

Psychometric Quality Evaluation of the U-GRT Figural Subtest Using the Rasch Model among Job Selection Participants

Evaluasi Kualitas Psikometrik Subtes Figural U-GRT Menggunakan Model Rasch pada Peserta Seleksi Kerja

Ibnu Firsty¹

University of North Sumatra

Email:

ibnufirsty@usu.ac.id

Dina Nazriani²

University of North Sumatra

Email:

nazriani.dina@usu.ac.id

Fahmi Ananda³

University of North Sumatra

Email:

fahmi.ananda@usu.ac.id

Rahma Fauzia⁴

University of North Sumatra

Email:

rahma.fauzia@usu.ac.id

Correspondence:

Ibnu Firsty

University of North Sumatra

Email: ibnufirsty@usu.ac.id

Abstract

This study aimed to evaluate the psychometric quality of the Figural subtest of the USU General Reasoning Test (U-GRT) using the Rasch Model to determine its suitability for employee selection. A descriptive quantitative approach was employed using secondary data obtained from 728 job applicants who completed a Computer-Based Test (CBT). Data were analyzed using the Rasch Model to examine item difficulty, item fit statistics (infit and outfit mean square), unidimensionality through Principal Component Analysis of Residuals (PCAR), person reliability, and score categorization based on person ability estimates transformed from logit measures into T-scores. The findings indicated that item difficulty ranged from -1.645 to 4.001 logits, with most items clustered at moderate difficulty levels. Nineteen of the twenty items met the Rasch fit criteria, while one item displayed an outfit statistic exceeding the recommended threshold and therefore requires further revision. The PCAR results supported the assumption of unidimensionality, whereas person reliability was classified as acceptable. Ability estimates were subsequently converted into five interpretive categories: very low, low, moderate, high, and very high. Overall, the Figural subtest of the U-GRT demonstrated adequate psychometric properties, although additional items representing extreme difficulty levels are recommended to improve measurement precision across the full range of ability.

Keywords : Rasch Model; Figural Reasoning; Psychometrics; Personnel Selection

Abstrak

Penelitian ini bertujuan mengevaluasi kualitas psikometrik subtes Figural pada USU General Reasoning Test (U-GRT) menggunakan Model Rasch sebagai dasar penilaian kelayakan instrumen dalam seleksi kerja. Penelitian menggunakan pendekatan kuantitatif deskriptif dengan data sekunder yang berasal dari 728 peserta seleksi kerja berbasis Computer-Based Test (CBT). Analisis dilakukan melalui Model Rasch untuk mengevaluasi tingkat kesukaran aitem, kesesuaian aitem terhadap model (infit dan outfit mean square), unidimensionalitas menggunakan Principal Component Analysis of Residuals (PCAR), reliabilitas person, serta penyusunan kategorisasi skor berdasarkan estimasi kemampuan dalam satuan logit yang ditransformasikan menjadi T-score. Hasil penelitian menunjukkan bahwa tingkat kesukaran aitem berada pada rentang -1,645 hingga 4,001 logit dengan sebagian besar aitem berada pada kategori kesukaran sedang. Sebanyak 19 dari 20 aitem memenuhi kriteria kesesuaian model Rasch, sedangkan satu aitem menunjukkan nilai outfit di atas batas yang direkomendasikan sehingga memerlukan evaluasi lebih lanjut. Analisis PCAR mendukung asumsi unidimensionalitas, sementara nilai person reliability berada pada kategori cukup. Kategorisasi skor menghasilkan lima tingkat kemampuan, yaitu sangat rendah, rendah, sedang, tinggi, dan sangat tinggi. Temuan ini menunjukkan bahwa subtes Figural U-GRT memiliki kualitas psikometrik yang memadai, meskipun pengembangan aitem pada tingkat kesukaran ekstrem masih diperlukan untuk meningkatkan presisi pengukuran.

Kata Kunci : Model Rasch; Penalaran Figural; Psikometri; Seleksi Kerja

Copyright (c) 2026 Ibnu Firsty et all

Received 02/07/2026

Revised 23/08/2026

Accepted 11/10/2026



INTRODUCTION

Psychological tests are among the most widely used instruments in the job selection process because they are capable of providing objective information regarding individual characteristics that cannot be directly observed. The use of evidence-based psychological instruments has become increasingly important as organizations face growing demands for accurate, fair, and accountable recruitment processes in obtaining human resources that are suited to job requirements (Kenku et al., 2026). Among various assessment approaches, the measurement of cognitive ability is still regarded as one of the most consistent predictors of job performance because it is associated with an individual's ability to understand new information, solve problems, and adapt to changes in the work environment (Suwarno, 2025). Psychological tests with good psychometric quality not only produce more accurate ability estimates but also reduce subjectivity in the selection decision-making process (Mönke et al., 2024). This condition places the evaluation of instrument quality as an inseparable part of modern psychological assessment practices. Selection decisions resulting from instruments that have not met psychometric standards have the potential to reduce recruitment effectiveness and increase the risk of mismatch between individual characteristics and job demands (Wahlkrantz & Melin, 2025).

General cognitive ability has long been recognized as a major predictor of job performance across various types of occupations and levels of task complexity (Vierula et al., 2021). A meta-analysis conducted by Schmidt and Hunter Roca (2021) reported that cognitive ability had a predictive validity of 0.51, which was higher than various other selection methods, such as unstructured interviews and work experience. These findings are reinforced by more recent research indicating that reasoning ability remains one of the best indicators for predicting individual success when facing dynamic and complex work (Landers et al., 2023). Based on cognitive ability theory, fluid reasoning is a major component that reflects an individual's capacity to recognize patterns, process new information, and solve problems without relying on previous experience (Gul & Ghazal, 2022). Akhtar (2025) explained that this ability is closely related to general intelligence and is therefore often used as a representation of intellectual ability in various selection instruments. This empirical evidence demonstrates that the measurement of reasoning ability continues to play a highly important role in supporting evidence-based job selection processes.

Figural reasoning is one of the important components of fluid reasoning that focuses on the ability to recognize visual patterns, understand spatial relationships, and draw logical conclusions through non-verbal stimuli (Gul & Ghazal, 2022). Unlike verbal and numerical tests, figural reasoning places greater emphasis on abstract thinking processes through the manipulation of visual information, thereby minimizing the influence of language, educational background, and culture (Sackett et al., 2022). These characteristics make figural tests more culture-fair and relevant for use in job selection processes involving

participants from diverse backgrounds (Akhtar, 2025). Noperi (2025) found that figural instruments had very good internal reliability and were strongly correlated with general intelligence ability, thereby providing an accurate representation of an individual's cognitive potential. Hamzah (2024) who positioned reasoning ability as one of the strongest predictors of job performance. This condition indicates that the psychometric quality of figural instruments is an important factor in producing accurate measurements and supporting decision-making in the employee selection process.

Although the use of cognitive ability tests continues to increase, the quality of instrument measurement faces increasingly complex challenges due to the development of digital technology and the increasing leakage of psychological test materials (Bailey, 2017). The dissemination of assessment materials through social media, online forums, and marketplaces has made it increasingly difficult to maintain instrument confidentiality, resulting in the potential for participants' scores to no longer reflect their actual abilities (Blum et al., 2016). Indonesia has more than 214 million internet users, or approximately 79% of the total population, a condition that accelerates the unauthorized distribution of test materials (Manikeswari & Kumar, 2024). Akhtar (2025) also stated that the use of cognitive ability instruments in Indonesia is still dominated by tests that have been used for a long time, such as Raven's Progressive Matrices and Culture Fair Intelligence Test (CFIT), thereby increasing the risk of declining measurement quality due to repeated use. This situation indicates that the development of new instruments needs to be accompanied by comprehensive psychometric evaluation to ensure that measurement quality is maintained when applied in the job selection process. This need is not only related to the development of new items but also to ensuring that each item is capable of consistently measuring the intended construct and is free from deviations in measurement characteristics (Preti et al., 2017).

In response to the need for cognitive ability instruments that are more suitable for the characteristics of Indonesian society, Universitas Sumatera Utara, through the Center for Community Psychological Services, developed the USU General Reasoning Test (U-GRT) as a measure of reasoning ability that can be used in various assessment contexts, including job selection. One of the components of the instrument is a figural subtest designed to measure the ability to recognize patterns, understand visual relationships, and perform abstract reasoning through non-verbal stimuli. The development of local instruments such as the U-GRT represents an important step toward reducing dependence on imported instruments that have long been used in psychological assessment practices in Indonesia (Akhtar, 2025). Nevertheless, the existence of an instrument does not automatically guarantee its measurement quality because the psychometric characteristics of each item need to be demonstrated through empirical testing before the instrument is widely used (Dijksma et al., 2023). Such evaluation is necessary to ensure that each item is capable of consistently measuring figural reasoning ability in the target

population. Adequate psychometric evidence is also a prerequisite for ensuring that assessment results can be used as a basis for decision-making that can be scientifically and professionally justified (Motta et al., 2022).

Several studies have applied the Rasch Model to evaluate the quality of cognitive ability instruments because this model is capable of estimating item difficulty and individual ability on the same measurement scale (Hobson, 2015; Lloyd, 2026; Tabane, 2017). This approach enables the identification of items that function according to the model, the testing of construct unidimensionality, and the evaluation of measurement reliability that cannot be adequately explained through Classical Test Theory. Several previous studies have reported that the application of the Rasch Model can improve instrument quality by identifying items that need to be retained, revised, or eliminated before being used in assessment practices (Suciati et al., 2026). However, psychometric studies of the U-GRT figural subtest among job selection participants remain very limited, particularly studies that integrate analyses of item difficulty, model fit, unidimensionality, measurement reliability, and score categorization within a single analytical framework. This limitation indicates that an empirical gap remains regarding the psychometric characteristics of instruments developed for the context of job selection in Indonesia. Addressing this gap is important so that the use of the U-GRT is supported by adequate empirical evidence in accordance with developments in modern psychological measurement.

The quality of a psychological instrument is determined not only by the clarity of the construct being measured but also by the ability of each item to accurately represent that construct. Denovan (2024) explained that good items should be capable of distinguishing individuals based on their level of ability and providing consistent measurement information. The Rasch approach provides a measurement framework that enables a more comprehensive evaluation of item characteristics through the estimation of item difficulty, item fit to the model, measurement reliability, and the dimensional structure of the instrument (Hapsan et al., 2026). This information serves as a basis for determining whether an instrument has met the psychometric characteristics required before being used in assessment practices. Instruments with good psychometric characteristics are expected to produce more objective ability estimates and reduce the possibility of errors in interpreting participants' scores. On this basis, psychometric evaluation constitutes an inseparable part of the development process for cognitive ability instruments.

Adequate psychometric characteristics also play an important role in producing meaningful score interpretations in the job selection process. Ability estimates obtained through the Rasch Model allow score categorization based on logit measures, which can subsequently be transformed into standard scores so that participants' abilities can be interpreted more objectively and consistently (Holzman, 2018). This approach not only provides information regarding an individual's ability position relative to item difficulty but also supports more accurate decision-making than the use of

raw scores alone (Wilson, 2023). Organizations require instruments that can accurately distinguish participants' abilities because assessment results serve as one of the bases for determining candidate suitability in the recruitment process (Mühling et al., 2015). If the psychometric characteristics of an instrument are inadequate, the resulting score interpretations may potentially lead to inaccurate selection decisions. This relationship indicates that evaluating the psychometric quality of the figural subtest is important not only from the perspective of instrument development but is also directly related to the quality of assessment practices in the job selection process.

This study was initiated from the need to obtain empirical evidence regarding the psychometric quality of the figural reasoning ability subtest of the U-GRT in the context of its use in the job selection process. The evaluation was conducted using the Rasch Model by considering several measurement characteristics, including item fit, unidimensionality, item difficulty, measurement reliability, and score categorization based on ability estimates in logit units. This approach enables evaluation not only based on instrument reliability but also by considering the characteristics of each item and the relationship between item difficulty and participants' ability. The evaluation results are expected to provide empirical evidence regarding the measurement quality of the U-GRT Figural subtest while also identifying items or measurement aspects that still require improvement. The findings of this study may subsequently serve as a basis for the development of the U-GRT Figural subtest and provide information for subsequent psychometric research employing the Rasch Model in the evaluation of cognitive ability instruments.

Based on the description above, this study aims to evaluate the psychometric quality of the U-GRT Figural subtest using the Rasch Model among job selection participants through analyses of item difficulty, item fit to the model, unidimensionality, measurement reliability, and score categorization. The results are expected to provide a scientific basis for the development and refinement of the U-GRT Figural subtest as an instrument for measuring figural reasoning ability in the job selection process. This study is also expected to enrich psychometric research, particularly regarding the application of the Rasch Model in the evaluation of cognitive ability instruments developed in Indonesia. In line with this objective, it is stated that the U-GRT Figural subtest has adequate psychometric quality based on the Rasch Model indicators, as demonstrated by the fulfillment of the unidimensionality assumption, the dominance of items that fit the measurement model, adequate measurement reliability, and an item difficulty distribution capable of representing variations in participants' abilities.

RESEARCH METHODS

This study employed a quantitative approach with a psychometric descriptive design to evaluate the quality of the figural subtest of the USU General Reasoning Test (U-GRT) using the Rasch Model. This psychometric study did not

involve independent or dependent variables but focused on evaluating the construct of figural reasoning as part of fluid intelligence, which encompasses the ability to recognize visual patterns, discriminate forms, understand spatial relationships, and solve problems through inductive and deductive reasoning (Stenulson, 2025). The evaluation was conducted on the psychometric characteristics of the instrument, including construct validity, reliability, item difficulty, item fit to the model, and the development of result categorization based on participants' ability estimates. The Rasch Model approach was selected because it is capable of producing estimates of individual ability and item difficulty on an objective and invariant interval scale, thereby providing more accurate interpretations than the Classical Test Theory approach (Castillo & Roca, 2021)

The research subjects consisted of 728 job selection participants who underwent psychological assessment using the U-GRT through the Center for Community Psychological Services (P3M) of Universitas Sumatera Utara and partner institutions. The sample was selected using a non-probability sampling technique with an accidental sampling (convenience sampling) approach based on the availability of assessment data that met the research requirements (Sugiyono & Lestari, 2021). This number exceeded the minimum recommendation for Rasch analysis, which is 200 respondents, and was therefore considered adequate to produce stable parameter estimates (Thakur, 2025). The research data consisted of secondary data obtained from the Computer-Based Test (CBT) administered by P3M of Universitas Sumatera Utara. The instrument analyzed was the U-GRT Figural Subtest, consisting of 20 items across four subtests, namely Figural Classifications (IF1), Figural Matrices (IF2), Deductive Figural (DF1), and Spatial Relations (DF2), with a dichotomous scoring system consisting of a score of 1 for correct answers and 0 for incorrect answers in accordance with the characteristics of the dichotomous

Rasch Model (Bailey, 2017). Considering that the data originated from a sensitive job selection process, the data used in this study were limited to information relevant to the psychometric analysis, and the research results were presented in aggregate without disclosing the individual identities of the participants.

Data analysis was conducted using Jamovi version 2.6.44 through descriptive analysis and the Rasch Model. Descriptive analysis was used to describe the characteristics of the respondents' score distribution, while Rasch analysis was used to evaluate item fit based on Infit and Outfit Mean Square (MNSQ) values with a criterion of $0.5 \leq \text{MNSQ} \leq 1.5$, test the assumption of unidimensionality through Principal Component Analysis of Residuals (PCAR), and estimate person reliability, person separation index, and item difficulty, which were visualized using a Wright Map (Wahlkrantz & Melin, 2025). Participants' ability estimates were expressed in logit units, then transformed into Z-scores using the formula $Z = (X - \mu)/\sigma$ and subsequently converted into T-scores using the formula $T = 50 + (10 \times Z)$ to establish ability categorization based on the distribution of standard scores (Castillo & Roca, 2021; Vierula et al., 2021). This approach enabled the interpretation of measurement results to be conducted more objectively and supported the evaluation of the psychometric quality of the U-GRT figural subtest as a job selection instrument (Stark et al., 2017)

RESULTS

In the Rasch Model approach, reliability not only reflects the internal consistency of the instrument but also describes the ability of the test to distinguish individual ability levels through the person reliability index (Hapsan et al., 2026; Holzman, 2018). Reliability evaluation in this study was conducted on the U-GRT Figural subtest using response data from 728 job selection participants..

Table 1. Reliability Analysis Results of the U-GRT Figural Subtest Based on the Rasch Model

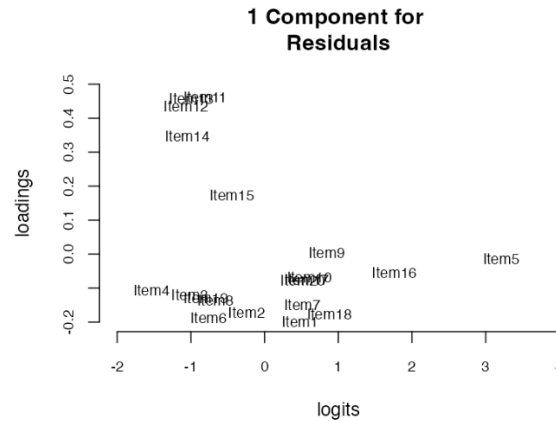
Parameter	Value	Interpretation
Squared Standard Deviation (SSD)	1,180	Respondent ability variation
Mean Squared Error (MSE)	0,306	Average measurement error
Person Reliability	0,742	Fair

Source: Results of the Rasch Model analysis using Jamovi version 2.6.44.

The analysis results showed that the U-GRT Figural subtest had a person reliability value of 0.742, which falls into the fair category. This value indicates that the instrument has adequate measurement consistency in distinguishing figural reasoning ability among job selection participants. The Squared Standard Deviation (SSD) value of 1.180 indicates variation in participants' abilities, while the Mean Squared Error (MSE) value of 0.306 indicates the magnitude of the average measurement error. A comparison of these two values shows that the variation in participants' abilities was greater than the measurement error. The U-GRT Figural subtest demonstrated a fair level of reliability in measuring figural reasoning ability in the sample of this study.

Furthermore, the construct validity of the U-GRT Figural subtest was evaluated through the Rasch Model approach by examining the internal structure of the instrument. The analysis was conducted to ensure that each item was capable of consistently representing the construct of figural reasoning ability. Evidence of validity in this study was obtained through testing unidimensionality, analyzing item fit to the model, and examining the distribution of item difficulty using the Wright Map. These three indicators provide an overview of the quality of the instrument's internal structure and serve as a basis for evaluating the suitability of the U-GRT Figural subtest as a measure of figural reasoning ability.

[7]



Picture 1.
Principal Component Analysis of Residuals (PCAR)

The results of the Principal Component Analysis of Residuals (PCAR) were used to evaluate the unidimensionality assumption of the U-GRT Figural subtest. The PCAR results showed a residual structure used to evaluate the presence of other dimensions beyond the primary construct being measured. Based on the analysis results, the U-GRT Figural subtest was found to meet the assumption of unidimensionality, allowing all items to be considered as representing the construct of figural reasoning ability within a single measurement dimension. The Rasch

Model analysis results showed that most items of the U-GRT Figural Subtest met the model fit criteria, with Infit and Outfit MNSQ values ranging from 0.5–1.5. Of the 20 items analyzed, 19 items (95.0%) met the fit criteria, while one item, namely Item 5, showed an Outfit MNSQ value of 3.437. Statistics on item difficulty and fit for all items are presented in Table 2.

Table 2. Item Difficulty and Fit Statistics of the U-GRT Figural Subtest

Item	Difficulty Level (Logit)	Infit MNSQ	Outfit MNSQ
Item 1	0,502	1,085	1,114
Item 2	-0,310	0,939	0,915
Item 3	-1,128	0,976	0,956
Item 4	-1,645	0,970	0,829
Item 5	4,001	1,051	3,437
Item 6	-0,858	0,965	0,964
Item 7	0,554	1,143	1,191
Item 8	-0,757	0,932	0,919
Item 9	0,952	1,099	1,369
Item 10	0,614	1,006	1,052
Item 11	-0,917	0,848	0,777
Item 12	-1,198	0,913	0,818
Item 13	-1,105	0,864	0,766
Item 14	-1,167	0,902	0,864
Item 15	-0,545	0,944	0,900
Item 16	1,734	0,900	0,942
Item 17	0,667	1,229	1,388
Item 18	1,002	1,121	1,429
Item 19	-0,902	1,021	0,989
Item 20	0,509	0,983	1,020

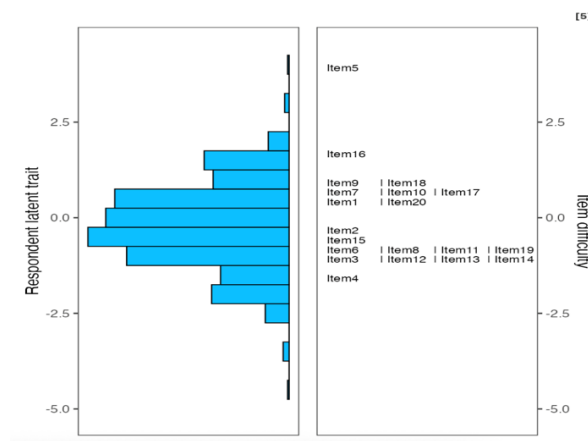
Source: Results of the Rasch Model analysis using Jamovi version 2.6.4

Based on Table 2, the Infit MNSQ values for all items ranged from 0.848–1.229, while the Outfit MNSQ values ranged from 0.766–1.429 for the 19 items that met the fit criteria. In general, these values indicate that participants' responses to most items were consistent with the expectations of the Rasch Model. In terms of difficulty level,

the item values ranged from -1.645 to 4.001 logits. Item 4 had the lowest difficulty level (-1.645 logits), while Item 5 had the highest difficulty level (4.001 logits). Most items were within the range of -1 to +1 logits, indicating that the distribution of difficulty levels tended to be concentrated at a moderate level.

Item 5 exhibited characteristics that require attention. This item had the highest difficulty level, namely 4.001 logits, with an Infit MNSQ value of 1.051, which remained within the range of 0.5–1.5. However, the Outfit MNSQ value of 3.437 indicated a considerable deviation in responses compared with the model expectations. This condition indicates that although participants' responses were generally consistent with the pattern measured by Infit, there were certain responses that were unexpected and contributed to the high Outfit value.

Therefore, Item 5 requires further evaluation at the instrument development stage, particularly in terms of stimulus characteristics, clarity of instructions, and the possibility of alternative interpretations of the stimulus or answer choices. Considering the relatively large misfit value, the item would be more appropriately revised and retested before being used in subsequent administrations rather than being retained directly without evaluation



Picture 2.
Wright Map of the U-GRT Figural Subtest

The Wright Map visualization shows that item difficulty ranged from -1.645 to 4.001 logits, with the majority of items concentrated within the interval of -1 to $+1$ logits. This distribution indicates that the U-GRT Figural subtest consists predominantly of items with moderate difficulty levels, meaning that measurement information is primarily obtained from respondents whose abilities fall within this range. On the other hand, the number of items with very low and very high difficulty levels is relatively limited, resulting in restricted measurement coverage for groups with extreme ability levels. Overall, the item distribution indicates that the measurement structure covers most of the respondents' ability range, although the distribution of item difficulty levels is not yet fully even.

Furthermore, as a practical outcome of the psychometric evaluation, this study developed score categorization for figural reasoning ability based on individual ability estimates (person ability) obtained through the Rasch Model. Ability estimates in logit units were then transformed into T-scores so that the measurement results could be more easily interpreted and used in assessment and job selection processes (Landers et al., 2023). The transformation into standard scores enables ability interpretation to be conducted more consistently because each category represents the individual's relative position within the reference group (Kenku et al., 2026; Mönke et al., 2024; Suwarno, 2025).

Table 3. Score Categorization of Figural Reasoning Ability of the U-GRT Figural Subtest

Kategori	Rentang T-score	Interpretasi
Very Low	< 35	Figural reasoning ability is far below average
Low	35–44	Figural reasoning ability is below average
Moderate	45–55	Figural reasoning ability is at the average level
High	56–65	Figural reasoning ability is above average
Very High	> 65	Figural reasoning ability is far above average

Source: Results of logit score transformation based on the Rasch Model analysis

The results showed that respondents' ability estimates could be interpreted into five categories, namely very low, low, moderate, high, and very high. This grouping provides a more systematic reference for interpreting figural reasoning ability, so that the scores obtained not only indicate the respondents' numerical positions but also their relative ability levels based on the distribution of abilities in

the research sample. This approach facilitates the presentation of assessment results to instrument users without eliminating the characteristics of Rasch Model-based measurement, which uses an interval scale. The score categorization produced in this study can be used as an initial basis for interpreting figural reasoning ability among job selection participants with characteristics similar to those of

the research sample (Gul & Ghazal, 2022). Nevertheless, the categories developed remain empirical based on the distribution of respondents' abilities in this study, so their application to broader populations requires further testing using more representative samples.

DISCUSSION

The psychometric quality of the Figural subtest of the USU General Reasoning Test (U-GRT) was evaluated using the Rasch Model as a modern measurement approach that enables the objective estimation of item parameters and respondents' abilities. This evaluation is important because psychological instruments used in the job selection process are required to have adequate evidence of validity and reliability so that the resulting decisions can be scientifically justified (Sackett et al., 2022). The results showed that the U-GRT Figural subtest had generally adequate psychometric characteristics, as indicated by the fulfillment of the unidimensionality assumption, the dominance of items that fit the model, and measurement reliability in the fair category. However, these results need to be interpreted cautiously because there was still one item that showed misfit, with an Outfit MNSQ value of 3.437, and person reliability of 0.742, which was in the fair category. These findings indicate that most items were able to consistently represent the construct of figural reasoning ability. This study also produced score categorization based on logit estimates transformed into T-scores, thereby providing a more practical basis for interpretation for instrument users.

Evidence of construct validity in this study was obtained through an analysis of the internal structure using the Rasch Model, which showed that the U-GRT Figural subtest met the assumption of unidimensionality. Ellingwood (2020), Lievens (2021), dan Woods (2020) found that this condition indicates that all items predominantly measure the same construct, namely figural reasoning ability, allowing score interpretation to be conducted within a consistent single ability dimension. These results were reinforced by the item fit analysis, which showed that 19 of the 20 items met the item fit criteria, while only one item exhibited a response pattern outside the expectations of the model. The dominance of items that fit the Rasch Model indicates that the internal structure of the instrument has functioned consistently in distinguishing the abilities of job selection participants. This finding is consistent with the view of Vella (2020) who stated that evidence of internal structure is one of the important components in establishing the construct validity of a psychological instrument. Nevertheless, construct validity is still regarded as a cumulative concept and therefore requires support from other sources of evidence, such as relationships with external variables and predictive validity (Preuß, 2025).

The distribution of item difficulty showed a relatively wide range of logit parameters, from -1.645 to 4.001 logits, with the majority of items falling within the interval of -1 to +1 logits. This distribution indicates that the U-GRT Figural subtest provides more measurement information for participants with moderate ability than for participants with

very low or very high ability. This condition was also evident in the Wright Map, which showed that the distribution of item difficulty relatively overlapped with the concentration of the abilities of the majority of respondents. Although there was one item with a very high difficulty level and an outfit value above the recommended threshold, its presence did not alter the general tendency that the instrument had sufficiently good targeting for the characteristics of the research sample. According to Breuer (2023) and Vo (2020), the correspondence between the distribution of respondents' abilities and item difficulty is one of the important indicators for producing stable ability estimates. These findings indicate that subsequent instrument development should be directed toward adding items with extreme difficulty levels so that measurement information is distributed more evenly across the ability continuum (Susongko et al., 2021).

The reliability analysis showed a person reliability value of 0.742, which was in the fair category. This value indicates that the U-GRT Figural subtest has been able to distinguish figural reasoning ability among participants relatively consistently, although measurement precision has not yet been optimal across the entire ability range. This condition is related to the distribution of item difficulty, which is still dominated by the moderate category, resulting in measurement information being more concentrated among the moderate-ability group. In the Rasch approach, reliability is influenced by the correspondence between item difficulty and respondents' abilities; therefore, an uneven distribution of items will affect the instrument's ability to distinguish individuals at extreme ability levels (Sinaga, 2024). The findings of this study support the views of Marastuti (2020) and Sukanto (2026) that reliability values in the fair category can still be accepted for research purposes and group assessment, particularly when the instrument is used as part of a selection process. This condition also indicates that improving instrument quality does not depend solely on the number of items but also on the variation in difficulty levels capable of representing the entire spectrum of respondents' abilities.

The use of the Rasch Model in this study provided more detailed information regarding the psychometric characteristics of the U-GRT Figural subtest compared with an approach based on total scores in Classical Test Theory. In addition to producing estimates of the difficulty level of each item, this approach enables participants' abilities to be estimated in logit units, so that the interpretation of measurement results no longer depends solely on raw scores (Tarigan & Fadillah, 2021). These estimates were subsequently transformed into T-scores to establish ability categories that are easier to use in interpreting assessment results. The development of logit-based categories provides a more consistent basis for interpretation because it is based on ability estimates calibrated on an interval scale (Rohmah & Widhiarso, 2019). This type of approach remains relatively limited in psychological instrument development research in Indonesia and can therefore expand the application of the Rasch Model in developing interpretive norms for cognitive ability instruments. These results also demonstrate that

psychometric evaluation does not stop at determining item quality but can be extended to produce a more objective score interpretation system for instrument users.

The results of the psychometric evaluation also provide several considerations that can be used as a basis for improving the U-GRT Figural subtest. The presence of one item that did not meet the model fit criteria indicates that a review of the stimulus, visual patterns, and item complexity is still necessary so that the measurement characteristics become more stable. The distribution of difficulty levels, which is still dominated by the moderate category, also indicates the need to add items with more varied difficulty levels so that the instrument can provide more accurate measurement information for participants with very low and very high abilities (S. Y. Lee & Hong, 2021). The development of an item bank through an automatic item generation approach may serve as an alternative for producing parallel items with comparable psychometric characteristics while also reducing the risk of psychological test material leakage (S. C. Lee et al., 2023). This step is relevant to the needs of job selection assessment, which requires instruments to have good measurement quality while also being capable of being used continuously across various test administrations. The evaluation process conducted in this study is expected to become part of the continuous development of the U-GRT so that the quality of measurement of figural reasoning ability can be further optimized.

CONCLUSION

The evaluation of the psychometric quality of the U-GRT Figural subtest using the Rasch Model showed that the instrument had generally adequate measurement characteristics for measuring figural reasoning ability among job selection participants. This was indicated by the fulfillment of the unidimensionality assumption, the dominance of items that fit the model, and reliability in the fair category. Nevertheless, the findings should be interpreted cautiously because the sample was obtained through convenience sampling, the data were derived from a single CBT platform, and the study did not examine external validity or predictive validity. In addition, one item showed misfit and had not yet been followed up through revision or retesting. Future studies are recommended to use more diverse samples from multiple institutions or assessment platforms, conduct external and predictive validity testing, and revise and retest items that show misfit. Thus, the psychometric evidence for the U-GRT Figural subtest can be strengthened, and the generalizability of the measurement results can be evaluated more broadly.

ACKNOWLEDGMENTS

The authors would like to thank the Center for Community Psychological Services (P3M) of Universitas Sumatera Utara and the partner institutions that provided access to the assessment data used in this study.

DECLARATION OF INTEREST

The authors declare that there is no conflict of interest related to this research or the publication of this article.

TRANSPARENCY OF DATA

The data used in this study were secondary data obtained from the psychological assessment process organized by the Center for Community Psychological Services (P3M) of Universitas Sumatera Utara. Because the data originated from a job selection process and are sensitive in nature, individual-level data are not publicly available. The use of and access to the data followed the provisions and permissions granted by the data provider

FUNDING

This study did not receive specific funding from government agencies, commercial institutions, or nonprofit organizations..

AI USAGE STATEMENT

The authors declare that no artificial intelligence (AI) tools were used in the research, data analysis, or preparation of this manuscript.

AUTHORS' CONTRIBUTIONS

Author 1: conceptualization, methodology, data analysis, investigation, and preparation of the initial manuscript.
 Author 2: supervision, validation, and review and editing of the manuscript.
 All authors: reviewed and approved the final version of the manuscript.

REFERENCES

- Akhtar, H. (2025). Measuring Process Factors of Fluid Reasoning Using Multidimensional Computerized Adaptive Testing. *Assessment*, 32(1), 32–47. <https://doi.org/10.1177/10731911241236351>
- Bailey, R. (2017). HR applications of psychometrics. *Psychometric Testing: Critical Perspectives*. <https://doi.org/10.1002/9781119183020.ch7>
- Blum, D., Holling, H., Galibert, M. S., & Forthmann, B. (2016). Task difficulty prediction of figural analogies. *Intelligence*. <https://www.sciencedirect.com/science/article/pii/S0160289616300927>
- Breuer, S. (2023). Find the Mistake Psychometric Properties of an Innovative Response Format for Figural Matrices Tasks. *Journal of Individual Differences*, 44(2), 83–96. <https://doi.org/10.1027/1614-0001/a000386>
- Castillo, C. C., & Roca, G. J. M. (2021). Effectiveness of psychometric tests for the selection of personnel in jobs in the retail sector. pdfs.semanticscholar.org/6de8/d6477bbf84f13c401e27e20329fd2400f20a.pdf
- Denovan, A., Dagnall, N., Drinkwater, K., & Escolà-Gascón, Á. (2024). The Inventory of Personality Organization-

- Reality Testing Subscale and Belief in Science Scale: Confirmatory factor and Rasch analysis of thinking style measures. In *Plos one*. journals.plos.org. <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0310055>
- Dijkma, I., Stuiver, M., & Lucas, C. (2023). Evaluating the psychometric properties of the Grit scale in Marine recruits using Rasch analysis. *BMJ Mil Health*. <https://militaryhealth.bmj.com/content/169/5/425.abstr>
- Ellingwood, H., Williams, K. M., Sitarenios, G., & ... (2020). Psychometric properties of a contextualized, actuarially informed assessment for law enforcement personnel selection: the M-PULSE Inventory. *Journal of Police and Criminal Psychology*. <https://doi.org/10.1007/s11896-018-9290-0>
- Gul, S., & Ghazal, S. (2022). Need of psychometrics for recruitment and selection in organizations: A qualitative perspective from a developing country. *Journal of Professional & Applied Psychology*. <https://iprp.com/ojs/index.php/jpap/article/view/96>
- Hamzah, H., Govindasamy, P., & Sudhir, J. J. (2024). Psychometric evaluation of the Cognitive Ability Assessment using Rasch analysis and exploratory factor analysis. In *Journal of Cognitive Sciences and Human Development*. publisher.unimas.my. <https://publisher.unimas.my/ojs/index.php/JCSHD/article/view/10611>
- Hapsan, A., Mulyatiningsih, E., Hidayati, K., & ... (2026). Assessing Candidates for the Indonesian Civil Service: An Analysis of Item Characteristics in the Academic Potential Test Using the Item Response Theory Approach. In *Вопросы государственного и муниципального управления*. cyberleninka.ru. <https://cyberleninka.ru/article/n/assessing-candidates-for-the-indonesian-civil-service-an-analysis-of-item-characteristics-in-the-academic-potential-test-using-the>
- Hobson, E. G. (2015). *Using the Rasch model in a computer adaptive testing application to enhance the measurement quality of emotional intelligence*. search.proquest.com. <https://search.proquest.com/openview/45d88f9e4b964634fb6b19ec99fdf4aa/1?pq-origsite=gscholar&cbl=2026366&diss=y>
- Holzman, M. A. (2018). *Evaluating rater effects in the context of ethical reasoning essay assessment: An application of the many-facets rasch measurement model*. search.proquest.com. <https://search.proquest.com/openview/7febfc134396f46f7e150bb0276dfado/1?pq-origsite=gscholar&cbl=18750&diss=y>
- Kenku, A. A., Yahaya, I., & Uzoigwe, T. L. (2026). Theoretical Perspectives on Psychometric Assessment as a Strategic Mechanism in Personnel Selection and Recruitment. *NIU Journal of Social Sciences*. <https://www.niujournals.ac.ug/ojs/index.php/niujoss/article/view/2446>
- Landers, R. N., Auer, E. M., Dunk, L., Langer, M., & Tran, K. N. (2023). A simulation of the impacts of machine learning to combine psychometric employee selection system predictors on performance prediction, adverse impact, and number of dropped predictors. *Personnel Psychology*. <https://doi.org/10.1111/peps.12587>
- Lee, S. C., Lee, Y. C., & Chiu, E. C. (2023). Psychometric validation of the Cognitive Abilities Screening Instrument using Rasch analysis in people with dementia. In *Medicine*. journals.lww.com. https://journals.lww.com/md-journal/fulltext/2023/08110/psychometric_validation_of_the_cognitive_abilities.2.aspx
- Lee, S. Y., & Hong, A. J. (2021). Psychometric investigation of the cultural intelligence scale using the Rasch measurement model in South Korea. In *Sustainability*. mdpi.com. <https://www.mdpi.com/2071-1050/13/6/3139>
- Lievens, F., & Sackett, P. R. (2017). The effects of predictor method factors on selection outcomes: A modular approach to personnel selection procedures. *Journal of Applied Psychology*. <https://psycnet.apa.org/record/2016-43607-001>
- Lievens, F., Sackett, P. R., & Zhang, C. (2021). Personnel selection: A longstanding story of impact at the individual, firm, and societal level. *European Journal of Work and Organizational Psychology*. <https://doi.org/10.1080/1359432X.2020.1849386>
- Lloyd, E. D. (2026). *A Comparison of Classical and Rasch Models for Analyzing Job Analysis Data*. search.proquest.com. <https://search.proquest.com/openview/52760e9636f5a67cdfaefd547a36dd8/1?pq-origsite=gscholar&cbl=18750&diss=y>
- Magro, B., & Vella, M. C. (2020). Understanding the Use of Psychometric Testing by Maltese Organisations for the Purpose of Recruitment and Selection. *MCAST Journal of Applied Research & Practice*. <https://journal.mcast.edu.mt/se/article/145691/en>
- Manikeswari, G. D., & Kumar, R. S. P. (2024). Psychometric testing-the game changer: a 360-degree view of psychometric testing and its applicability in HR functions. *International Journal of Business Excellence*. <https://doi.org/10.1504/IJBEX.2024.139112>
- Marastuti, A., Anggoro, W. J., Marvianto, R. D., Afghani, A., & Azzam, A. (2020). Perbandingan properti psikometri antara tes PAPs berbentuk computer-based dan paper and pencil test. In *Jurnal Psikologi UGM*. academia.edu. <https://www.academia.edu/download/71090313/pdf.pdf>
- Mönke, F. W., Roulin, N., & Lievens, F. (2024). Validity of social media assessments in personnel selection: A systematic review of the initial evidence. *European Journal of Psychological Assessment*. <https://psycnet.apa.org/record/2025-48832-001>
- Motta, C. da, Castilho, P., Pato, M. T., & Carvalho, C. B. (2022). Rasch model analysis of the Situational Test of Emotional Understanding–brief in a large Portuguese sample. *Current Psychology*.

- <https://doi.org/10.1007/s12144-020-00862-6>
- Mühling, A., Ruf, A., & Hubwieser, P. (2015). Design and first results of a psychometric test for measuring basic programming abilities. *Proceedings of the Workshop in Primary and Secondary Computing Education*. <https://doi.org/10.1145/2818314.2818320>
- Noperi, H., Diani, R., & Irwandani, I. (2025). Psychometric analysis of the critical thinking skills instrument in physics learning using the Rasch model. *Indonesian Journal of Science and Mathematics Education*. <https://ejournal.radenintan.ac.id/index.php/IJSME/article/view/28770>
- Preti, A., Vellante, M., & Petretto, D. R. (2017). The psychometric properties of the “Reading the Mind in the Eyes” Test: An item response theory (IRT) analysis. *Cognitive Neuropsychiatry*. <https://doi.org/10.1080/13546805.2017.1300091>
- Preuß, J. (2025). Using Cognitive Diagnostic Models to Evaluate the Two-Process Theory of Matrix Reasoning. *Journal of Intelligence*, 13(2). <https://doi.org/10.3390/jintelligence13020022>
- Rohmah, Y. S., & Widhiarso, W. (2019). Pengujian validitas konstruk gadjah mada scholastic test (GMST) melalui analisis faktor eksploratori. In *Jurnal Psikologi Universitas Gajah Mada, December*. researchgate.net. https://www.researchgate.net/profile/Wahyu-Widhiarso/publication/337902024_Pengujian_Validitas_Konstruk_Gadjah_Mada_Scholastic_Test_GMST_Melalui_Analisis_Faktor_Eksploratori/links/5df160c592851c836475e5ed/Pengujian-Validitas-Konstruk-Gadjah-Mada-Scholastic-Test-GMST-Melalui-Analisis-Faktor-Eksploratori.pdf
- Sackett, P. R., Zhang, C., Berry, C. M., & ... (2022). Revisiting meta-analytic estimates of validity in personnel selection: Addressing systematic overcorrection for restriction of range. *Journal of Applied Psychology*. <https://psycnet.apa.org/fulltext/2022-17327-001.html>
- Sinaga, Y. C. (2024). *Evaluasi Psikometri Aitem Induktif Figural pada USU General Reasoning Test (UGRT)*. repositori.usu.ac.id. <https://repositori.usu.ac.id/handle/123456789/97537>
- Stark, S., Chernyshenko, O. S., & Drasgow, F. (2017). Modern psychometric theory to support personnel assessment and selection. In *Handbook of Employee Selection*. <https://doi.org/10.4324/9781315690193-42>
- Stenulson, D. (2025). *Rasch Model Utilization Towards Developing an Instrument for Measuring Creative Drive*. search.proquest.com. <https://search.proquest.com/openview/03ec18ca4ea4ec5384632121666155b0/1?pq-origsite=gscholar&cbl=18750&diss=y>
- Suciati, I., Walanda, D. K., Idris, M., Napitupulu, M., Anggraini, A., Lefrida, R., & Walanda, R. A. (2026). Psychometric Evaluation of Culturally Responsive Numeracy and Problem-Solving Assessment: A Rasch Modeling Approach. *Jurnal Pendidikan MIPA*. <https://jpmipa.fkip.unila.ac.id/index.php/jpmipa/article/view/1160>
- Sugiyono, & Lestari, P. (2021). *Metode Penelitian Komunikasi (Kuantitatif, Kualitatif, dan Cara Mudah Menulis Artikel pada Jurnal Internasional)* (Sunarto (ed.); 1). CV. Alfabeta Bandung.
- Sukamto, A., Bahrani, B., & Nurvayanti, N. (2026). Tes Kemampuan Akademik (TKA) sebagai inovasi evaluasi pendidikan: Sebuah tinjauan naratif. *Journal of Instructional and Development Researches*. <https://journal.iel-education.org/index.php/JIDeR/article/view/695>
- Susongko, P., Ratu, I. M., & Hayati, M. N. (2021). Tes kemampuan berfikir tingkat tinggi untuk mengukur literasi lingkungan hidup bagi siswa sekolah menengah pertama dengan aplikasi model rasch. *PSEJ (Pancasakti Science Education Journal)*. <https://www.sciencejournal.org/index.php/PSEJ/article/view/127>
- Suwarno, S. (2025). Integration of Psychometric Testing in Employee Recruitment and Selection to Enhance Job Fit. *Journal of Management, Economic, and Accounting*. <http://jurnalunived.com/index.php/JMEA/article/view/1149>
- Tabane, L. M. (2017). *Revising the Industrial Test Battery Using the Rasch Model for Measurement*. search.proquest.com. <https://search.proquest.com/openview/3b8e7cd276cc2aba95a46ae16907d821/1?pq-origsite=gscholar&cbl=2026366&diss=y>
- Tarigan, M., & Fadillah, F. (2021). Analisis item response theory Raven's coloured progressive matrices pada sampel anak usia dini. *PSIKODIMENSIA: Kajian Ilmiah Psikologi*. <https://journal.unika.ac.id/index.php/psi/article/view/3101>
- Thakur, T. (2025). *Rasch measurement theory: A complete course*. books.google.com. <https://books.google.com/books?hl=en&lr=&id=K5lIEQAAQBAJ&oi=fnd&pg=PT25&dq=rasch+model+figural+reasoning+psychometrics+personnel+selection&ots=GIEGlfzEwr&sig=n-AobsoouOO2l5jDk9Z8ubQh5Js>
- Vierula, J., Talman, K., & Hupli, M. (2021). Development and psychometric testing of Reasoning Skills test for nursing student selection: An item response theory approach. *Journal of Advanced Nursing*. <https://doi.org/10.1111/jan.14799>
- Vo, D. Van. (2020). Development of inductive reasoning in students across school grade levels. *Thinking Skills and Creativity*, 37. <https://doi.org/10.1016/j.tsc.2020.100699>
- Wahlkrantz, E., & Melin, J. (2025). Fair Assessment of Logical Ability in Military Selection: A Rasch Analysis of a Paper Fold and Cut Test. *SAGE Open*. <https://doi.org/10.1177/21582440251372266>
- Wilson, M. (2023). *Constructing measures: An item response modeling approach*. taylorfrancis.com. <https://doi.org/10.4324/9781003286929>
- Woods, S. A., Ahmed, S., Nikolaou, I., & Costa, A. C. (2020). Personnel selection in the digital age: A review of validity and applicant reactions, and future research challenges. *European Journal of Work and Organizational Psychology*.