based on the analysis results presented in Table 1, the polynomial IRT model was evaluated by considering five main criteria: AIC, BIC, M2 statistics, RMSEA, TLI, and CFI. Each parameter provides an important indication of the model's fit to the empirical data used. The GRM showed the best performance compared to the GPCM and PCM. The AIC value of GRM is 93600.630, lower than GPCM (94364.090) and PCM (94741.560), indicating better model efficiency. The BIC value of GRM is also smaller, at 94118.790, compared to GPCM (94882.250) and PCM (95178.210), indicating that GRM has the best balance between model complexity and data fit. The M2 statistic, developed by Maydeu-Olivares and Joe (2006), was used to measure the model's fit to the data by evaluating item residuals. The M2 values for all models showed significant results (p = 0.000), signifying a difference between the observed data and the theoretical model. However, the interpretation of M2 requires support from other parameters, such as RMSEA, to provide a more complete picture of model fit. The RMSEA value for GRM is 0.051, which is smaller than that of GPCM (0.061) and PCM (0.087). This value indicates a low level of estimation error and better model fit by the accepted RMSEA standard (<0.06). In addition, the TLI and CFI values for GRM were 0.850 and 0.899, respectively, much higher compared to GPCM (TLI = 0.785, CFI = 0.855) and PCM (TLI = 0.561, CFI = 0.570). The high TLI and CFI values indicate that the GRM fits the empirical data.

GRM showed the best performance in accommodating polynomial response data compared to other models. The superiority of GRM is seen in the low AIC and BIC values and the RMSEA, TLI, and CFI values that indicate optimal model fit. M2 statistics provided additional insights, although significant results in all models emphasized combining information from multiple parameters to determine the best model. The use of GRM provides strong justification for analyzing this data, supporting the accuracy and validity of the measurement results theoretically and empirically. Further analysis using the GRM model obtained chi-square model accuracy, chi-square free degree, chi-square probability and RMSEA for each, as shown in Table 2.

10.21831/pep.v29i1.84659

Table 2. Item Fit

Items	χ^2	df	RMSEA	P
Q15	201.034	149	0.012	0.003
Q10	243.951	187	0.011	0.003
Q9	212.090	161	0.011	0.004
Q11	203.707	163	0.010	0.017
Q8	218.907	183	0.009	0.036
Q14	190.810	162	0.008	0.060
Q7	186.642	161	0.008	0.081
Q6	178.291	157	0.007	0.117
Q3	189.864	168	0.007	0.119
Q1	182.932	165	0.007	0.161
Q13	175.462	162	0.006	0.222
Q2	181.993	169	0.006	0.234
Q12	155.874	148	0.005	0.313
Q4	155.295	153	0.002	0.433
Q5	171.952	177	0.000	0.593

Based on the criteria for the accuracy of each Item with the data presented in Table 2, of the 15 GCBS items, five items have a chi-square probability <0.05 (ideal>0.05), namely Q15, Q10, Q9, Q11, Q8. However, all items have an RMSEA value <0.05, indicating that all GCB scale items represent the conceptual model (by the theory). The a_j parameter is a numerical representation of an item's psychological uncertainty in the context of data structure. In general, a higher parameter value indicates that the Item has a more well-defined meaning, and conversely, a lower indicates the Item has a less well-defined meaning.

Table 3. Item Parameter Estimation

Items	a	b1	b2	b3	b4	b5
Q1	2.152	-4.272	-1.25	-0.691	-0.283	0.573
Q2	2.199	-3.338	-0.9	-0.229	0.255	0.976
Q3	1.836	-3.954	0.164	0.639	1.162	1.784
Q4	2.508	-3.465	-0.557	0.021	0.489	1.32
Q5	1.826	-3.833	-1.193	-0.575	-0.110	0.884
Q6	2.407	-3.614	-0.912	-0.357	0.082	0.82
Q7	2.364	-3.503	-0.536	0.059	0.464	1.145
Q8	1.721	-3.986	-0.192	0.237	0.741	1.276
Q9	2.152	-3.503	-0.143	0.455	0.901	1.552
Q10	1.409	-1.696	-1.001	-0.386	0.791	NA
Q11	2.210	-3.522	-1.223	-0.609	0.002	0.885
Q12	2.833	-3.061	-0.49	0.033	0.463	1.105
Q13	1.981	-3.53	0.041	0.532	1.116	1.719
Q14	2.295	-3.93	-0.828	-0.231	0.239	1.008
Q15	1.747	-5.278	-2.349	-1.79	-1.248	-0.163

Based on the results of estimating item parameters using the EAP estimator, GRM using the R program produces item differentiation parameters that fluctuate between 1.721 and 2.833. The discriminating power threshold criteria are 0.01-0.34, very weak discriminating power, 0.35

ISSN (print) 2685-7111 | ISSN (online) 2338-6061

10.21831/pep.v29i1.84659

-0.64 low discriminating power, 0.65-1.34 medium discriminating power, 0.65-1.34, which means that medium discriminating power and 1.35-1.69 means that high discriminating power, 1.70 or more means that very high discriminating power (Baker, 2001). Referring to these criteria, item Q10 has medium discriminating power, while the other items have high discriminating power. EAP estimation results in GRM for the category of threshold parameters and location parameters for each Item are shown by b5 parameters to indicate the position of instrument items on the scale of latent traits or specific abilities. Location parameters with negative values indicate that the category is an easy choice according to respondents; location parameters with 0 to 1 indicate that the response category is relatively tricky. Location parameters> 1 indicate that the response category is a difficult choice.

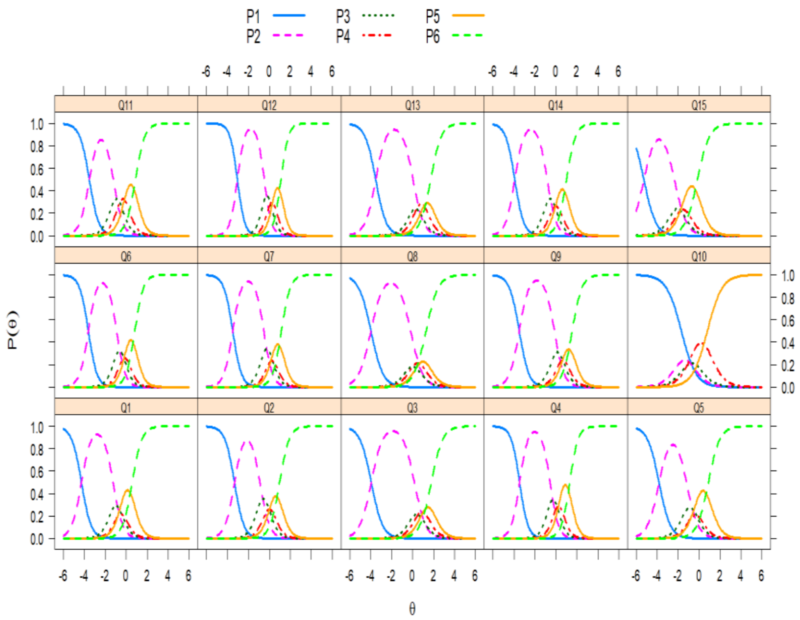

Figure 1. Item Category Information Curve

Category b1 in Item 1 through Item 15 has a difficulty parameter that varies between -5.278 to -1.696, which means that response category b1 is relatively more straightforward for respondents to choose. Response category b2 has parameters ranging between -2.349 to 0.041; category 2 includes easy options on almost all items, except item 3 and item 13, with parameters between 0.041 and 0.164 with a medium category. Response category b3 has parameters that vary between -1.79 and 0.639; in category 3, seven items are categorized as easy, while the remaining eight items are categorized as moderate. Response category 4 has a parameter location between -1.247 to 1.162. Response category four on Item Q1, Item Q5 and Q15 are classified as easy; on Q3 and Q13, response category four is classified as complex. While the other items' response category four is classified as moderate. The parameter location for Category 5 varies between -0.163 to 1.784. The difficulty level of response category five on Item Q15 is classified as easy, on items Q1, Q2, Q5, Q6, and Q11 as moderate, while the rest are classified as complex. In item 9, response category 5, no one chose, so the R program produces a parameter location that is not available (NA).

The category characteristic curve (CCC) or option characteristic curve (OCC) is an extension of the Item characteristic curve (ICC) specifically for polytomous items. Polytomous items have more than two response categories. The OCC curve shows the probability of a test

ISSN (print) 2685-7111 | ISSN (online) 2338-6061

10.21831/pep.v29i1.84659

taker or respondent choosing one of the specific response options as a function of the latent trait.

The OCC curve in Figure 1 shows that each Item has six categories. Program R, by default, performs regrouping for the responses given by the respondents. Category P1 indicates that the respondent did not answer, while P2 is not valid, P3 is probably not true, P4 is undecided, P5 is probably true, and P6 is true. Ability characterized by θ indicates the respondent's interest or level of understanding of conspiracy theories, -2 to -6 indicates that the respondent has low understanding, interest, or participation in conspiracy theories, and -2 to 2 indicates moderate understanding, interest or participation in conspiracy theories. In contrast, categories 2 to 6 indicate understanding, interest or participation in conspiracy theories. Options P1 and P2 (blue and magenta colours) were selected by respondents who had a low understanding, interest or participation in conspiracy theories; Option P3 (dark green colour), option P4 (red colour) and P5 (yellow colour) were selected mainly by respondents who had a moderate understanding, interest or participation in conspiracy theories. Meanwhile, option P6 (green colour) was chosen mainly by respondents with a high understanding, interest or participation in conspiracy theories.

Overall, the response categories (answer options) of each GCBS can measure respondents' understanding, interest and participation in conspiracy theories; this is evident from each option of each Item chosen by respondents according to their understanding, interest or participation in conspiracy theories, P1-P3 spread on the left side (negative), P4-P6 spread on the right side (positive), and none of the items have options that respondents misinterpret. Furthermore, we can also assess the item information function (IIF) curve generated by GRM with the R program. The item information function curve shows the amount of information each Item can explain on various latent traits at various attribute levels (Muraki, 1992).

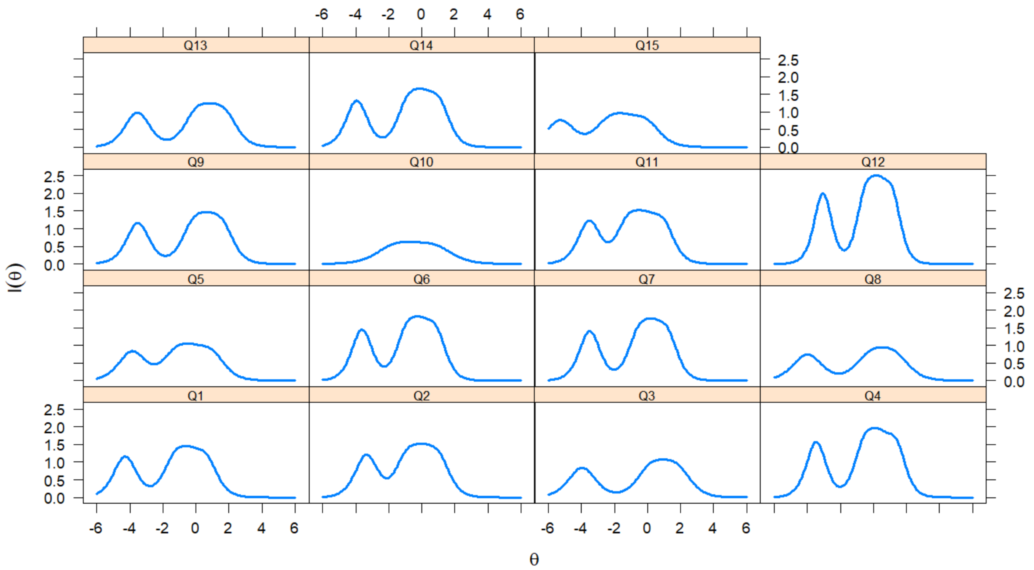

Figure 2. Item Information Function (IIF) Curve

Based on Figure 2, almost all items provide optimal information at θ -4 and $\theta = 0$ with explainable information between 1 to 2.5 and provide low information at θ 4, except Item Q6, which only optimally explains information at $\theta = 0$ with the resulting information function around 0.6. The IIF curves appear to vary, indicating that the polytomous options on each Item function well, except for Item Q6, which is less functional. Item Q6 only functions and provides optimal information, if delivered to respondents with medium ability. In contrast, other items

ISSN (print) 2685-7111 | ISSN (online) 2338-6061

10.21831/pep.v29i1.84659

will provide optimal information if delivered to respondents with low, medium and high ability. Almost all items on the GCBS provide relatively low information at abilities -4, -2 and +4, indicating that the GCBS items should be more suitable for measuring conspiracy theory beliefs in respondents with medium to low knowledge.

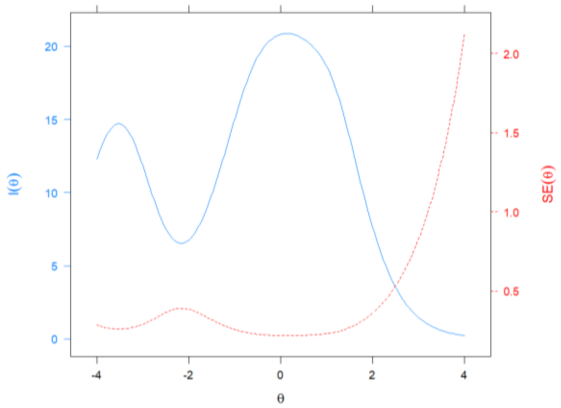

Figure 3. Information Curve and Measurement Standard Error

The information and standard error of measurement curves in Figure 3 show the contribution of GCBS in providing information on various latent traits with the lowest measurement error. Visually, it can be seen that the GCBS as a whole provides optimal information function at $\theta =$ -4 to about 2.5; this can be seen from the high IIF and low SE. The GCBS cannot provide good information at abilities 2.6 to 4. This means that the items on the GCBS are very suitable for measuring respondents' understanding, interest and participation with low to moderate abilities ($\theta =$ -4 to 0). We can estimate various reliability coefficients In CTT, while in Polytomous IRT modelling using GRM, we estimate reliability using the empirical marginal reliability approach. Empirical marginal reliability is the correlation between latent traitsθ and standard errors. We can assess empirical marginal reliability using the correlation between the standard error and the curve θ through Program R.

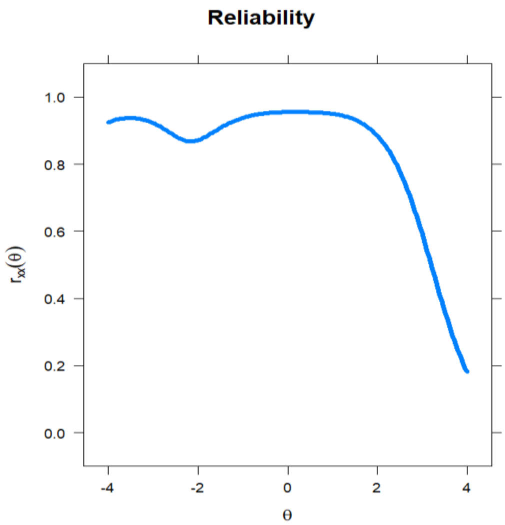

Figure 4. Empirical Marginal Reliability Curve

Based on the empirical marginal reliability curve in Figure 4, the curve has an optimum point around $4 \leq \theta \leq 2.8$, with a marginal reliability coefficient of 0.936. This shows that the

ISSN (print) 2685-7111 | ISSN (online) 2338-6061

10.21831/pep.v29i1.84659

GCBS scale is a measuring tool that has good consistency. Finally, we can interpret the distribution of respondents' abilities generated through GRM modelling using the R program; the distribution can be seen from the following histogram.

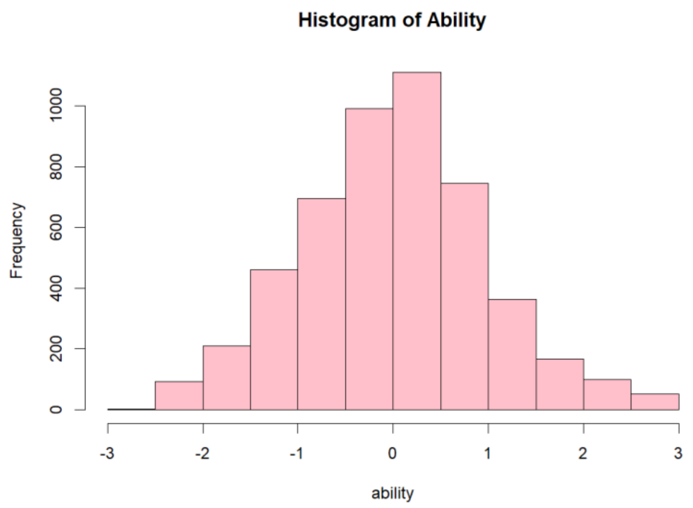

Figure 5. Distribution Histogram of Ability

The ability distribution histogram generated by GRM using the R program in Figure 5 visually shows that the majority of respondents who gave responses on the GCBS were low to medium-ability *students* ($-3 \leq \theta \leq 0.5$), or the majority of students who lack a strong interest in conspiracy theories.

Discussion

The results of this study indicate that the GRM is the most appropriate polytomous IRT model to analyze the psychometric properties of the GCBS. This finding aligns with the characteristics of GRM, which is theoretically and empirically recommended for ordinal scaled data such as Likert scales, as it can capture response differences more accurately (Samejima, 1969). All items in the scale showed appropriate parameters, with most having high power to discriminate. The findings of this study indicate that the scale can effectively identify variations in the level of belief in conspiracy theories. One Item, Q10, showed moderate discrepancy but still functioned well in the context of construct measurement. These results reinforce the structural validity and internal consistency reported by Brotherton et al. (2013) and reconfirmed by Drinkwater et al. (2020) through the CTT approach. By applying GRM, this study provides a more precise approach to evaluating item quality, particularly in the context of non-cognitive scales, as also suggested by Edwards & Kilpatrick (1948), who emphasized the importance of discrimination parameters in validating attitude-based scales.

This finding supports the study of Brotherton et al. (2013), who previously used the CTT approach to assess the validity and reliability of the GCBS. The CTT approach successfully identified that the scale could distinguish individuals with low, medium, and high abilities in understanding conspiracy theories. However, as Drinkwater et al. (2020) explained, CTT has significant limitations as item parameter estimates depend highly on sample size. GRM showed an advantage in this study as it produces item parameter estimates independent of sample size, providing more stable and accurate results. The response categories on the GCBS show a varied distribution of difficulty levels. Response categories 1 and 2 are more frequently selected by respondents with low understanding of conspiracy theories, while categories 4 and 5 reflect higher levels of understanding. This response distribution reflects the discrimination ability of the scale in measuring the level of belief in conspiracy theories, as described by Kaplan (2024), who highlighted the importance of instruments that can reflect a spectrum of beliefs based on individual psychological factors.

The items on the GCBS also show optimal information functions at low to medium ability levels (-4 to 0). This finding is consistent with the results of Lantian et al. (2016), which showed that the GCBS was designed to reach individuals with a moderate understanding of conspiracy theories. Item Q6 was an exception as it provided optimal information only at a moderate level of ability, indicating that this Item has room for improvement in measuring ability across a broader spectrum of ability. However, the item information curves (IIFs) show that all items significantly contribute to the scale's information function. The empirical marginal reliability coefficient of 0.936 indicates excellent consistency in measurement, as reported by Retnawati (2014), who emphasized the importance of high reliability in IRT-based analysis. Thus, the results of this study confirm that the GCBS does not require item deletion or calibration and is reliable for measuring beliefs in conspiracy theories.

This study also provides practical implications for the development of psychometric instruments. The GRM-based analysis allows researchers to identify items that provide optimal information at different ability levels of respondents, thereby increasing the efficiency and validity of the instrument. This study confirms that GRM is a more sophisticated approach than CTT, as Hambleton et al. (1991) proposed, especially in populations with heterogeneous ability distributions. Overall, this study verifies the results of Brotherton et al. (2013) using the CTT approach but with the additional advantages IRT offers, particularly GRM. This finding is relevant to the need to develop more adaptive and responsive instruments to changing social contexts, including crises such as the COVID-19 pandemic, as expressed by Van Prooijen and Douglas (2017). Based on these findings, the GCBS can be considered a robust and flexible measurement tool for various social and educational psychology research contexts.

CONCLUSION

The results of this study show that the polytomous IRT approach, specifically the GRM, is the most appropriate method for analyzing the psychometric properties of the GCBS. The GRM showed superior model fit and could identify items with high distinctiveness, which were informative for respondents with low to moderate conspiratorial beliefs. This finding addresses the main objective of the research, which is to re-evaluate the performance of the GCBS with a more modern and precise approach than the previous classical approach. These results' implications indicate that the use of IRT can improve measurement accuracy in the study of complex social beliefs and open up opportunities for applying GCBS across various cultural contexts and populations. In addition, this approach reinforces the importance of item-based analysis in developing non-cognitive instruments, particularly in social psychology and critical digital literacy. The main contribution of this study lies in the comprehensive application of the polytomous IRT model to the original GCBS developer dataset, which has not been done in many previous studies. Thus, this study not only expands the understanding of the GCBS's validity and reliability but also provides a relevant methodological reference for other belief and attitude scale research in the current information age.

Conflict of interests

There are no known conflicts of interest associated with this publication.

10.21831/pep.v29i1.84659

REFERENCES

Baker, F. B. (2001). The basics of item response theory. ERIC.

Brotherton, R., French, C. C., & Pickering, A. D. (2013). Measuring Belief in Conspiracy Theories: The Generic Conspiracist Beliefs Scale. Frontiers in Psychology, 4. https://doi.org/10.3389/fpsyg.2013.00279

Cangur, S., & Ercan, I. (2015). Comparison of Model Fit Indices Used in Structural Equation Modeling Under Multivariate Normality. Journal of Modern Applied Statistical Methods, 14(1), 152–167. https://doi.org/10.22237/jmasm/1430453580

Chen, W.-H., & Thissen, D. (1997). Local dependence indexes for item pairs using item response theory. Journal of Educational and Behavioral Statistics, 22(3), 265–289. https://psycnet.apa.org/doi/10.2307/1165285

Creswell, J. W., & Creswell, J. D. (2017). Research design: Qualitative, quantitative, and mixed methods approaches. Sage publications.

Dai, S., Vo, T. T., Kehinde, O. J., He, H., Xue, Y., Demir, C., & Wang, X. (2021). Performance of Polytomous IRT Models With Rating Scale Data: An Investigation Over Sample Size, Instrument Length, and Missing Data. Frontiers in Education, 6. https://doi.org/10.3389/feduc.2021.721963

Desjardins, C. D., & Bulut, O. (2018). Handbook of educational measurement and psychometrics using R. CRC Press.

Drinkwater, K. G., Dagnall, N., Denovan, A., & Neave, N. (2020). Psychometric assessment of the Generic Conspiracist Beliefs Scale. PLOS ONE, 15(3), e0230365. https://doi.org/10.1371/journal.pone.0230365

Edwards, A. L., & Kilpatrick, F. P. (1948). A Technique for the Construction of Attitude Scales. Journal of Applied Psychology, 32(4), 374–384. https://doi.org/10.1037/H0057313

Embretson, S. E., & Reise, S. P. (2000). Item response theory for psychologists. Lawrence Erlbaum Associates, Publishers.

Hambleton, R. K., & Swaminathan, H. (1985). Item response theory: Principles and applications. Nijhof'f Publishing.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). Fundamentals of item response theory (Vol. 2). Sage.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal, 6(1), 1–55. https://doi.org/https://doi.org/10.1080/10705519909540118

Kaplan, C. (2024). Relationships Between Pseudoscience, Conspiracies, Paranormal Ideas, and Personality Factors.

Kaplan, R. M., & Saccuzzo, D. P. (2017). Psychological testing: Principles, applications, and issues. Cengage Learning.

Kim, J. (2018). Extensions and applications of item explanatory models to polytomous data in item response theory. University of California, Berkeley.

ISSN (print) 2685-7111 | ISSN (online) 2338-6061

10.21831/pep.v29i1.84659

Lantian, A., Muller, D., Nurra, C., & Douglas, K. M. (2016). Measuring belief in conspiracy theories: Validation of a French and English single-item scale. International Review of Social Psychology, 29(1), 1–14.

Masters, G. N. (2016). Partial credit model. In Handbook of Item Response Theory, Volume One (pp. 137–154). Chapman and Hall/CRC.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. Applied Psychological Measurement, 16(2), 159–176. https://doi.org/10.1177/014662169201600206

Muraki, E. (1993). Information functions of the generalized partial credit model. ETS Research Report Series, 1993(1), i–12.

Ostini, R., & Nering, M. L. (2006). Polytomous item response theory models (Issue 144). Sage.

Pennycook, G., Cannon, T. D., & Rand, D. G. (2018). Prior exposure increases perceived accuracy of fake news. Journal of Experimental Psychology: General, 147(12), 1865.

Retnawati, H. (2014). Teori respons butir dan penerapannya: Untuk peneliti, praktisi pengukuran dan pengujian, mahasiswa pascasarjana. Nuha Medika.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. Psychometrika Monograph Supplement.

Samejima, F. (1997). Graded response model. In Handbook of modern item response theory (pp. 85–100). Springer.

Shapiro, G. K., Holding, A., Perez, S., Amsel, R., & Rosberger, Z. (2016). Validation of the vaccine conspiracy beliefs scale. Papillomavirus Research, 2, 167–172. https://doi.org/10.1016/j.pvr.2016.09.001

Sözer, E., & Kahraman, N. (2021). Investigation of Psychometric Properties of Likert Items with Same Categories Using Polytomous Item Response Theory Models. Journal of Measurement and Evaluation in Education and Psychology, 12(2), 129–146.

Team, R. C. (2013). R: A language and environment for statistical computing.

Van Prooijen, J.-W., & Douglas, K. M. (2017). Conspiracy theories as part of history: The role of societal crisis situations. Memory Studies, 10(3), 323–333.

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.

ISSN (print) 2685-7111 | ISSN (online) 2338-6061