JCSHD

COGNITIVE SCIENCES AND HUMAN DEVELOPMENT

Psychometric evaluation of the Cognitive Ability Assessment using Rasch analysis and exploratory factor analysis

Hazalizah Hamzah*[1], Priyalatha Govindasamy[1], Johnathan Jaya Sudhir[2], Asyraf Wajdi Mohtar[2] & Syara Shazanna Zulkifli[1]

[1]Faculty of Human Development, Universiti Pendidikan Sultan Idris, 35900, Perak, Malaysia.

[2]CXS Analytics Sdn. Bhd., 59200, Kuala Lumpur, Malaysia.

ABSTRACT

Despite the growing emphasis on employability, there is a lack of culturally relevant and psychometrically robust tools to assess the cognitive abilities of Malaysian undergraduates. This study examined the psychometric properties of the Cognitive Ability Assessment (CAA), a 50-item instrument designed to measure employability-related cognitive abilities in this population. A cross-sectional study with 278 students from a public university examined the Cognitive Ability Assessment (CAA) using Rasch modelling and Exploratory Factor Analysis (EFA) to assess item functioning, dimensionality, reliability, and construct validity. Rasch analysis showed acceptable item fit (infit MNSQ = 0.73–1.32) and a broad difficulty range (−4.60 to 5.16 logits), with unidimensionality supported for the Quantitative (46.1%), Fluid (40.0%), and Spatial (60.4%) domains, but Comprehension explained only 26.1% of variance, indicating multidimensionality. Differential item functioning identified 10 items with large DIF and 4 with intermediate DIF across academic programmes. EFA explained 30–48% of the variance after item refinement. Internal consistency was moderate (α = 0.54–0.67), with acceptable model fit for most domains (CFI = 0.94–0.97), except Fluid (CFI = 0.83). The findings suggest that the CAA shows promising measurement precision but needs refinement to achieve structural coherence and subgroup fairness. Future research should confirm its factor structure, test invariance across populations, and assess predictive validity for employment outcomes.

Keywords: cognitive ability assessment, exploratory factor analysis, Malaysian undergraduates, psychometric properties, Rasch analysis

ARTICLE INFO

Email address: hazalizah@fpm.upsi.edu.my (Hazalizah Hamzah)

*Corresponding author

https://doi.org/10.33736/jcshd.10611.2026

e-ISSN: 2550-1623

Manuscript received: 22 August 2025; Accepted: 9 March 2026; Date of publication: 31 March 2026

1 INTRODUCTION

The rapid evolution of the labour market, driven by technological advancements from the First Industrial Revolution to the present era of digital transformation, has significantly reshaped job structures and workforce expectations. Contemporary economies are increasingly characterised by automation, artificial intelligence, and digitally mediated work environments, requiring employees to demonstrate advanced cognitive, analytical, and adaptive competencies (Schwab, 2016). Recent global labour market analyses further indicate that complex problem-solving, critical thinking, and digital literacy are among the most in-demand competencies for emerging occupations (Tee et al., 2024). As technological integration accelerates, certain roles diminish while new categories of employment emerge, redefining employability across global and regional contexts (Frey & Osborne, 2017).

In Malaysia, graduate employability has become a strategic national priority. Government–industry collaborations, including initiatives such as MyTalent, aim to align higher education outcomes with labour market demands as the country transitions toward a knowledge-based economy (Mustapha & Abdullah, 2004). Aligning curricula with industry expectations is widely recognised as essential for workforce integration (Ismail & Hassan, 2019). Recent empirical evidence suggests that employers in Malaysia continue to report gaps in graduates' digital, communication, and higher-order thinking skills, underscoring the need for stronger alignment between educational preparation and workplace expectations (Saleh & Abdul Wahab, 2025). These findings echo broader international scholarship, which emphasises that employability increasingly depends on cognitive, interpersonal, and adaptive competencies cultivated through higher education (Roslan et al., 2024).

To operationalise the assessment of such competencies, employers frequently rely on standardised aptitude and cognitive ability tests. Instruments such as the Watson-Glaser Critical Thinking Appraisal, SHL General Ability Test, Cognitive Criteria Ability Test, and Raven's Progressive Matrices are widely utilised in recruitment settings to evaluate reasoning and problem-solving abilities (Bartram, 2005; Raven et al., 2000). These tools provide efficiency, objectivity, and predictive validity in identifying candidates likely to perform effectively in cognitively demanding roles (Arthur et al., 2006). However, many widely used instruments were developed in Western contexts and may not fully account for Malaysia's linguistic, cultural, and educational diversity.

When assessment content assumes specific linguistic familiarity, contextual references, or educational practices that differ from local experiences, construct irrelevant variance may be introduced (Djiwandono, 2006; Schmitt & Kuljanin, 2008). Such misalignment may disproportionately disadvantage candidates from diverse socioeconomic and linguistic backgrounds, including indigenous communities and those educated in vernacular systems (Cheong et al., 2016). Recent scholarship further highlights that culturally misaligned assessments can undermine fairness and weaken the validity of selection decisions (Ercikan et al., 2023). These concerns underscore the importance of developing locally contextualised instruments that are culturally appropriate while maintaining rigorous psychometric standards (Bartram, 2005; Djiwandono, 2006).

The Cognitive Ability Assessment (CAA) was developed to address the lack of culturally relevant, psychometrically robust tools for Malaysian undergraduates. Designed within the Malaysian educational and linguistic context, the CAA measures key cognitive domains aligned with contemporary employability requirements, reflecting local curricular exposure and cultural nuances. Empirical evidence regarding its reliability, dimensional coherence, construct validity, and fairness across academic backgrounds remains limited. Based on the Cattell–Horn–Carroll theory of cognitive abilities, widely regarded as the most comprehensive and empirically supported model of cognitive structure (Schneider & McGrew, 2018), the CAA comprises 50 items across four domains: Comprehension-Knowledge (13 items), Quantitative Knowledge (13 items), Fluid Knowledge (12 items), and Spatial Knowledge (12 items). Items are categorised by difficulty—easy, intermediate, and hard—to capture a wide spectrum of cognitive abilities, from foundational skills to advanced problem-solving, allowing nuanced evaluation of students' strengths and developmental needs.

The development of the CAA involved a multidisciplinary team of experts, including psychologists specialising in clinical, industrial, and organisational contexts, psychometricians with expertise in item design and validation, and language specialists proficient in both English and Malay. This collaboration ensured that the test was psychometrically sound and culturally and linguistically relevant to Malaysian students. Each item was reviewed and revised multiple times to ensure clarity, appropriateness, and alignment with the cognitive skills expected of undergraduate students. For example, comprehension knowledge items were carefully crafted to assess general knowledge and verbal reasoning within a Malaysian cultural framework. In contrast, Quantitative domain items focused on real-world numerical applications relevant to industry and higher education curricula.

Additionally, the test design adhered to best practices in psychometrics, including item difficulty calibration and discrimination analysis, ensuring that each item contributes meaningfully to the overall assessment. The inclusion of the Fluid Knowledge and Spatial Knowledge domains expands the scope of the CAA, enabling a comprehensive assessment of competencies highly valued in contemporary labour markets, including abstract reasoning, problem-solving in novel contexts, and spatial visualisation, particularly in STEM-related occupations. This comprehensive approach makes it a reliable tool for assessing undergraduate students' cognitive abilities while also addressing the specific needs of employers and educators.

The collaborative input from diverse fields ensured that the CAA aligns with academic expectations and employer requirements, bridging the gap between higher education and the labour market. The involvement of language specialists ensured the test was linguistically accurate and free of cultural biases, making it equally accessible to English- and Malay-speaking participants. This meticulous development process underscores the instrument's commitment to inclusivity, precision, and relevance, enhancing its utility as a pre-employment and educational assessment tool.

The development of a new assessment instrument does not, in itself, ensure measurement quality. Contemporary measurement standards emphasise that instruments intended for educational or organisational decision-making must demonstrate clear empirical evidence of reliability, structural

validity, and fairness before application (Bond & Fox, 2015). At present, no published empirical data have established the psychometric properties of the CAA, including item functioning, dimensionality, or subgroup invariance. The absence of such evidence limits confidence in the interpretation and practical application of scores, representing a critical empirical gap.

To address this gap, the present study undertakes a comprehensive psychometric evaluation of the newly developed instrument using Rasch analysis and exploratory factor analysis. Rasch analysis was conducted to examine item difficulty, item fit, person-ability estimation, dimensionality, and differential item functioning, thereby providing evidence of reliability, measurement precision, and subgroup fairness at the item level. An exploratory factor analysis was conducted to investigate the instrument's underlying latent structure and determine whether the empirical factor configuration aligns with its theoretical framework. By integrating these complementary approaches, the study aims to establish foundational evidence of the CAA's reliability and validity and to identify areas requiring refinement to strengthen its measurement quality and applied defensibility.

2 METHODS

2.1 Design

A cross-sectional survey design was employed to collect and analyse quantitative data, enabling the identification of the internal consistency and factor structure of the Cognitive Ability Assessment. This design enabled efficient data collection at a single time point, making it particularly suitable for assessing how well the instrument's items measured the intended cognitive domains (Creswell & Creswell, 2018). By simultaneously capturing responses from a diverse group of participants, the approach provided a snapshot of performance across the selected sample, facilitating the examination of patterns and relationships within the data (Sedgwick, 2014).

2.2 Participants

The final sample of 278 undergraduate students was considered adequate for both Rasch analysis and exploratory factor analysis. For Rasch modelling, sample sizes exceeding 250 respondents generally produce stable item difficulty estimates within ±0.5 logits, particularly for dichotomous response formats (Bond & Fox, 2015; Linacre, 1994). For exploratory factor analysis, recommended subject-to-item ratios typically range from 5:1 to 10:1 to ensure stable factor extraction and reliable loading estimates (Costello & Osborne, 2005; Tabachnick & Fidell, 2013). Given that the Cognitive Ability Assessment (CAA) comprised 50 items, the present sample met the minimum requirements for both Rasch calibration stability and factor-analytic adequacy and was deemed appropriate for evaluating item-level functioning and the instrument's underlying factor structure.

The sample predominantly comprised female participants, primarily Malay. First-year students made up the largest portion of the cohort, followed by sophomores. Participants were enrolled in both Psychology and non-Psychology degree programmes, with an average age in the early 20s. The average CGPA of 3.71 indicated a high level of academic proficiency. Table 1 provides further details on the participants' demographic composition.

Table 1. Demographic characteristics of the participants.

Characteristic	Category	*n*	%
Gender	Male	50	18.00
	Female	228	82.00
Race	Malay	219	78.80
	Non-Malay	59	21.20
Year	Freshman	186	66.90
	Sophomore	56	20.00
	Junior	20	7.00
	Senior	16	6.00
Programme	Non-psychology	114	41.00
	Psychology	164	59.00
		M	*SD*
Age		21.24	1.58
CGPA		3.71	0.19

2.3 Instrument

The Cognitive Ability Assessment (CAA), a previously developed 50-item instrument, was used to assess cognitive abilities. The CAA comprises four domains: Comprehension-Knowledge (13 items), Quantitative Knowledge (13 items), Fluid Knowledge (12 items), and Spatial Knowledge (12 items). The instrument is designed to capture a broad spectrum of cognitive abilities, from foundational skills to advanced problem-solving. Incomplete or invalid responses were removed before analysis

2.4 Data Analysis

An EFA was conducted to investigate the data's underlying structure and assess the construct validity of the Cognitive Ability Assessment. This technique allowed the identification of latent factors that explain the relationships among the observed items. To accommodate the binary response format of the data, tetrachoric correlations were used, as these are appropriate for dichotomous variables and provide a more accurate representation of the relationships between items (Flora & Curran, 2004). The maximum likelihood extraction method was applied to estimate factor loadings and efficiently model the structure. This method is widely recognised for its robustness and suitability in estimating parameters when normality assumptions are met or closely approximated (Fabrigar et al., 1999). Varimax rotation was employed to enhance the interpretability of the factor structure. This orthogonal rotation method minimises cross-loadings and simplifies the factor matrix, ensuring that items load strongly onto one factor while having minimal loadings on others (Thurstone, 1947). The number of factors to retain was determined using a combination of eigenvalues greater than 1, scree plot analysis, and theoretical considerations. These criteria balanced empirical evidence with conceptual alignment with the CHC framework, guiding the design of the Cognitive Ability Assessment.

Items with factor loadings above 0.30 were considered significant contributors to their respective factors (Tabachnick & Fidell, 2013). This threshold ensures that each item has a meaningful relationship with its factor while excluding weaker associations. Cross-loadings were carefully examined, and items with significant cross-loadings or ambiguous associations were reviewed to maintain factor clarity and alignment with theoretical expectations. The psych package in R (Revelle, 2026) was utilised for the analysis, offering comprehensive tools for tetrachoric correlations, factor extraction, and rotation. Using R ensured precision and reproducibility in the analysis, with tetrachoric correlation critical for rigorous EFA. The results provided strong evidence for the Cognitive Ability Assessment's structural validity, confirming its ability to measure the intended cognitive domains. Additionally, the factor structure aligns well with the CHC model's theoretical underpinnings, further supporting the instrument's robustness.

2.4.1 Rasch Analysis

Rasch analysis, a probabilistic model, was employed to evaluate item difficulty and individuals' abilities on a shared scale, allowing for direct comparison (Tennant & Conaghan, 2007). This method transforms ordinal qualitative data into interval-level measures on a linear logit scale, accounting for variations in item difficulty (Chien et al., 2008). Item difficulty reflects each item's relative challenge within the scale, with negative values indicating more accessible items and positive values indicating more difficult ones. Person estimates, expressed in logits, represent individuals' abilities; positive estimates indicate higher cognitive ability, and negative estimates indicate lower ability.

In this study, Rasch analysis provided evidence on the quality of questionnaire items and the instrument's internal structure. Joint Maximum Likelihood estimation assessed fit statistics, unidimensionality, and reliability for persons and items. Critical aspects of the analysis included item measures, fit statistics, principal component analysis of residuals (PCA-R), and item-person maps to assess unidimensionality and item performance. Each dimension was analysed individually for fit statistics, PCA-R, item-person distribution, and DIF, comprehensively evaluating the instrument's quality. The analysis was conducted using Winstep software (version 5.2.3) to assess the validity and reliability of the CAA by following specific criteria for fit statistics and unidimensionality (Linacre, 2025). Item fit was evaluated using the mean square (MNSQ) values for infit and outfit, with acceptable ranges of 0.5-1.5, as values within this range indicate productive measurement (Bond & Fox, 2015). Items falling outside this range were flagged for potential revision or removal due to misfit. Unidimensionality was evaluated using principal component analysis (PCA) of the residuals. The eigenvalue of the first contrast needed to be below 2.0 to confirm that the test primarily measured a single cognitive construct (Smith, 2002).

Differential item functioning (DIF) was analysed between students from psychology and non-psychology programmes to ensure fairness and validity. The Mantel-Haenszel statistic was used, with negligible DIF indicated by effect sizes below 0.43, moderate DIF between 0.43 and 0.64, and large DIF above 0.64 (Linacre, 2026). This ensured that potential biases across subgroups were addressed, supporting the test's fairness across different educational backgrounds. The person-item map was also reviewed to examine the distribution of item difficulty relative to participants' cognitive abilities, ensuring comprehensive coverage across ability levels.

3 RESULTS

3.1 PCA Residual Analysis

In the first contrast, the unidimensionality of the constructs was evaluated using both raw variance explained and unexplained variance (Table 2). The Quantitative (46.1%), Fluid (40%), and Spatial (60.4%) constructs all exceeded the 40% threshold for raw variance explained, confirming their unidimensionality. Additionally, the unexplained variance for these constructs was within the acceptable range of 1.5 to 2.2 eigenvalue units, further supporting this conclusion. In contrast, the Comprehension construct had a raw variance of only 26.1%, which falls below the unidimensionality threshold, suggesting the presence of multiple latent traits. Although the unexplained variance for this construct was also low (within the acceptable range), the insufficient raw variance explained indicates that it remains multidimensional. Therefore, while the Quantitative, Fluid, and Spatial constructs are confirmed as unidimensional, the Comprehension construct demonstrates multidimensionality, as multiple latent traits influence it.

Table 2. Summary of raw variance and unexplained variance for the four dimensions in CAA.

Estimates	Dimension			
	Quantitative	Comprehension	Fluid	Spatial
Raw variance explained by the measure	46.1	26.1	40	60.4
Unexplained variance in the 1st contrast	2.2	1.9	2.1	1.5

3.1.1 Item-Person Map

The item-person map provides valuable insights into the alignment between test items and respondents' abilities (Figure 1). The person mean is 1.33, while the item mean is 0.00, resulting in a difference of 1.33. Since this difference exceeds the 0.5 threshold, the test items are relatively easier for the respondents. The higher mean suggests that, on average, the respondents possess a higher ability level than the test items' average difficulty. Although the items are well-distributed across the difficulty spectrum, most students are clustered above the average item difficulty. This indicates that most test takers found the items easier, highlighting a potential mismatch between item difficulty and respondent ability. To better target the higher-ability group, it may be necessary to incorporate more challenging items into the test.

3.1.2 Individual Item Performance

The Rasch dataset analysis provides detailed insights into the performance of individual items, focusing on critical statistics such as the Measure, Infit, and Outfit Mean Square (MSQ), and the Point Measure Correlation (PTME). The following sections present the performance of items across the four dimensions of the Cognitive Ability Assessment (CAA), providing an overview of the measures, fit, and item polarity for each dimension.

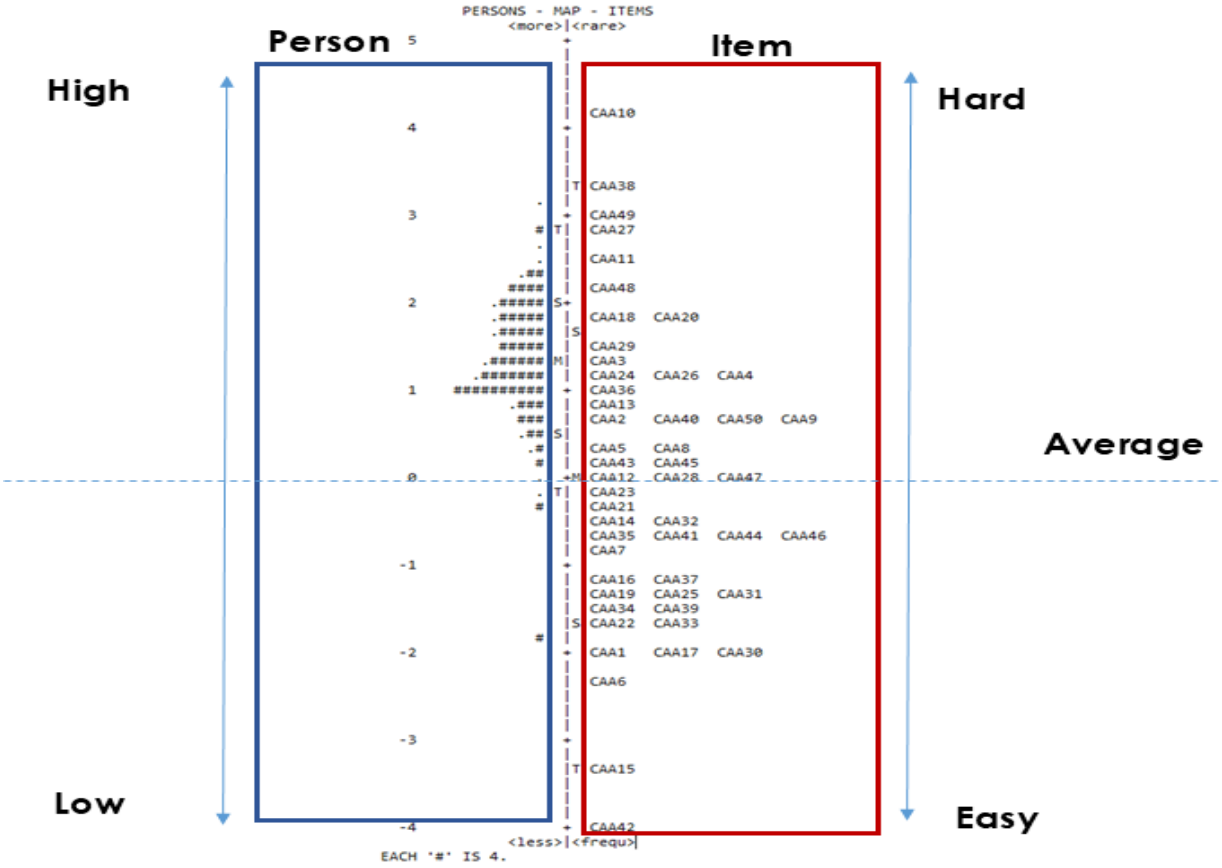

Figure 1. Overall distribution of the item-person map.

3.1.3 Item Measures

The item measures across the four dimensions: Quantitative, Comprehension, Fluid, and Spatial vary in difficulty, reflecting the diversity in item complexity and respondent performance (Table 3). In the Quantitative dimension, items CAA27 (3.59) and CAA11 (3.32) are the most challenging, while CAA15 (-3.12) and CAA39 (-1.35) are the easiest. The Comprehension dimension similarly shows variation, with CAA48 (1.77) and CAA20 (1.46) as more difficult, and CAA16 (-1.67) and CAA44 (-1.25) as easier. In the Fluid dimension, CAA49 (3.54) and CAA29 (1.75) represent the most challenging items, whereas CAA17 (-2.20) and CAA25 (-1.41) are easier. Lastly, the Spatial dimension includes the most difficult items, with CAA10 (5.16) and CAA38 (4.08) at the higher end of difficulty, while CAA42 (-4.6) and CAA30 (-2.28) are among the easiest. Overall, the difficulty of items across dimensions demonstrates a broad range, allowing for a comprehensive assessment of respondent capabilities.

Table 3. Item measures by dimension.

Quantitative		Comprehension		Fluid		Spatial	
Item	Measure	Item	Measure	Item	Measure	Item	Measure
CAA1	-1.9	CAA2	0.15	CAA5	0.46	CAA6	-2.64
CAA3	1.77	CAA4	0.68	CAA9	0.77	CAA10	5.16
CAA7	-0.51	CAA8	-0.25	CAA13	0.91	CAA14	-0.42
CAA11	3.32	CAA12	-0.6	CAA17	-2.2	CAA18	2.44
CAA15	-3.12	CAA16	-1.67	CAA21	-0.32	CAA22	-1.91
CAA19	-1.11	CAA20	1.46	CAA25	-1.41	CAA26	1.46
CAA23	0.11	CAA24	0.72	CAA29	1.75	CAA30	-2.28
CAA27	3.59	CAA28	-0.55	CAA33	-1.78	CAA34	-1.61
CAA31	-1.16	CAA32	-1.03	CAA37	-1.2	CAA38	4.08
CAA35	-0.43	CAA36	0.45	CAA41	-0.78	CAA42	-4.6
CAA39	-1.35	CAA40	0.12	CAA45	0.27	CAA46	-0.64
CAA43	0.52	CAA44	-1.25	CAA49	3.54	CAA50	0.97
CAA47	0.27	CAA48	1.77				

3.1.4 Item Fit

The item fit analysis is based on Mean Square (MNSQ) values for each dimension: Quantitative, Comprehension, Fluid, and Spatial, as reported in Table 4. The result indicates that most items fall within the acceptable range (0.6-1.4) for the Infit MNSQ, suggesting a good fit for the Rasch model. In the Quantitative dimension, items like CAA11 (1.08) and CAA27 (1.10) exhibit slightly higher but acceptable fit, while CAA15 (0.77) and CAA31 (0.79) indicate lower Infit values, reflecting a tighter fit. The Comprehension dimension shows a similar pattern, with most items fitting well, though CAA16 (1.13) stands out as having a slightly higher value. Most items show a good fit for the Fluid dimension, although CAA45 (1.31) and CAA49 (1.22) are slightly outside the ideal range, indicating a possible misfit. The Spatial dimension generally displays acceptable Infit values, though CAA14 (1.32) and CAA38 (1.13) show a marginally higher-than-desired fit. While most items fit well within the model's expectations, a few items in each dimension may need further investigation to improve their alignment with the intended constructs.

3.1.5 Point Measure Correlation (PTME)

The Point Measure Correlation (PTME) results for the Quantitative, Comprehension, Fluid, and Spatial dimensions provide insights into the alignment of items with the underlying latent traits (Table 5). In the Quantitative dimension, most items align well, with PTME values between 0.2 and 0.8, but CAA15 has a lower PTME of 0.32, indicating weaker alignment. In the Comprehension dimension, items such as CAA20 (0.53) and CAA12 (0.47) show strong alignment, while CAA16 exhibits a notably low PTME value of 0.06, suggesting poor correlation with the construct. In the Fluid dimension, items like CAA13 (0.57) and CAA9 (0.55) display strong alignment, but CAA45 (0.18) and CAA49 (0.23) indicate weaker alignment. In the Spatial dimension, CAA50 (0.50) and CAA18 (0.49) align well, but items CAA10 (0.22) and CAA42

(0.20) show weaker correlations. While most items across dimensions align well with the latent traits, some show weaker correlations and may require further revision.

Table 4. Item fit (MNSQ) by dimension.

Quantitative		Comprehension		Fluid		Spatial	
Item	INFIT (MSQ)	Item	INFIT (MSQ)	Item	INFIT (MSQ)	Item	INFIT (MSQ)
CAA1	0.99	CAA2	1.03	CAA5	0.77	CAA6	0.73
CAA3	1	CAA4	0.95	CAA9	0.83	CAA10	1.07
CAA7	0.98	CAA8	1.09	CAA13	0.81	CAA14	1.32
CAA11	1.08	CAA12	0.86	CAA17	0.78	CAA18	0.95
CAA15	0.77	CAA16	1.13	CAA21	0.97	CAA22	0.8
CAA19	0.93	CAA20	0.9	CAA25	0.97	CAA26	0.86
CAA23	0.96	CAA24	0.96	CAA29	1.13	CAA30	0.72
CAA27	1.1	CAA28	0.94	CAA33	0.87	CAA34	0.94
CAA31	0.79	CAA32	1.03	CAA37	0.98	CAA38	1.13
CAA35	0.94	CAA36	1.1	CAA41	1	CAA42	1.07
CAA39	0.87	CAA40	1	CAA45	1.31	CAA46	0.96
CAA43	1.03	CAA44	0.93	CAA49	1.22	CAA50	1
CAA47	1.07	CAA48	1.07				

Table 5. Distribution of point measure correlation (PTME).

Quantitative		Comprehension		Fluid		Spatial	
Item	PTME	Item	PTME	Item	PTME	Item	PTME
CAA1	0.26	CAA2	0.37	CAA5	0.6	CAA6	0.43
CAA3	0.47	CAA4	0.45	CAA9	0.55	CAA10	0.22
CAA7	0.38	CAA8	0.29	CAA13	0.57	CAA14	0.28
CAA11	0.39	CAA12	0.47	CAA17	0.38	CAA18	0.49
CAA15	0.32	CAA16	0.06	CAA21	0.41	CAA22	0.45
CAA19	0.37	CAA20	0.53	CAA25	0.32	CAA26	0.58
CAA23	0.43	CAA24	0.45	CAA29	0.39	CAA30	0.45
CAA27	0.31	CAA28	0.39	CAA33	0.34	CAA34	0.43
CAA31	0.43	CAA32	0.28	CAA37	0.33	CAA38	0.26
CAA35	0.39	CAA36	0.32	CAA41	0.36	CAA42	0.2
CAA39	0.35	CAA40	0.39	CAA45	0.18	CAA46	0.49
CAA43	0.42	CAA44	0.35	CAA49	0.23	CAA50	0.5
CAA47	0.36	CAA48	0.38				

3.1.6 Differential Item Functioning

The Differential Item Functioning (DIF) analysis of the CAA flagged several items as having significant bias between Psychology and Non-Psychology students (Table 6). Of the 50 items analysed, 10 exhibited large DIF, while 4 showed intermediate DIF. Large DIF was more prevalent, particularly in the Comprehension (Comp) and Quantitative (Quant) dimensions, with the Comprehension dimension showing the highest number of DIF-flagged items. These findings suggest that students from Psychology programmes performed better on certain comprehension and quantitative items than their non-psychology counterparts, as reflected in higher measure scores for Psychology students. Among the flagged items, CAA2, CAA8, and CAA16 (from the Comprehension dimension), along with CAA35 (from the Quantitative dimension), showed significant DIF favouring Non-Psychology students. Conversely, items such as CAA12, CAA20, and CAA44 (Comprehension), and CAA45 (Fluid Knowledge) favoured Psychology students. These results highlight that the Comprehension dimension presents the most prominent DIF effects, suggesting that language may play a critical role in these disparities. Similarly, the Quantitative dimension showed a mix of large and intermediate DIF, indicating areas where bias might influence student performance. Interestingly, the Fluid and Spatial dimensions had fewer items flagged, with most showing either negligible DIF or favouring psychology students when large DIF was present. The analysis emphasises the need to revise problematic items, particularly those with a systematic bias in favour of one group. This suggests that while the CAA demonstrates overall reliability, certain items may require further refinement to improve its validity across diverse student groups.

Table 6. Summary of differential item functioning result.

Item	Measure (0)	Measure (1)	M-H Size	*p*	Remark	Dimension	Favor
CAA1	-1.23	-3.15	0.22	0.009	Negligible	Quant	-
CAA2	1.11	0.27	0.65	0.005	Large	Comp	Non-psychology
CAA8	0.73	-0.12	0.91	0.003	Large	Comp	Non-psychology
CAA9	1.15	0.29	0.5	0.014	Intermediate	Fluid	Non-psychology
CAA10	3.64	4.83	0.97	0.078	Large	Spat	Psychology
CAA12	-1	0.39	0.83	<.001	Large	Comp	Psychology
CAA16	-0.34	-2.05	1.13	0.002	Large	Comp	Non-psychology
CAA18	1.53	2.18	0.44	0.044	Negligible	Spat	-
CAA19	-0.9	-1.67	0.11	0.030	Negligible	Quant	-
CAA20	0.91	2.58	1.78	<.001	Large	Comp	Psychology
CAA24	0.07	1.88	1.4	<.001	Intermediate	Comp	Psychology
CAA27	3.63	2.34	0.56	0.003	Intermediate	Quant	Non-psychology
CAA28	-0.48	0.25	0.63	0.031	Intermediate	Comp	Psychology
CAA29	2.08	1.1	0.85	0.001	Large	Fluid	Non-psychology
CAA35	-0.1	-1.29	1.15	<.001	Large	Quant	Non-psychology
CAA44	-1.86	-0.24	1.2	0.008	Large	Comp	Psychology
CAA45	-0.56	0.66	0.72	<.001	Large	Fluid	Psychology

Note. 0 = Non-Psychology programme; 1= Psychology programme.

3.1.7 Exploratory Factor Analysis on each Dimension

The variance explained and the number of items deleted for each dimension reflect the optimisation of the factor structure during analysis (Table 7). After removing three items, the ‘Spatial' dimension explained the highest proportion of variance (48%), indicating that the retained items strongly capture the underlying construct. The ‘Quantitative' dimension, with 42% of variance explained and three items deleted, demonstrates a relatively high level of variance explained after item refinement. The ‘Fluid' dimension explained 40% of the variance. However, it required deleting six items, suggesting that many of its original items did not contribute effectively to the construct, possibly reflecting issues with item quality or coherence. The ‘Comprehension' dimension accounted for 30% of the variance after three items were deleted, yielding the lowest explained variance. This may indicate that further refinement or additional items are needed to capture the construct more effectively.

Table 7. Per cent of variance explained by dimension after removing items.

Dimension	Per cent of variance explained	Number of items deleted	Items deleted
Quantitative	42	3	CAA17, 43, 47
Comprehension	30	3	CAA8, 32, 36
Fluid	40	6	CAA29, 33, 37, 41, 45, 49
Spatial	48	3	CAA10, 14, 38

3.1.8 Unidimensionality Test

The unidimensionality, Cronbach's alpha, and Comparative Fit Index (CFI) results suggest varying levels of reliability and model fit across different constructs (Table 8). The ‘Spatial' construct demonstrated the strongest internal consistency ($\alpha = 0.67$) and unidimensionality index (0.66), along with a high model fit (CFI = 0.94), indicating that its items are reliably measuring a single underlying factor. Similarly, the ‘Quantitative' construct showed adequate unidimensionality (0.61) and a strong model fit (CFI = 0.96), though its Cronbach's alpha (0.65) suggests moderate reliability. ‘Comprehension' also exhibited strong model fit (CFI = 0.97) but lower internal consistency ($\alpha = 0.54$), signalling potential issues with item coherence. Conversely, the ‘Fluid' construct had the lowest unidimensionality (0.48) and CFI (0.83), paired with moderate reliability ($\alpha = 0.60$), suggesting that its items may not fully align with a singular latent factor, and further item revision or deletion may be necessary to improve construct validity.

Table 8. Unidimensionality, reliability and model fit index by dimension.

Dimension	Unidimensionality index	Cronbach alpha	CFI index
Quantitative	0.61	0.65	0.96
Comprehension	0.65	0.54	0.97
Fluid	0.48	0.60	0.83
Spatial	0.66	0.67	0.94

4 DISCUSSION

Rasch analysis was conducted to evaluate the performance of individual items within the CAA and their overall contribution to the instrument's dimensions. The primary aim was to assess how well the items aligned with the theoretical framework and the extent to which they measured the four targeted cognitive domains: Quantitative, Comprehension, Fluid Knowledge, and Spatial Knowledge. To complement this, a unidimensionality test and EFA were employed to investigate the instrument's internal structure, providing a comprehensive assessment of its psychometric properties. Combining these methods ensured triangulation of results, enhancing the robustness and reliability of the evaluation (Bond & Fox, 2015).

The Rasch analysis revealed two key findings. First, items demonstrated a diverse range of difficulty levels, from easy to moderate to hard, across the four dimensions. This range ensured the instrument captured a broad spectrum of respondents' abilities, catering to both low- and high-performing individuals. Second, the distribution of item difficulty aligned with the instrument's intended design, ensuring a comprehensive evaluation of cognitive abilities. Including items at varying difficulty levels enabled the CAA to effectively assess individuals with diverse cognitive capabilities, thereby supporting its construct validity (DeVellis, 2016).

Despite these strengths, the analysis flagged several problematic items that require further scrutiny. Specifically, Item 27 from the Quantitative dimension, Items 45 and 49 from the Fluid Knowledge dimension, and Items 10 and 38 from the Spatial Knowledge dimension exhibited inconsistencies. For example, Item 27, designed to be easy, had a high rate of incorrect responses because its options were poorly constructed, confusing respondents. This issue caused the item to be misclassified as difficult. Similarly, other flagged items were found to lack clarity or to fail to match their intended difficulty level, making them either too easy or overly challenging. These findings highlight the critical role of item clarity and alignment in ensuring that the instrument performs as intended (Linacre, 1994). Addressing these issues through rephrasing and redesigning response options will improve the instrument's precision and accuracy.

The flagged items also revealed evidence of differential item functioning, with responses varying by students' academic majors and prior experiences. While such differences could indicate potential bias, they also provide valuable insights into how the instrument interacts with diverse populations. For example, students with specific academic backgrounds may perform differently on certain items, suggesting opportunities to tailor the instrument or develop training programmes to bridge gaps. The CAA can be refined by addressing these variations to ensure fairness while leveraging these insights to enhance its utility (Kline, 2015).

The EFA provided additional insights by examining the CAA's latent structure. While the theoretical model initially proposed a clear factor structure, the empirical results revealed some misalignments. Such discrepancies are not uncommon in EFA, as the method is inherently data-driven and can uncover latent variables that the theoretical framework may not fully capture (Fabrigar & Wegener, 2012). According to Brown (2015), these misalignments often highlight underlying constructs that may need further exploration or refinement. However, any deviations

must be carefully evaluated to ensure that they remain within acceptable bounds and preserve the integrity of the theoretical model (DeVellis, 2016).

The explained variance and model fit indices improved significantly when problematic items were removed. The final dimensions retained more than three items each, which aligns with established guidelines for scale development (Nunnally & Bernstein, 1994). This adjustment reinforced the instrument's robustness and ensured that the remaining items provided a valid representation of the theoretical constructs. These findings underscore the importance of iterative refinement and ongoing validation to improve the instrument's accuracy and usability.

The CAA demonstrates several strengths that establish it as a valuable tool for assessing cognitive abilities. The distribution of items across varying difficulty levels enables the instrument to comprehensively evaluate abilities, ensuring its applicability to a wide range of respondents. Furthermore, aligning most items with their respective domains provides evidence of the instrument's construct validity. The robust methodology employed, including Rasch analysis and EFA, ensures the reliability of the findings and establishes a strong foundation for the instrument's future use. However, the findings also highlight areas for improvement. The flagged items emphasise greater clarity in item design and response options. Moreover, increasing the sample size and ensuring greater diversity will enhance the generalizability of the results. Validation efforts should address these flagged items while aligning the empirical findings with the theoretical model. These steps will improve the instrument's psychometric properties and ensure its fairness and applicability across diverse populations (AERA, APA, & NCME, 2014).

In conclusion, the CAA has proven to be a reliable and valid tool for assessing employability-related cognitive abilities. While some flagged items require revision, the instrument effectively captures essential cognitive constructs, demonstrating its potential for accurate and comprehensive cognitive assessment. The diversity in item difficulty and the strong alignment with psychometric standards ensure that the CAA is well-suited for use in educational and organisational contexts. Continuing refinement and validation will further enhance its reliability, ensuring that the instrument remains a robust and equitable tool for cognitive evaluation in diverse settings.

ACKNOWLEDGEMENTS

This research is a collaborative effort between CXS Analytics Sdn. Bhd. and Universiti Pendidikan Sultan Idris. The authors are grateful to CXS Analytics Sdn. Bhd. for their funding and consultation on the project, Talent Corporation Malaysia Berhad for their assistance in data collection, and Universiti Pendidikan Sultan Idris for their contributions to the success of the project and the publication of this article. This article is part of the consultation projects registered under UPSI (Codes: 2023-0082-PP-02 and 2024-0231-PP-02).

AUTHOR CONTRIBUTIONS

The first author conceptualised the research idea and conducted the data collection. The second author managed the dataset, conducted the data analysis, and prepared the initial draft of the manuscript. The third and fourth authors contributed to the study design, provided funding and resources, and assisted in the review and interpretation of the findings. The final author supported the data collection and analysis. All authors reviewed the final manuscript and approved it prior to submission.

CONFLICT OF INTEREST

The authors declare that there are no conflicts of interest related to this study.

DATA AVAILABILITY STATEMENT

The data supporting the findings of this study are available from the corresponding author upon reasonable request. Access to the data is subject to approval from both the authors and CXS Analytics due to data ownership and confidentiality considerations.

ETHICAL STATEMENT

This study used anonymised data obtained from a consultation project. Participants provided informed consent prior to participation, and no identifiable information was available to the authors during the analysis.

FUNDING

This research was funded by CXS Analytics Sdn. Bhd. under consultation projects registered with Universiti Pendidikan Sultan Idris (Codes: 2023-0082-PP-02 and 2024-0231-PP-02).

REFERENCES

Arthur, W. J., Day, E. A., McNelly, T. L., & Edens, P. S. (2003). A meta-analysis of the criterion-related validity of assessment centre dimensions. *Personnel Psychology, 56*(1), 125–153. https://doi.org/10.1111/j.1744-6570.2003.tb00146.x

Bartram, D. (2005). The great eight competencies: A criterion-centric approach to validation. *Journal of Applied Psychology*, 90(6), 1185–1203. https://doi.org/10.1037/0021-9010.90.6.1185

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

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). Guilford Press.

Cheong, K. C., Hill, C., Fernandez-Chung, R., & Leong, Y. C. (2016). Employing the 'unemployable': Employer perceptions of Malaysian graduates. *Studies in Higher Education*, *41*(12), 2253–2270. https://doi.org/10.1080/03075079.2015.1034260

Chien, T. W., Hsu, S. Y., Chein, T., Guo, H. R., & Su, S. B. (2008). Using Rasch analysis to validate the revised PSQI to assess sleep disorders in Taiwan's hi-tech workers. *Community Mental Health Journal*, *44*, 417–425. https://doi.org/10.1007/s10597-008-9144-9

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation, 10*(7), 1–9. https://doi.org/10.7275/jyj1-4868

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). Sage Publications.

DeVellis, R. F. (2016). *Scale development: Theory and applications* (4th ed.). Sage.

Djiwandono, P. I. (2006). Cultural bias in language testing. *TEFLIN Journal*, *17*(1), 81–88. https://doi.org/10.15639/teflinjournal.v17i1/85-93

Ercikan, K., Por, H. H., & Guo, H. (2023). Cross-cultural validity and comparability in assessments of complex constructs. In N. Foster & M. Piacentini (Eds.), *Innovating assessments to measure and support complex skills* (pp. 190–210). OECD Publishing. https://doi.org/10.1787/e5f3e341-en

Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. *American Journal of Theoretical and Applied Statistics, 5*(1), 1–4. https://doi.org/10.11648/j.ajtas.20160501.11

Fabrigar, L. R., & Wegener, D. T. (2012). *Exploratory factor analysis*. Oxford University Press.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*(3), 272–299. https://doi.org/10.1037/1082-989X.4.3.272

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods, 9*(4), 466–491. https://doi.org/10.1037/1082-989X.9.4.466

Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, *114,* 254–280. https://doi.org/10.1016/j.techfore.2016.08.019

Ismail, A. A., & Hassan, R. (2019). Technical competencies in digital technology towards Industrial Revolution 4.0. *Journal of Technical Education and Training*, *11*(3), 55–62. https://doi.org/10.30880/jtet.2019.11.03.008

Kline, R. B. (2015). *Principles and practice of structural equation modelling* (4th ed.). Guilford Press.

Linacre, J. M. (1994). *Sample size and item calibration stability.* Rasch Measurement Transactions. https://www.rasch.org/rmt/rmt74m.htm

Linacre, J. M. (2026). *A user's guide to Winsteps Ministep: Rasch-model computer programs*. Winsteps.com. https://www.winsteps.com/a/Winsteps-Manual.pdf

Linacre, M. (2025). *Winsteps now does CMLE and JMLE: Multiple-choice, rating scale and partial credit Rasch analysis*. Winsteps.com. https://www.winsteps.com/winsteps.htm

Mustapha, R., & Abdullah, A. (2004). Malaysia transitions toward a knowledge-based economy. *The Journal of Technology Studies, 30*(3), 51–61. https://doi.org/10.21061/jots.v30i3.a.8

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.

Raven, J., Raven, J. C., & Court, J. H. (2000). *Manual for Raven's Progressive Matrices and Vocabulary Scales*. Harcourt Assessment.

Revelle, W. (2026). *Psych: Procedures for psychological, psychometric, and personality research*. https://CRAN.R-project.org/package=psych

Roslan, F. N., Mohd Fuzi, N., Idris, N., Che Hashim, H. I., Abd Razak, S. S., & Ong, S. Y. Y. (2024). Factors affecting graduate employability in Malaysian public university. *International Journal of Academic Research in Business & Social Sciences, 14*(12), 2382–2387. http://dx.doi.org/10.6007/IJARBSS/v14-i12/24198

Saleh, H., & Abdul Wahab, N. A. (2025). Employers' perspectives on Malaysian graduates' skills: A contemporary study. *Journal of TVET and Technology Review, 3*(1), 16–23. https://doi.org/10.30880/jttr.2025.03.01.002

Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review, 18*(4), 210–222. https://doi.org/10.1016/j.hrmr.2008.03.003

Schneider, W. J., & McGrew, K. S. (2018). The Cattell–Horn–Carroll theory of cognitive abilities. In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (4th ed., pp. 73–163). The Guilford Press.

Schwab, K. (2016). *The fourth industrial revolution*. World Economic Forum.

Sedgwick, P. (2014). Cross-sectional studies: Advantages and disadvantages. *BMJ*, *348*, g2276. https://doi.org/10.1136/bmj.g2276

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

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Pearson Education.

Tee, P. K., Wong, L. C., Dada, M., Song, B. L., & Ng, C. P. (2024). Demand for digital skills, skill gaps and graduate employability: Evidence from employers in Malaysia. *F1000Research*, *13*, 389. https://doi.org/10.12688/f1000research.148514.1

Tennant, A., & Conaghan, P. G. (2007). The Rasch measurement model in rheumatology: What is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis & Rheumatism*, *57*(8), 1358–1362. https://doi.org/10.1002/art.23108

Thurstone, L. L. (1947). *Multiple-factor analysis: A development and expansion of the vectors of mind.* University of Chicago Press.

Wang, T. (2010). Comparative evaluation of survey methods. In J. Sheth & N. Malhotra (Eds.), *Wiley International Encyclopedia of Marketing*. Wiley. https://doi.org/10.1002/9781444316568.wiem02043