e-ISSN 2987-6680
Transforma Jurnal Manajemen
Universitas Pasundan
DOI:https://doi.org/10.56457/tjm.v4i2.427

Analysis of the Characteristics of the Questionnaire Items for the Evaluation of LMS ASN with Integrity Using the Graded Response Model

[1]Rezky Budiman, [2]Yohanes Nugraha, [3]Arif Triyanto, [4]Atar Koswara,[5]Siswantoyo
Universitas Negeri Yogyakarta
E-mail: rezkybudiman.2026@student.uny.ac.id

ABSTRACT
This study aims to analyze the item characteristics of a questionnaire evaluating the "ASN with Integrity" (Integrity-Oriented Civil Servant) Learning Management System (LMS) using the Graded Response Model (GRM). A quantitative psychometric approach was employed, using response data from 13 questionnaire items, each measured on a five-category ordinal scale. The initial dataset consisted of 24,217 respondents, with 24,214 responses included in the GRM calibration process. The analysis encompasses checking model assumptions, estimating discrimination and threshold parameters, evaluating item fit, and analyzing the information function along with the Standard Error of Measurement (SEM). The results indicate that the model generally demonstrates a good fit, with an RMSEA of 0.034, CFI of 0.977, TLI of 0.966, and SRMSR of 0.089. All items exhibit positive and relatively high discrimination, with 'a' parameters ranging from 3.18 to 5.88. The threshold parameters (b1–b4) for all items are ordered sequentially, ensuring no reversals in the response categories. Item-level fit, assessed via RMSEA, also shows favorable results, with values ranging from 0.018 to 0.030. However, a maximum residual Q3 value of approximately 0.480 suggests potential local dependence between certain pairs of items, which warrants further investigation. The information function reveals that the instrument provides the most robust information and measurement precision at lower levels of the latent attribute—up to approximately $\theta = 0$—while precision decreases at higher, positive levels of the latent attribute. This is associated with a "ceiling effect," evidenced by the 12,229 respondents who assigned a score of 5 to all 13 items. Based on these findings, no items require immediate removal. GIO5 is a priority for review as it possesses the lowest discrimination power, while items with very high discrimination power should be examined for potential content redundancy. Overall, the instrument demonstrates good psychometric quality; However, further refinement is required to enhance its measurement capability at high levels of the latent attribute and to ensure local independence among the items.
Keywords:*Graded Response Model*, Item Response Theory, item characteristics, LMS, ASN with Integrity, psychometrics, information function

INTRODUCTION

Digital transformation in the implementation of civil servant competency development encourages the use of Learning Management Systems (LMS) as a learning tool, material management, assessment, and learning process monitoring. In the context of ASN with Integrity, the LMS not only functions as a medium for delivering material but also becomes part of the learning ecosystem that needs to be systematically evaluated so that the resulting information can truly be used to understand the experiences and responses of participants. Good evaluation requires instruments that can produce adequate measurement evidence and support score interpretation according to their intended use. This principle aligns with the Standards for Educational and Psychological Testing, which place validity, reliability, and fairness as essential components in the development and use of measurement instruments (AERA, APA, & NCME, 2014).

Copyright© Authors
(cc) BY

In practice, learning evaluation questionnaires often use a graded response scale, for example, from strongly disagree to strongly agree. This type of data is ordinal, so the information contained in each response category cannot always be adequately explained solely through classical statistics such as means, item-total correlations, or reliability coefficients. Item Response Theory (IRT)-based analysis offers a more detailed approach because it models the relationship between responses to an item and the underlying latent attribute (theta, θ). With this approach, the characteristics of each item can be studied at multiple levels of the latent attribute, rather than solely based on the total score.

For items with more than two response categories, the Graded Response Model (GRM) is a relevant IRT model. GRMs are developed for polytomous data and model the probability of a respondent being in or above each response category based on the level of a latent attribute. This approach allows researchers to gain insight into item discrimination and the threshold position between categories. More recent GRM tutorials emphasize that model application requires attention to unidimensionality, local independence, and the monotonicity of the relationship between latent attributes and the probability of higher category responses (DeMars, 2024).

The use of GRM is crucial in evaluating the LMS ASN Berintegrity questionnaire because the quality of an item cannot be assessed solely from the distribution of responses. Two items can have nearly identical means but differ in their ability to differentiate respondents at a particular latent attribute level. Similarly, response categories that lack a regular threshold pattern may indicate that the category selection is not functioning as designed. Thus, discrimination and threshold parameters can provide a more substantive picture of how each item performs.

In addition to item parameters, model fit and information function are important components of psychometric evaluation. Model-data fit helps determine whether an item's empirical response pattern can still be adequately explained by the GRM, while the information function indicates the degree of measurement precision at various θ positions. In IRT, greater information is associated with smaller measurement errors at a given level of the latent attribute. Therefore, examining item information and test information can help determine whether an instrument is more informative for respondents with low, medium, high, or specific ranges of attributes.

A number of IRT literature places unidimensionality, local independence, and monotonicity as the main assumptions of ordinal IRT models. Violation of these assumptions can affect parameter estimation and result interpretation. The literature also shows that unidimensionality does not automatically guarantee local independence, as two items can still have residual correlation after controlling for the influence of latent attributes (Cappelleri et al., 2019; DeMars, 2024). Therefore, assumption analysis needs to be placed as a preliminary step before interpreting GRM parameters.

Based on the above description, this study aims to analyze the characteristics of the questionnaire items for evaluating the LMS for ASN with Integrity using a quantitative psychometric approach. The study does not stop at the question of whether the questionnaire scores are reliable, but rather seeks to explain how each item functions in measuring latent attributes. The analysis focuses on fulfilling GRM assumptions, discriminatory power parameters, category threshold parameters, item suitability, information functions, and identifying items that need to be reviewed or refined. The results of this study are expected to provide an empirical basis for improving the quality of the instrument so that its use in LMS evaluation becomes more accurate and informative.

LITERATURE REVIEW

Copyright© Authors
(cc) BY

LMS evaluation is essentially the process of collecting and interpreting evidence regarding the quality of the learning experience, ease of use, engagement, system support, and other aspects of the evaluation construct. When evaluating using questionnaires, each response is viewed as a manifestation of the latent attribute being measured. Because research decisions and conclusions depend on the interpretation of scores, the quality of the instrument needs to be supported by psychometric evidence appropriate to its intended use (AERA, APA, & NCME, 2014).

Item Response Theory views participant responses as a function of the latent attribute (θ) and item characteristics. Conceptually, it can be written as $P(X_i=k|\theta)$, which is the probability that a respondent with a latent attribute level θ will respond in category k for item i. Unlike approaches that only look at the total score, IRT allows for analysis of item characteristics and measurement accuracy at different levels of the latent attribute.

In IRT models, θ is typically placed on a continuous scale. Item parameters then describe the location and discrimination of the item on the continuum. This interpretation makes the analysis more informative, especially for instruments with graded response categories.

The GRM is a polytomous IRT model for ordered response categories. This model is suitable for data such as Likert scales because the categories have a substantive order, for example, 1 = strongly disagree to 5 = strongly agree. For item i with discrimination parameters a_i and threshold b_{ik}, the cumulative probability of selecting category k or higher can be expressed as:

$$P(X_i \geq k \mid \theta) = 1 / [1 + \exp\{-a_i(\theta - b_{ik})\}].$$

The exact probability in category k is obtained from the difference of two cumulative probabilities:

$$P(X_i = k \mid \theta) = P(X_i \geq k \mid \theta) - P(X_i \geq k+1 \mid \theta).$$

If there are K response categories, then each item has one a parameter and K−1 threshold parameters. For example, in five response categories, there is one a and four thresholds (b1–b4). Thresholds should increase sequentially in a well-functioning model, because higher category boundaries are expected to be reached at higher levels of the latent attribute.

The parameter a describes the item's ability to differentiate between respondents with different levels of the latent attribute. Visually, higher a values result in a steeper cumulative curve. Items with high a tend to exhibit more rapid changes in response probability as θ changes. However, interpretation should not rely solely on a single numerical threshold; construct context, model fit, and item information also need to be considered.

The threshold parameter b indicates the location on the θ continuum corresponding to the probability shift between categories. On a five-category scale, there are four thresholds. For example, b1 represents the boundary between category 1 and a higher category, while b4 corresponds to the shift to the highest category. A negative threshold indicates that the boundary is reached at an θ below the scale mean, while a positive threshold indicates a need for an θ higher than the mean.

Examination of threshold order is important because dysfunctional response categories can produce irregular thresholds. In such situations, researchers need to consider whether the item wording or response categories need to be reviewed.First, unidimensionality means that item responses are primarily explained by a single dominant latent attribute. This can be examined using exploratory or confirmatory factor analysis, parallel analysis, and factor structure examination. Second, local independence means that after θ is controlled, responses to one item have no substantial residual dependence on other items. Third, monotonicity means that the probability of selecting a higher category does not decrease as θ increases. These three assumptions are essential foundations for interpreting ordinal IRT models (Cappelleri et al., 2019; DeMars, 2024).

Copyright© Authors

In this study, model assumptions are treated as evidence that needs to be verified, not merely a formality. The results of the assumption check will be used to determine whether the GRM parameter estimates are interpretable. If any indication of violation is found, researchers need to investigate possible causes, such as the presence of multiple constructs, item redundancy, wording similarity, or dysfunctional response categories.

Item fit indicates the extent to which empirical response patterns are consistent with model-predicted responses. One approach that can be used is an item-fit statistic such as S-X^2, depending on the software and model implementation. Items that do not fit should be reviewed alongside substantive information, rather than immediately removed.

The item information function, $I_i(\theta)$, indicates the amount of information an item provides at a given level θ. The test information function is the sum of the information of all items:

$$I_T(\theta) = \Sigma I_i(\theta).$$

The relationship between information and the standard error of measurement can be written as:

$$SE(\theta) = 1 / \sqrt{I_T(\theta)}.$$

Thus, when the test information is high, the measurement error tends to be smaller. The information curve can be used to determine whether the instrument is most precise on the latent attribute.

This study uses questionnaire responses as observational data analyzed using GRM. The initial stage is an assumption check. Once the assumptions are adequate, the parameters a and b are estimated, and then item fit and information function are checked. The results of all analyses are used in an integrated manner to determine which items to retain, review, or refine.

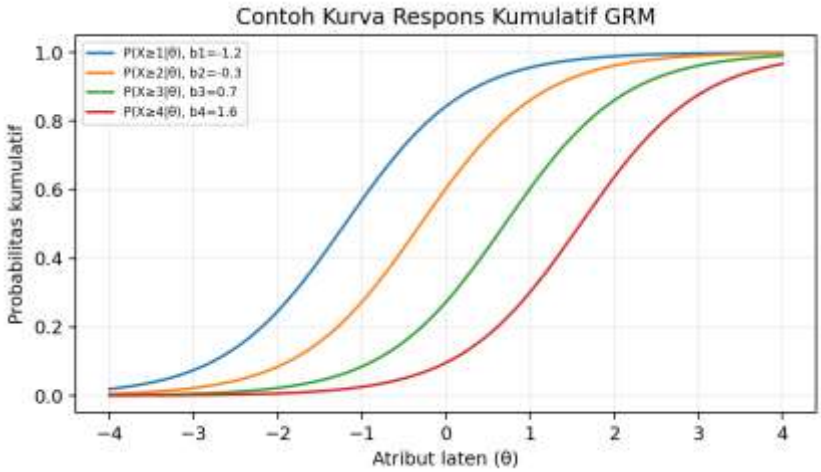
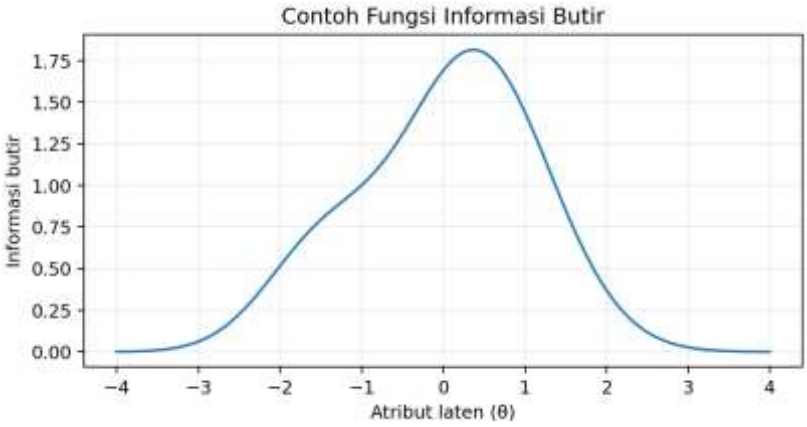


Figure 2.1. Example of cumulative response curves in GRM for four thresholds.



Copyright© Authors
(cc) BY

Figure 2.2. Example of item information function at various levels of latent attributes.

Table 2.1. Example of GRM parameter interpretation

Parameter	Meaning	Simple interpretation
a	Distinguishing power	The larger it is, the sharper the item's ability to differentiate respondents.
b1–b4	Category threshold	Shows the location of movement between categories on the θ scale.
Item fit	Model fit	Indicates whether the response pattern conforms to the GRM.
IIF	Item information	Indicates the level of grain accuracy at a certain θ.
TIF	Test information	The amount of information on all items at a given θ.
SE(θ)	Measurement error	The smaller, the more precise the measurement.

Psychometric literature indicates that GRM is widely used for instruments with Likert-type responses or ordinal categories. GRM research and tutorials emphasize the estimation of a single discrimination parameter and a number of thresholds depending on the number of response categories, as well as examination of category functions, item fit, and information functions. Recent studies have also shown that information curves can be used to accurately visualize the range of latent attributes being measured.

METHOD

This study employed a quantitative approach with a psychometric design. The quantitative approach was used because the research data consisted of participants' numerical responses to each item in the ASN Integrity LMS evaluation questionnaire, which were then analyzed using statistical techniques and Item Response Theory (IRT) modeling. Meanwhile, the psychometric design was used because the research focused on evaluating the characteristics and measurement quality of each questionnaire item, rather than simply describing respondents' levels of satisfaction or perceptions of the LMS.

The psychometric approach in this study is aimed at obtaining empirical evidence regarding item quality, response category function, discriminatory power, threshold parameters, model fit, and measurement accuracy. Thus, the primary object of analysis is not simply the respondent's total score, but the relationship between responses to each item and the latent attribute underlying that response pattern. This latent attribute is represented by the parameter θ, which describes the respondent's position on the construct measured by the LMS evaluation questionnaire.

Item characteristic analysis was conducted using the Graded Response Model (GRM), an IRT model for polytomous data with sequential response categories. The selection of the GRM was based on the characteristics of questionnaires that use graded response categories, such as the Likert scale. In the GRM, each item is analyzed based on the discriminatory power parameter (a) and the category threshold parameter (b). The a parameter is used to determine the item's ability to differentiate respondents with different levels of latent attributes, while the b parameter is used to describe the position of the threshold for movement between response categories on the latent attribute continuum.

Mathematically, the cumulative probability that a respondent with latent ability or attribute θ responds in category k or higher can be expressed as:

Copyright© Authors

$$P(X_i \geq k \mid \theta) = 1 / [1 + \exp\{-a_i(\theta - b_{ik})\}]$$

with a_i as the discriminating power parameter of the i-th item, b_ik as the threshold parameter of the k-th category, and θ as the respondent's latent attribute. Based on the cumulative probability, the probability of a respondent being in a category is obtained through the difference in the cumulative probabilities of two consecutive categories.

The psychometric analysis in this study was conducted in stages. The first stage was an examination of data characteristics and the fulfillment of model assumptions. This examination included unidimensionality and local independence, as IRT parameter estimates are based on the assumption that items primarily measure the same latent attribute and that responses to one item are independent of responses to other items after the latent attribute is controlled.

The second stage is item parameter estimation using the GRM. At this stage, the a value and a number of b parameters are obtained for each item according to the number of response categories. The estimation results are then used to determine the discriminatory ability of each item and the pattern of movement between response categories.

The third stage is the analysis of category functions and item characteristics. This analysis can be visualized using the Item Characteristic Curve (ICH) or category response curve, which allows us to understand how the probability of each category changes along the θ continuum. This visualization is important because it provides a more understandable picture of the function of each response category.

The fourth stage is item fit analysis. This analysis is conducted to determine whether the empirical response patterns for each item are consistent with the patterns predicted by the GRM. Items that exhibit discrepancies require further review, taking into account item parameters, response category characteristics, and statement substance.

The fifth stage is information function analysis. The information function is used to determine the level of accuracy of each item in measuring the latent attribute at various θ values. The total test information is obtained by summing the information on all items:

$$I_T(\theta) = \Sigma I_i(\theta)$$

The relationship between information and measurement error can be expressed as:

$$SE(\theta) = 1 / \sqrt{I_T(\theta)}$$

Thus, latent attribute sections with high levels of information will yield relatively smaller measurement errors. This analysis allows researchers to determine whether the questionnaire is more accurate in measuring respondents with low, medium, high, or specific levels of the latent attribute.

The final stage is item evaluation and classification. Decisions regarding item quality are not based solely on a single parameter, but are made by considering discriminatory power, the order and position of threshold parameters, response category function, model fit, information function, and measurement error. Based on the overall evidence, items can be categorized as retained, needing review, or needing improvement.

With this approach, this study is expected to produce a more comprehensive psychometric evaluation of the ASN Integrity LMS evaluation questionnaire. The results not only show how respondents responded to the questionnaire but also provide information on which items provide the best measurement, at which latent attribute level an item performs optimally, and which items require improvement. This approach can ultimately support the development of LMS evaluation instruments that are more accurate, informative, and have a strong measurement basis.

Table 3.1 Stages of Psychometric Analysis

Stage	Analysis	Main output

Copyright© Authors

1	GRM data and assumptions check	Unidimensionality and local independence
2	GRM model estimation	Parameters a and b
3	Response category analysis	Category response curve/ICC
4	Item fit analysis	The suitability index of each item
5	Information analysis	IIF, TIF, and SE(θ)
6	Item evaluation	Items are maintained, reviewed, or refined

RESULTS AND DISCUSSION

The analysis began by comparing several factor model specifications to determine the structure that best fit the responses to the 13 questionnaire items. Four models were compared: one-factor, two-factor, three-factor, and four-factor. Comparisons were made using log-likelihood values, the Akaike Information Criterion (AIC), and the Bayesian Information Criterion (BIC). In model selection, lower AIC and BIC values indicate a better balance between model fit and model complexity.

Model	LogLik	AIC	BIC
1 Factor	-126262.7	252655.4	253181.5
2 Factors	-121461.1	243076.3	243699.6
3 Factors	-121501.9	243179.8	243892.1
4 Factors	-120821.9	241839.8	242633.1

Table 4.1 Comparison of models based on Logic, AIC, and BIC

Based on Table 4.1, the four-factor model yields an AIC of 241,839.8 and a BIC of 242,633.1. These two values are the lowest compared to the one-, two-, and three-factor models. Furthermore, the four-factor model has the highest (less negative) log-likelihood value, at -120,821.9. Therefore, based on the information criteria used, the four-factor model was selected as the most appropriate model among the compared model specifications.

These results also provide a methodological basis for further analysis using a multidimensional GRM with four factors. The choice of four factors was not solely due to the larger number of factors, but rather because the model demonstrated improved fit, as reflected in the decrease in AIC and BIC. In a psychometric context, the selected model must be interpreted in conjunction with the construct theory and the substantive meaning of each factor, so that the final decision is not solely statistical.

Model comparisons showed that increasing complexity from a one-factor model to a four-factor model resulted in improved overall fit. The one-factor model had the highest AIC and BIC, while the four-factor model produced the lowest. These findings suggest that respondents' response patterns are more accurately represented by a multidimensional structure rather than being imposed as a single dimension.

Substantively, these results are important because questionnaires comprising multiple construct aspects can produce responses influenced by more than one dimension of ability or latent characteristic. Therefore, using a four-factor model allows for a more detailed study of the characteristics of each item on the relevant dimensions.

Once the four-factor model has been established, the analysis continues by calculating the item information function. The information function indicates how much an item contributes to providing information about the respondent's latent ability position. The greater the information value at a given ability level, the smaller the expected measurement error in that ability area.

Copyright© Authors
(cc) BY

In multidimensional models, item information needs to be read with respect to the dimensions/factors being analyzed. Because multidimensional information functions require predetermined directions or angles, calculations are performed on a series of theta values for each factor. This procedure also explains why the initial command, which only inputs one-dimensional theta, resulted in an error message. Once the angles/factors are defined, information can be calculated and summarized for each item.

Item	Factor 1	Factor 2	Factor 3	Factor 4	Total
GIO9	5,602	27,163	16,893	18,908	68,565
GIO2	9,920	11,773	22,255	22,052	66,000
GIO4	8,196	11,171	15,573	27,512	62,452
GIO1	8,067	11,098	21,527	16,664	57,355
GIO8	4,894	21,854	13,605	16,882	57,235
GIO7	4,426	13,343	10,798	15,766	44,333
GIO5	5,873	6,939	9,812	19,080	41,704
GIO6	4,275	9,582	9,254	15,521	38,632
GIO10	2,505	12,645	8,815	11,019	34,984
GIO3	3,717	7,361	9,831	10,927	31,835
GIO11	0.601	5,635	4.144	6,721	17,102
GIO12	0.095	4,618	6,039	6,039	16,792
GIO13	0.000	3,878	3,878	3,878	11,633

Table 4.2 Maximum information rating of each item on the four factors

Table 4.2 shows that GIO9 has the highest maximum total information, namely 68.565. This value is followed by GIO2 at 66.000 and GIO4 at 62.452. Conversely, GIO13 has the lowest total information, namely 11.633, followed by GIO12 at 16.792 and GIO11 at 17.102. Thus, based on the maximum information generated in this analysis, GIO9, GIO2, and GIO4 are the most informative items in measuring the construct as a whole.

Interpretation of Information Based on Factors

Factor	Highest information	Mark	Lowest information	Mark
Factor 1	GIO2	9,920	GIO13	0.000
Factor 2	GIO9	27,163	GIO13	3,878
Factor 3	GIO2	22,255	GIO13	3,878
Factor 4	GIO4	27,512	GIO13	3,878

Table 4.3 Items with maximum and minimum information on each factor

In Factor 1, GIO2 provided the highest information at 9.920. In Factor 2, GIO9 provided the highest information at 27.163. In Factor 3, GIO2 again provided the highest information at 22.255. Meanwhile, in Factor 4, GIO4 provided the highest information at 27.512. This pattern indicates that the information contribution is not distributed equally across all items and factors.

These findings can be used to evaluate the instrument. Items that provide high levels of information on a particular factor can be retained as important indicators for that dimension. Conversely, items with very low levels of information require further review rather than immediate deletion. The decision to delete the item must consider the item's content, discrimination parameters, category parameters, construct representation, and the instrument's intended use.

The ranking results show significant variation in the informational capabilities of the items. GIO9 is the item with the highest total information, making it relatively strong

Copyright© Authors

in providing information about the respondent's latent position in the model used. GIO2 and GIO4 also demonstrate high informational contributions and occupy the following positions.

On the other hand, GIO13 had the lowest total information. This indicates that in the theta region used in the evaluation, this item contributed relatively little information compared to other items. Psychometrically, this condition could be caused by low discrimination power, the position of the category parameters not matching the distribution of respondents' abilities, or item characteristics that were less able to differentiate respondents in certain ability regions. However, the specific cause cannot be determined solely from the maximum information value.

Overall, the analysis results support the use of the four-factor model as a basis for further analysis of item characteristics. Item information can be used to determine which items contribute most to the measurement of each dimension. This information can also be used as a basis for instrument refinement, particularly if further research aims to improve measurement accuracy across specific ability ranges.

However, the information function should not be the sole basis for assessing item quality. A comprehensive evaluation should incorporate GRM parameter estimates, response category characteristic curves, item/model fit tests, examination of dimensionality and local independence assumptions, and consideration of content validity. This approach allows for statistically and substantively stronger decisions about whether to retain, revise, or delete an item.

Summary of Results Based on the Order of Research Objectives

No.	Objective	Results	Meaning
1	Determining the most appropriate model structure	The four-factor model had an AIC of 241839.8 and a BIC of 242633.1, both the lowest; a LogLike of -120821.9 was the highest (less negative).	The four-factor model was selected for further analysis.
2	Analyzing item information in multidimensional GRM models	Maximum information was calculated on four factors; there was variation in information contribution between items.	Each item has a different contribution to each factor.
3	Identifying the most informative items	GIO9 has the highest total information of 68,565; followed by GIO2 with 66,000 and GIO4 with 62,452.	These three items are the most informative indicator candidates based on these results.
4	Identifying items with relatively low information	GIO13 has the lowest total information of 11,633; GIO12 16,792 and GIO11 17,102.	Items need further review before a revision/deletion decision is made.

Table 4.4 Summary of research results based on objectives

CONCLUSIONS

Based on the results of the item characteristic analysis using the multidimensional Graded Response Model (GRM), it can be concluded that the four-factor structure is the most appropriate model among the compared models. The four-factor model produces the lowest AIC and BIC values, 241839.8 and 242633.1, respectively, and a log-likelihood value of -120821.9. These results indicate that the four-factor multidimensional structure provides a better fit to the data than the one-, two-, and three-factor models.

Copyright© Authors
(cc) BY

The information function analysis shows that each item's ability to provide information to the construct is not the same. GIO9 has the highest maximum total information of 68.565, followed by GIO2 at 66.000 and GIO4 at 62.452. At the factor level, GIO2 is the most informative item on Factors 1 and 3, GIO9 is the most informative on Factor 2, while GIO4 is the most informative on Factor 4.

In contrast, GIO13 had the lowest maximum total information at 11,633. GIO12 and GIO11 were also in the relatively low information group. These results indicate that some items make a smaller contribution to the measurement and therefore require attention in the next stage of instrument evaluation and development.

Overall, the multidimensional GRM approach provides more detailed information about measurement quality than simply looking at the total score. The analysis can identify the most informative items and the dimensions on which they provide the most information. Therefore, the research results can be used as an empirical basis for refining evaluation instruments.

Bibliography

AERA, APA, & NCME. (2014). Standards for educational and psychological testing. American Educational Research Association.

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

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

Samejima, F. (1997). Graded response model. In W. J. van der Linden & R. K. Hambleton (Eds.), Handbook of modern item response theory. Springer.

Rizopoulos, D. (2006). ltm: An R package for latent variable modeling and item response theory analyses. Journal of Statistical Software, 17(5), 1–25.

Frontiers in Education. (2021). Performance of polytomous IRT models with rating scale data: An investigation over sample size, instrument length, and missing data.

Frontiers in Psychology. (2026). Psychometric evaluation of a four-item scale derived from the PHQ-9 and GAD-7 among community adults based on multivariate generalizability theory and item response theory.

Kementerian Pendayagunaan Aparatur Negara dan Reformasi Birokrasi. (2026). Peluncuran nasional program e-learning ASN Berintegritas.

Catatan metodologis: contoh grafik pada BAB II bersifat ilustratif dan bukan hasil estimasi data penelitian. Nilai parameter pada grafik dibuat untuk menjelaskan konsep GRM secara visual.

Baker, F. B. (2001). The basics of item response theory (2nd ed.). ERIC Clearinghouse on Assessment and Evaluation.

de Ayala, R. J. (2009). The theory and practice of item response theory. Guilford Press.

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

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. Psychometric Monograph No. 17. Psychometric Society.

Kementerian Pendayagunaan Aparatur Negara dan Reformasi Birokrasi. (2017). Peraturan Menteri Pendayagunaan Aparatur Negara dan Reformasi Birokrasi Nomor 38 Tahun 2017 tentang Standar Kompetensi Jabatan Aparatur Sipil Negara. https://peraturan.bpk.go.id/Download/123418/PERMENPAN%20NOMOR%2038%20TAHUN%202017.pdf

Kementerian Pendayagunaan Aparatur Negara dan Reformasi Birokrasi. (2021). Peraturan Menteri Pendayagunaan Aparatur Negara dan Reformasi Birokrasi Nomor 90 Tahun 2021 tentang Pembangunan dan Evaluasi Zona Integritas Menuju Wilayah Bebas dari

Copyright© Authors

(cc) BY

Korupsi dan Wilayah Birokrasi Bersih dan Melayani di Instansi Pemerintah. https://peraturan.bpk.go.id/Details/202223/permen-pan-rb-no-90-tahun-2021

Kementerian Pendayagunaan Aparatur Negara dan Reformasi Birokrasi. (2024). Peraturan Menteri Pendayagunaan Aparatur Negara dan Reformasi Birokrasi Nomor 5 Tahun 2024 tentang Perubahan atas Peraturan Menteri Pendayagunaan Aparatur Negara dan Reformasi Birokrasi Nomor 90 Tahun 2021 tentang Pembangunan dan Evaluasi Zona Integritas Menuju Wilayah Bebas dari Korupsi dan Wilayah Birokrasi Bersih dan Melayani di Instansi Pemerintah. https://peraturan.go.id/id/permenpanrb-no-5-tahun-2024

Kementerian Perhubungan. (2025). Peraturan Menteri Perhubungan Republik Indonesia Nomor PM 4 Tahun 2025 tentang Organisasi dan Tata Kerja Kementerian Perhubungan. https://peraturan.go.id/id/permenhub-no-4-tahun-2025

Copyright© Authors