Print ISSN 2094-1064
Online ISSN 2244-0437

doi: http://dx.doi.org/10.7828/ljher.v19i2.1607

Enhancing Real-World Relevance of Visual-Spatial Career Aptitude Testing: Development and Validation in a Malaysian Context

SABRENA GABRIELLA AROSH
ORCID NO. 0000-0003-2979-7431
E1100226@helplive.edu.my
Department of Psychology, HELP University,
Kuala Lumpur, Malaysia

MARIA FELICITAS M. MAMAUAG
ORCID NO. 0000-0002-3323-7683
mmamauag@ldcu.edu.ph
Office of the University Research and Coordination
Liceo de Cagayan University
Cagayan de Oro City, Philippines

ABSTRACT

This study describes the creation of a Visual-Spatial Aptitude Test, a crucial component of the Multiple Aptitude Test development project at a Malaysian private university. Despite the value of aptitude testing in career guidance, it has been overshadowed by interest and personality assessments. In response, the HELP Career Readiness Evaluation System (HELP-CaRES) was devised, encompassing career readiness, employability skills, personality, interests, and aptitude. However, initial versions faced challenges in accurately assessing students aged 17-25. This study addresses this age group's scarcity of visual-spatial items, developing a 30-item instrument called MAT-D (VS) to measure various spatial skills and enhance their real-world relevance in aptitude testing. Confirmatory factor analysis and Rasch analysis on 149 undergraduates demonstrated good model fit and reliability. Subsequent refinements, involving 203 undergraduates, exhibited improved reliability and model fit. Empirical support for Carroll's three-factor theory in an Asian context emerged. The study proposes further revisions, norm establishment, and application of item response theory models for continued enhancement. The need for further studies underscores the commitment to ensuring that the newly developed tests are

meaningful and applicable to actual career contexts in Asia. Despite the focus on test development, the goal remains to produce assessments that effectively inform career guidance and decision-making processes.

Keywords: aptitude testing, visual-spatial aptitudes, career guidance, confirmatory factor analysis, Rasch analysis

INTRODUCTION

The historical path of career counseling in Asia underscores the crucial need for cultivating a competent workforce, yet the reliance on Western models raises concerns about cultural bias. This issue is particularly pertinent in Malaysia, where the emerging field of career counseling lacks locally validated assessment tools. The limited focus on aptitude testing, compounded by the absence of locally tailored assessments at the school level, poses significant challenges to effective career guidance. Addressing this gap, a local university's career counseling center initiated the HELP-Career Readiness Evaluation System (HELP-CaRES), aiming to develop a holistic measure of career readiness for Malaysians aged 14-25 (Mamauag et al., 2016). However, challenges arose in aligning visual-spatial items with the appropriate difficulty levels for tertiary students (17-25). Consequently, this study attempted to develop a visual-spatial aptitude scale tailored for tertiary students aged 17-25, employing rigorous psychometric methods to ensure validity and reliability.

Defining Aptitudes

Aptitudes, distinct from achievements, predict abilities rather than measuring taught knowledge (Cohen & Swerdlik, 2018). Gottfredson (2003) defines ability as successful task performance given opportunity and motivation, encompassing cognitive abilities such as mental manipulation of facts (p. 117). Hierarchical models, notably Carroll's three-stratum theory, integrate cognitive ability insights (Anastasi & Urbina, 1997). Stratum III or g-factor predicts job performance across various roles, while Stratum II abilities like mathematical reasoning influence degree choices (Gottfredson, 2003; Kell, Lubinski, & Benbow, 2013). Abilities align with interests, guiding career decisions, particularly for gifted individuals (Gottfredson, 2003; Kell et al., 2013). Non-cognitive factors minimally impact job performance predictions compared to mental abilities (Gottfredson, 2003).

Gottfredson (2003) recommended assessing verbal, spatial-mechanical, mathematical reasoning, and clerical speed as minimum aptitudes relevant to various professions. Spatial abilities were highlighted by Hegarty and Waller (2005) and Newcombe and Shipley (2015), emphasizing their importance in daily activities, creativity, and academic performance, particularly in mathematics and sciences. Studies by Yilmaz (2009) and Zhang and Lin (2015) further supported the significance of spatial skills in advanced mathematical understanding and arithmetic outcomes. The multidimensionality of spatial ability, as described by Hegarty & Waller (2005) and Linn and Peterson (1985, as cited in Yilmaz, 2009), aligns with Carroll's (1993) hierarchical model. Magno's (2009) taxonomy of test items provided a framework for delineating various constructs and creating aptitude test items, informing the development of the Multiple Aptitude Test (MAT) forms and their operational definitions (Magno, 2009).

Visual-Spatial Aptitudes

Figure-ground perception discerns figural boundaries from backgrounds, utilizing cues for recognition. This study defines it as the capacity to employ figure-ground cues effectively. Object assembly combines visualization with problem-solving, requiring test takers to analyze and reconstruct disassembled images into cohesive wholes. Progressive series tasks demand analytical reasoning and visualization skills to discern abstract relations and infer missing elements accurately, defining them as the ability to uncover and apply abstract rules within geometric sequences. Surface development is regarded as a subset of spatial visualization, crucial in drafting, physics, and mechanical courses. Visual discrimination entails identifying similarities or differences between objects, crucial for tasks requiring exact matching or differentiation.

Topology's exclusion from the current study's scale development reflects the conceptual richness and varied definitions found in the literature. Accurate assessment of visual-spatial aptitude is widely recognized as valuable in educational and vocational settings, yet defining its sub-factors remains problematic. Inconsistencies in terminology and classification hinder clear delineation and effective operationalization within aptitude assessments. Carroll's (1993) factor analysis aimed to establish clear sub-factors, but subsequent efforts have been limited. Accurate definitions are crucial for addressing the ongoing challenge of defining visual-spatial ability.

Gottfredson (2003) suggested assessing minimum aptitudes like verbal, spatial-mechanical, mathematical reasoning, and clerical speed across professions. Spatial abilities, emphasized by Hegarty and Waller (2005) and Newcombe and Shipley (2015), correlate with scientific and technical aptitudes (Vernon, 1969, as cited in Anastasi & Urbina, 1997; Yilmaz, 2009; Zhang & Lin, 2015). Carroll's hierarchical model (1993) identified five spatial ability clusters, supporting Magno's (2009) taxonomy for test item development. Object assembly, progressive series tasks, and surface development represent subsets of spatial visualization, engaging problem-solving, and visualization (Ivie & Embretson, 2010; Blum et al., 2016; Magno, 2009). Visual discrimination aids in figure distinctions, crucial in reading development (Carroll, 1993; Catts et al., 2001). However, topology's multifaceted nature challenges its inclusion in this study's scale development (Butner et al., 2015; de Freitas & McCarthy, 2014; Godoy & Rodríguez, 2004). Defining visual-spatial sub-factors remains inconsistent (D'Oliveira, 2004), complicating terminology standardization and test development (Carroll, 1993; D'Oliveira, 2004).

Trends in Aptitude Testing

Recent trends in aptitude testing highlight the increasing recognition of the importance of visual-spatial skills in various academic and professional domains (Farias et al., 2024; Pinna et al., 2021). Aptitude tests play a crucial role in identifying individuals' strengths and potentials, particularly in fields where visual-spatial abilities are highly valued, such as STEM disciplines, architecture, and design (Geer et al., 2019; Zhu et al., 2023).

Clare (2023) further emphasizes the importance of interdisciplinary collaboration in understanding visual-spatial skills, suggesting that insights from psychology, education, neuroscience, and computer science are essential for a comprehensive understanding of the topic. Moreover, Bartlett (2023) discusses how interdisciplinary approaches can lead to innovative research methodologies and practical applications in visual-spatial assessment.

For instance, Geer et al. (2024) conducted a meta-analytic review on spatial anxiety and spatial skills, highlighting the strengths and limitations of this method in synthesizing research findings. Similarly, Gonthier, Harma, and Gavornikova-Baligand (2024) discuss the challenges of longitudinal research in understanding the development of reasoning performance. Robust research methods such as meta-analysis and longitudinal investigation, however, have been a limited discussion on the potential challenges associated with these methodologies in psychometrics.

Ensuring real-world relevance and innovative approaches to aptitude testing will certainly enhance its effectiveness and utility in career guidance and decision-making processes (Cherry, 2023; Mercer, 2021). Recent studies by Hassock & Hill (2022) and Healy (2023) have demonstrated the positive impact of contextually relevant aptitude tests on career outcomes and academic success.

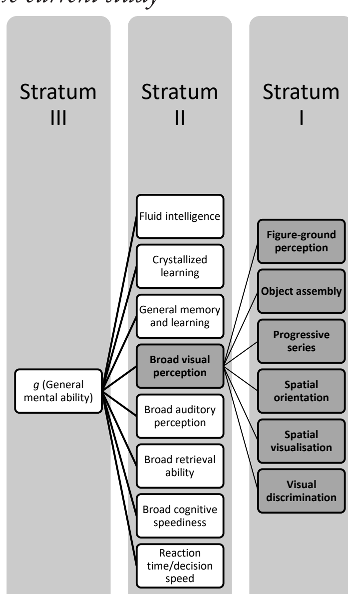
This study, therefore, fills a critical gap in the literature by developing and validating visual-spatial aptitude tests specifically tailored to the Malaysian context. The inclusion of culturally relevant content and types of visual-spatial aptitude like spatial visualization in the new tests sets them apart from existing assessments, making them more suitable for Malaysian students and aligning with local career aspirations and opportunities (Denker et al., 2023; Seemiller et al., 2023).

FRAMEWORK

The conceptual framework of this study is rooted in Carroll's (1993) three-stratum theory of cognitive abilities, which provides a hierarchical structure for understanding and measuring various cognitive constructs. In this framework, specific abilities are situated at stratum I, forming the foundational layer, while broader abilities, such as visual perception, are located at stratum II. This hierarchical arrangement suggests that constructs at different levels are interrelated, with more specific abilities contributing to broader cognitive functions.

Figure 1

Conceptual framework of constructs to be included in the MAT-D (VS) where shaded constructs were used in the current study

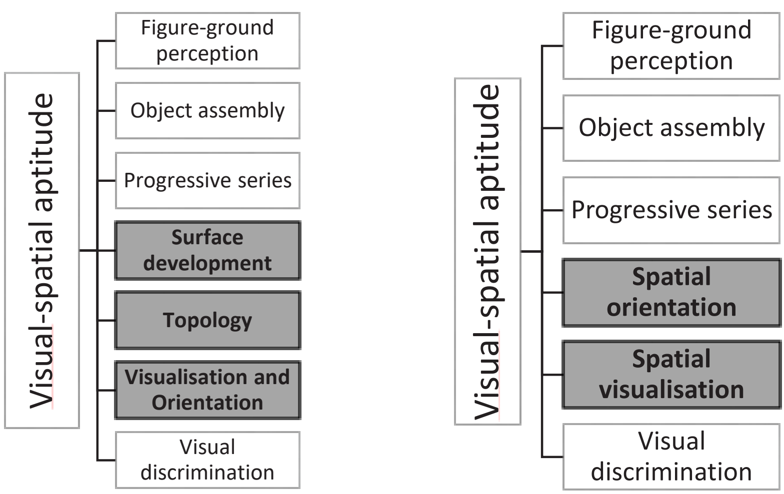


To ensure clarity in construct identification and item development, the study follows DeVellis's (2003) guidelines. This involves refining and specifying constructs before creating measurement items, ensuring that they accurately reflect the intended abilities. Moreover, the exclusion of topology from the visual-spatial scale of the Multiple Aptitude Test-Form D (MAT-D), along with the separate treatment of visualization and orientation, reflects a deliberate effort to align the measurement with the theoretical underpinnings of Carroll's model.

Moreover, the integration of surface development into the spatial visualization measurement further enhances the alignment with Carroll's theory. By restructuring the measurement approach in this manner, the study aims to adhere closely to theoretical frameworks while also incorporating recommendations for construct refinement, as advocated by Yilmaz (2009).

Figure 2

Change in measured constructs between MAT [left] and MAT-D (VS) [right] as shown by shaded constructs



Overall, the conceptual framework outlined in this study serves as a strong foundation, effectively steering the selection and fine-tuning of constructs while ensuring alignment between theoretical underpinnings and empirical measurement strategies. This assertion finds support in the work of Pinna, Conti, and Porcheddu (2021), which draws from Gestalt psychology to underscore the significance of contrast polarity in perceptual organization. Furthermore, Ishikawa

(2021) underscores the critical role of spatial thinking and cognitive mapping, contributing to a comprehensive theoretical framework for understanding visual-spatial abilities. This framework is sustained by Carroll's (1993) three-stratum theory of cognitive abilities, which remains a cornerstone in contemporary research in psychological testing and assessment.

METHODOLOGY

This section provides a concise overview of the study's design, the target population, sample characteristics, and the methodology employed. It begins with a detailed description of the MAT-D(VS) scale, followed by an account of the pilot study and its outcomes. This further outlines the research procedure and discusses the data analysis techniques utilized.

Research Design

Employing a quantitative research design, this study adopts a cross-sectional survey approach. The cross-sectional survey method entails administering the scale to participants only once, ensuring a large dataset is gathered within a short timeframe (Sedgwick, 2014). This methodological choice minimizes participant attrition and ensures data collected is representative of the population's diversity in terms of age, gender, and race (Fraenkel, Wallen, & Hyun, 2012).

Furthermore, the study opts for online distribution over traditional paper-and-pen administration due to several advantages. The online platform offers greater convenience for respondents, enables cost-effectiveness, facilitates quick data collection, and ensures data integrity by preventing non-response (Fraenkel, Wallen, & Hyun, 2012; Qualtrics, 2017).

Population and Sample

The MAT-D(VS) scale targets secondary and tertiary education students aged between 14 to 25 years old, specifically foundation or sixth form and first-year undergraduate students in local public and private tertiary institutions. This demographic aligns with the scale's purpose as a higher-ability alternative to previous versions. The intended age range of participants is between 17 to 25 years old. The sample comprises 203 first-year undergraduate students from the Department of Psychology at a local private university. This sample size meets the requirements for Rasch model analysis and confirmatory factor analysis, as recommended by Şahin and Anıl (2017) and Hair et al. (2010), respectively. Participants were recruited through convenience sampling, with extra credit incentives offered for participation. Despite potential limitations in generalizability, convenience sampling was deemed appropriate given its alignment with the study's objectives and the accessibility of the target population (Bordens & Abbott, 2011).

Instrument

The MAT-D(VS) is a 30-item scale designed to assess visual-spatial aptitude, comprising six subscales, each containing five items. Participants receive scores for each subscale, calculated as the sum of correctly answered items, as well as a total visual-spatial score, obtained by summing all subscale scores. Each item presents four response options, with one correct answer. Preceding the presentation of items in each subscale is an information sheet briefing participants on the subscale's nature, along with a sample item (Appendix A). Given the proprietary nature of the MAT-D(VS) development, the full scale is not included in the appendices.

Initial Development

The development process of the MAT-D(VS) adhered to guidelines outlined by DeVellis (2003), encompassing eight sequential steps. However, deviations from DeVellis' framework occurred, allowing for iterative amendments based on item evaluation. Construct identification involved a thorough literature review to define the constructs under study, aligning with DeVellis' construct identification step. Operational definitions were then developed based on the literature review findings, with slight modifications from preceding MAT versions. Notably, the construct of spatial visualization and orientation was disaggregated into spatial visualization and spatial orientation because of ambiguity in the literature. Additionally, topology was excluded from the MAT-D(VS) due to its complex and mathematically intensive nature.

Item Generation

Item generation was carried out by two researchers, each assigned three constructs based on the operational definitions. Items were designed to align with the defined parameters of the constructs, hand-drawn on graph paper, and digitized using appropriate software. Each item underwent review by the research team and an assessment consultant to ensure content validity, consistent with DeVellis' recommendations. The tool was designed and laid out using an online platform called Qualtrics.

Procedure

Participants received the scale link via email with instructions to complete it in a quiet environment without distractions, emphasizing the requirement to complete it in a single session. Although the estimated completion time ranged from 45 to 60 minutes, no time limit was imposed. Items were presented through Qualtrics, with participants starting on a page displaying the informed consent document, followed by a general information sheet introducing the scale sections.

Subsequent pages presented information sheets and subscales, starting with figure-ground perception. Participants were reminded that they could not backtrack once they submitted responses. Each subscale's information page preceded its corresponding items. Participants selected responses based on their perception of the correctness of each item, followed by providing demographic information.

Upon completion, participants were thanked for their participation. No score reports were provided due to the study's developmental nature, but scores were shared with the lecturer for possible extra credit allocation. Participants could also direct questions via email.

Data Analysis

Psychometric properties of MAT-D(VS) were assessed following Furr's and DeVellis's recommendations. Construct validity was evaluated through confirmatory factor analysis using Mplus, prioritizing its importance over content, criterion, and construct validity. This method examines the relationship between items and the underlying construct, allowing for the removal of items that do not align.

For reliability assessment, Rasch model analysis was conducted using Quest to determine item and person reliability. This method identifies and eliminates items and persons not fitting the measurement model, ensuring accurate scale psychometrics. Additionally, distractor analysis was performed to assess the quality of distractors post-removal of non-fitting cases.

Both factor analysis and Rasch analysis were employed to establish validity and reliability. Factor analysis evaluates construct validity, while Rasch analysis assesses scale reliability and item quality. Raw data, recorded in Excel, were transformed to be compatible with Quest and Mplus for analysis.

The distribution of total scores approximated a normal distribution, supporting the scale's reliability and validity. Skewness and kurtosis fell within acceptable ranges, confirming normality assumptions. The Shapiro-Wilk test further supported the normal distribution assumption.

Measurement Model: Confirmatory Factor Analysis

Confirmatory Factor Analysis (CFA) serves as a statistical technique grounded in exploratory factor analysis principles, often employed to confirm or establish expected correlations and covariances between constructs based on theory predictions or prior analyses (DeVellis, 2003). CFA's advantages lie in its flexibility, allowing for mixed correlated and uncorrelated factors within a single model, aligning with the underlying theory (DeVellis, 2003). It is the method of choice when the internal structure of the scale is clear, crucial for reliability and validity assessment (Furr, 2011).

In this study, CFA is selected as the MAT-D(VS) model structure was previously determined in initial MAT versions. CFA refines scale development by evaluating the measurement model's goodness-of-fit and examining factor loadings, indicating each item's contribution to the intended construct (Furr, 2011). Model fit is determined by the degree of consistency between the hypothesized model and collected data, with various fit indices employed for evaluation.

Method of Estimation

For binary data, like variables with only two response categories, tetrachoric correlations are utilized instead of Pearson correlations (Beauducel & Herzberg, 2006). Maximum Likelihood (ML) estimation, a popular choice, is suitable for normally distributed data (Beauducel & Herzberg, 2006). Weighted Least Squares Mean and Variance adjusted (WLSMV) estimation, introduced by Muthén and colleagues, is preferred for smaller samples, offering less bias compared to earlier methods (Beauducel & Herzberg, 2006).

Fit Indices

Key indices determining goodness-of-fit include the chi-square test, Root Mean Square Error of Approximation (RMSEA), and Comparative Fit Index (CFI). The chi-square test measures the difference between sample covariance and model-fitted matrices, with smaller, non-significant values indicating good fit (Hu & Bentler, 1999). RMSEA, a popular index favoring parsimony, suggests good fit at values below 0.06 (Hu & Bentler, 1999). CFI, complementary to RMSEA, indicates good fit at values equal to or greater than 0.95 (Hu & Bentler, 1999).

Measurement Model: Item Response Theory

Item Response Theory (IRT) serves as a psychometric theory employed in test development, assessing latent traits or abilities (Baker, 2001). IRT assumes respondents possess varying degrees of underlying ability, with responses linked to these abilities, aligning with principles of invariant measurement (Bond & Fox, 2015).

Rasch Model

Among IRT models, the Rasch model is prominent for its ability to attain invariant measures, essential for independent comparisons between test items and respondents (Englehard, 2013). The model runs item calibration and person proficiency estimation, ensuring objectivity (Yu, 2017). Data transformation to logits facilitates independent comparisons, with Rasch models typically estimating difficulty parameters, represented as theta (θ), for items and persons (Yu, 2017).

Fit Statistics

Fit indices like infit and outfit mean-square values assess item misfit, with infit mean-squares emphasizing information-weighted statistics (Linacre, 2002). Standardized fit statistics provide insights into misfit likelihood, guiding item selection and assessment (Bond & Fox, 2015). Mean-square values falling between 0.8 to 1.2 indicate optimal fit, essential for accurate measurement (Wright & Linacre, 1994).

Distractor Analysis

Promoting accurate measurement, Siroky and Di Leonardi (2015) offer tips for refining test items, including analyzing distractors, echoing Haladyna's (2004) assertion of a link between test scores and distractor choice. Distractor analysis optimizes the number of distractors per item, guides item retention or revision, and diagnoses item performance issues (Haladyna, 2004).

Distractor analysis provides insights into test takers' abilities based on their distractor choices (Irvin et al., 2012). High-ability individuals tend to select the correct answer, while lower-ability individuals choose distractors more often (Irvin et al., 2012). Effective distractors differentiate between high and low-ability test takers, correlating negatively with total test scores (DiBattista & Kurzawa, 2011).

For a distractor to be effective, it must be plausible and attractive to lower-ability test takers, contributing to its discriminatory power (Siroky & Di Leonardi, 2015). A wrong distractor increases the chances of guessing the correct answer, undermining the effectiveness of the test (Haladyna, 2004).

Pilot Study

Conducted with 149 first-year undergraduate psychology students, the pilot study employed convenience sampling due to time and accessibility constraints (Bordens & Abbott, 2011). Although limited in generalizability, this method allowed for preliminary data collection and analysis, which are sample-independent (Yu, 2017).

Using a cross-sectional survey design similar to the main study, data collection occurred in a single administration, ensuring a representative sample across various demographics (Sedgwick, 2014). Participants signed up through an online portal and were offered extra credit for participation, resulting in a high response rate (Bordens & Abbott, 2011).

Results of the Pilot Study

In the pilot study, the construct validity of items was assessed through confirmatory factor analysis (CFA) using Mplus software (Version 7.0). Mplus was selected due to its capability to estimate models for binary items based on tetrachoric correlations. Goodness-of-fit for the six-factor model of visual-spatial aptitude was evaluated using chi-square, Root Mean Square Error of Approximation (RMSEA), and Comparative Fit Index (CFI).

The chi-square test indicated good model fit ($\chi 2(390) = 401.416$, $p = .334$). This finding was supported by RMSEA (RMSEA = .014, 90% CI [0.000, 0.033]) and a probability of RMSEA < .05 of 1.00, suggesting close model fit. CFI also indicated good model fit (CFI = .982). These results suggest that the six-factor model adequately represents the latent constructs.

Regression coefficients were analyzed to determine the significance of each item within the model. Standardized regression coefficients were examined, with 22 items showing significant loading within the model ($p < .05$). However, one item (SV2) exhibited a significant negative loading. Despite this, SV2 was retained for further analysis.

Subsequent analysis of squared multiple correlations revealed that 15 items significantly contributed to the variance explained by the constructs ($p < .05$). SV2, despite its earlier negative loading, still made a significant contribution. Items that did not fit the model structure or contribute significantly to the constructs were removed, resulting in 15 retained items.

Correlations between constructs were examined, with all constructs showing significant correlations with visual-spatial aptitude except for one construct

(OA), leading to its removal.

The reliability of the scale was assessed through Rasch analysis using Quest software. Reliability measures, including item reliability, person reliability, and internal consistency, were moderate to high. Although internal consistency fell slightly below the recommended range, this may be attributed to items not loading within the model.

Distractor analysis was conducted on the 15 retained items to assess distractor quality. Results indicated that all distractors were functioning well, except for one item (SO1) and one distractor in another item (SV2). Modifications were made to these items accordingly.

In summary, based on the outcomes of the three main analyses, 15 items were replaced with new items. Amendments were made to certain items, and the revised MAT-D(VS) maintained the format of five items per subscale for the six constructs.

RESULTS AND DISCUSSION

This section presents the findings derived from the confirmatory factor analysis, Rasch analysis, and distractor analysis conducted on the MAT-D(VS). The discussion encompasses the model fit of the MAT-D(VS) concerning fit indices, examination of factor loadings for individual items, Rasch analysis results, and recommendations for further scale development.

Confirmatory Factor Analysis

In assessing the construct validity of the MAT-D(VS), a six-factor model of visual-spatial aptitude was tested using Mplus (Version 7.0; Muthén & Muthén, 2012). The chi-square goodness-of-fit statistic, RMSEA, and CFI were employed as fit indices. Good model fit was determined by $p>0.05$ for the chi-square test, RMSEA<0.06, and CFI>0.95.

Results indicated good model fit, supported by the chi-square test ($\chi2(390) = 384.100$, $p = .575$), RMSEA (RMSEA<.001, 90% CI [0.000, 0.023], PCLOSE=1.00), and CFI (CFI=1.000). Thus, the six-factor structure of visual-spatial aptitude was deemed a suitable representation of the latent construct. Standardized regression coefficients were examined to determine the factor loadings of individual items, as depicted in Table 1.

Table 1

Standardised Regression Coefficients for Items in MAT-D(VS) [Revised]

Item No.	Item Name	Estimate	Standard Error	Two-tailed p-value
1	FGP1	0.502	0.094	0.000**
2	FGP2	0.479	0.106	0.000**
3	FGP3	0.513	0.099	0.000**
4	FGP4	0.829	0.079	0.000**
5	FGP5	0.525	0.093	0.000**
6	OA1	0.037	0.136	0.784
7	OA2	0.855	0.230	0.000**
8	OA3	0.126	0.169	0.456
9	OA4	-0.055	0.145	0.707
10	OA5	0.373	0.131	0.004*
11	PS1	0.100	0.106	0.343
12	PS2	0.432	0.089	0.000**
13	PS3	0.532	0.097	0.000**
14	PS4	0.644	0.086	0.000**
15	PS5	0.450	0.104	0.000**
16	SO1	0.472	0.114	0.000**
17	SO2	0.475	0.093	0.000**
18	SO3	0.439	0.110	0.000**
19	SO4	-0.217	0.104	0.037*
20	SO5	0.412	0.098	0.000**
21	SV1	0.375	0.098	0.000**
22	SV2	0.109	0.108	0.316
23	SV3	0.629	0.096	0.000**
24	SV4	0.610	0.088	0.000**
25	SV5	0.487	0.089	0.000**
26	VD1	0.747	0.061	0.000**
27	VD2	0.628	0.079	0.000**
28	VD3	0.837	0.049	0.000**
29	VD4	0.809	0.058	0.000**
30	VD5	0.845	0.057	0.000**

Note. *. $p < .05$. **. $p < .001$.

Twenty-five items loaded significantly within the model at p<.05, with 23 items loading significantly at p<.001. One item (SO4) exhibited negative factor loadings but was retained for further analysis. Table 2 shows the Squared multiple correlation values were examined to determine the contribution of individual items to the measurement model.

Table 2

Squared Multiple Correlations for Items in MAT-D(VS) [Revised]

Item No.	Item Name	Estimate	Standard Error	Two-tailed p-value
1	FGP1	0.252	0.094	0.007*
2	FGP2	0.230	0.101	0.023*
3	FGP3	0.263	0.102	0.010*
4	FGP4	0.687	0.132	0.000**
5	FGP5	0.276	0.097	0.005*
7	OA2	0.730	0.392	0.063
10	OA5	0.139	0.097	0.153
12	PS2	0.186	0.076	0.015*
13	PS3	0.284	0.104	0.006*
14	PS4	0.415	0.110	0.000**
15	PS5	0.202	0.094	0.031*
16	SO1	0.222	0.108	0.039*
17	SO2	0.226	0.088	0.011*
18	SO3	0.193	0.096	0.045*
19	SO4	0.047	0.045	0.297
20	SO5	0.169	0.081	0.036*
21	SV1	0.140	0.073	0.056
23	SV3	0.395	0.120	0.001*
24	SV4	0.373	0.107	0.001*
25	SV5	0.238	0.087	0.006*
26	VD1	0.557	0.091	0.000**
27	VD2	0.394	0.099	0.000**
28	VD3	0.700	0.082	0.000**
29	VD4	0.654	0.094	0.000**
30	VD5	0.714	0.097	0.000**

Note. *. $p < .05$. **. $p <. 001$.

Twenty-one items significantly contributed to the variance explained by the constructs at p<.05. Item SO4, with negative factor loadings, did not significantly contribute to the variance explained and was disregarded. Consequently, 21 items were retained for Rasch and distractor analysis, as detailed in Table 3.

Table 3

Items Retained after Confirmatory Factor Analysis on MAT-D(VS) [Revised]

FGP	PS	SO	SV	VD
FGP1	PS2	SO1	SV3	VD1
FGP2	PS3	SO2	SV4	VD2
FGP3	PS4	SO3	SV5	VD3
FGP4	PS5	SO5		VD4
FGP5				VD5

Correlations between visual-spatial aptitude constructs were examined to determine their relationship to the underlying construct. Table 4 displays the standardized correlation coefficients, revealing moderate to high correlations among most constructs. However, the OA construct exhibited weaker statistical significance compared to others. Notably, the correlation between FGP and SO indicated a potential issue of multicollinearity.

Table 4

Correlations between Constructs within the MAT-D(VS) [Revised]

	FGP	OA	PS	SV	SO	VD
FGP	-					
OA	0.388*	-				
PS	0.854**	0.673*	-			
SV	0.674**	0.881*	0.917**	-		
SO	1.001**	0.578*	0.993**	0.929**	-	
VD	0.783**	0.501*	0.818**	0.790**	0.742**	-

Note. *. $p < .05$. **. $p <. 001$.

These findings provide valuable insights into the validity and structure of the MAT-D(VS), guiding future steps in scale development and refinement.

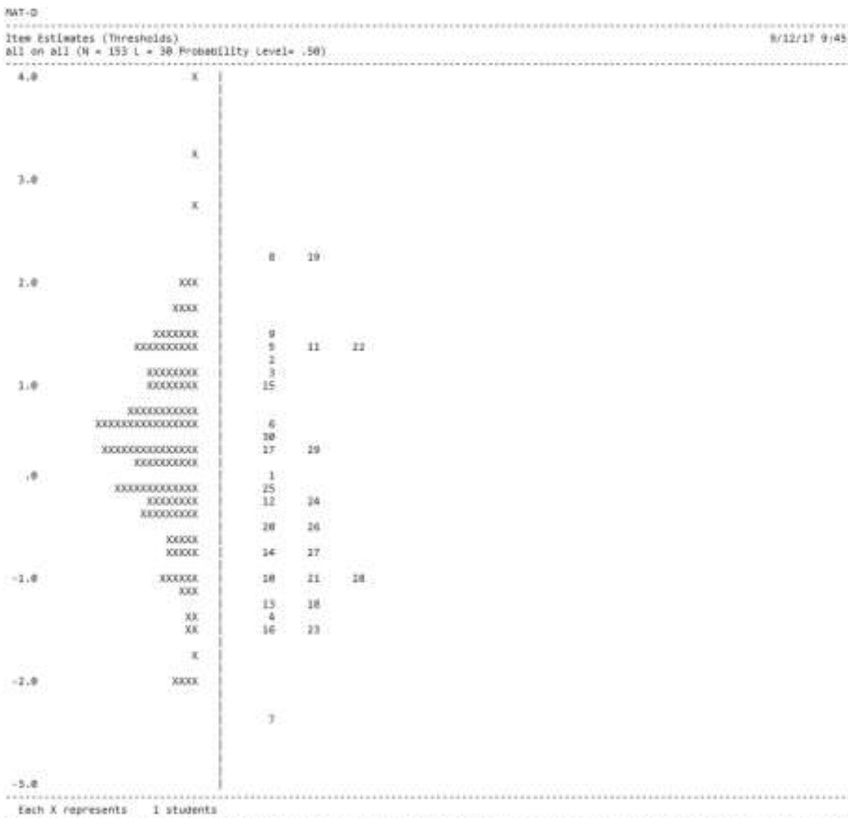
Rasch Analysis

Rasch analysis, facilitated by Quest (Version 2.1; Adams & Khoo, 1996), was conducted to assess the reliability of the MAT-D(VS) scale. Initially, Quest demonstrated a good fit of items to the model, with minor exceptions such as item SO4. Subsequent analyses involved the removal of misfitting cases to enhance the psychometric properties of the items.

After refining the sample size, all items demonstrated a good fit to the model, with satisfactory infit mean-square values. Item reliability was high (0.97), while person reliability was fair (0.78), and internal consistency was deemed fair (α=0.77). The mean score of the valid cases in the model was 16.72 out of 30 (SD=4.87).

Figure 3

Wright map for MAT-D(VS) [Revised]



The Wright map as shown in Figure 3, further illustrates that the test items generally correspond well with the abilities of the examinees, except for item OA2, which appears to be considerably below the general ability measured.

Table 5 provides insights into the fit values, t-statistics, and threshold values for all items. Notably, the majority of items exhibited a satisfactory fit to the Rasch model. Moreover, based on the thresholds and standard error values, the retained items were classified according to their perceived difficulty.

Table 5

Threshold Values, Infit, and Outfit Statistics for MAT-D(VS) [Revised]

Item No.	Item Name	Thresholds	Infit Mean Squared (INFIT MNSQ)	Outfit Mean Squared (OUTFIT MNSQ)	Infit t-statistic	Outfit t-statistic
1	FGP1	0.03	0.99	1.00	-0.1	0.1
2	FGP2	1.14	0.97	1.03	-0.3	0.3
3	FGP3	1.10	0.98	1.10	-0.3	0.7
4	FGP4	-1.34	0.87	0.67	-1.0	-1.5
5	FGP5	1.31	1.02	0.99	0.3	0.0
6	OA1	0.60	1.25	1.47	3.9	3.3
7	OA2	-2.42	0.95	0.93	-0.1	0.0
8	OA3	2.21	0.98	1.10	-0.1	0.5
9	OA4	1.46	1.13	1.38	1.3	2.0
10	OA5	-0.91	1.11	1.01	1.1	0.1
11	PS1	1.35	1.09	1.12	1.1	0.7
12	PS2	-0.25	1.04	1.05	0.6	0.4
13	PS3	-1.24	0.95	0.88	-0.4	-0.5
14	PS4	-0.69	0.89	0.85	-1.3	-0.9
15	PS5	0.97	0.99	1.06	-0.1	-0.4
16	SO1	-1.48	1.00	1.09	0.0	0.4
17	SO2	0.36	1.00	0.95	0.0	-0.3
18	SO3	-1.20	1.05	1.05	0.4	0.3
19	SO4	2.21	1.26	1.44	1.7	1.6
20	SO5	-0.51	1.10	1.18	1.2	1.2
21	SV1	-0.99	1.10	1.10	0.9	0.5
22	SV2	1.31	1.10	1.29	1.1	1.7
23	SV3	-1.43	0.94	0.76	-0.4	-1.0
24	SV4	-0.19	0.91	0.87	-1.4	-0.9
25	SV5	-0.04	1.00	0.96	0.1	-0.3
26	VD1	-0.51	0.90	0.79	-1.3	-1.4
27	VD2	-0.65	0.91	0.85	-1.0	-0.9
28	VD3	-0.91	0.79	0.66	-2.2	-1.9
29	VD4	0.27	0.87	0.82	-2.3	-1.5
30	VD5	0.48	0.84	0.77	-2.8	-1.9

Table 6 indicates that most items fell into the categories of 'easy' or 'moderate,' while a few were classified as 'very easy' or 'difficult.'

Table 6

Thresholds, Standard Error Values, and Perceived Difficulty for 21 Retained Items

Item No.	Item Name	Thresholds	Standard Error	Perceived Difficulty
1	FGP1	0.03	0.18	Moderate
2	FGP2	1.14	0.19	Difficult
3	FGP3	1.10	0.19	Difficult
4	FGP4	-1.34	0.22	Very easy
5	FGP5	1.31	0.19	Difficult
12	PS2	-0.25	0.18	Easy
13	PS3	-1.24	0.22	Very easy
14	PS4	-0.69	0.19	Easy
15	PS5	0.97	0.18	Moderate
16	SO1	-1.48	0.23	Very easy
17	SO2	0.36	0.18	Moderate
18	SO3	-1.20	0.21	Very easy
20	SO5	-0.51	0.19	Easy
23	SV3	-1.43	0.23	Very easy
24	SV4	-0.19	0.18	Easy
25	SV5	-0.04	0.18	Easy
26	VD1	-0.51	0.19	Easy
27	VD2	-0.65	0.19	Easy
28	VD3	-0.91	0.20	Easy
29	VD4	0.27	0.18	Moderate
30	VD5	0.48	0.19	Moderate

Distractor Analysis

The 21 retained items underwent a distractor analysis to evaluate the qualities of distractors and identify areas for refinement. Results indicated that the majority of items contained well-functioning distractors, as evidenced by their positive point-biserial correlation values with the answer key.

However, two items, FGP3 and SO3, required amendments due to problematic distractors. Specifically, distractor A for FGP3 and distractor C for SO3 needed adjustments to enhance their effectiveness. Overall, 19 items were retained without amendments, while two items required minor adjustments to their distractors, ensuring the integrity and reliability of the MAT-D(VS) scale.

Comparison with Pilot Study Results

Comparing the actual study with the pilot study revealed improvements in the psychometric properties of the scale. Model fit improved across all indices in the actual study, with more items showing significant factor loadings and contributions to explained variance.

Inter-construct correlations improved in the actual study, with significant correlations between all constructs. However, some weak correlations and multicollinearity persisted.

In the Rasch analysis, a slightly better fit was observed in the pilot study, but both rounds of analysis showed acceptable values. Reliability values improved from the pilot to the actual study.

In the distractor analysis, 19 items were retained without changes in the actual study, compared to 13 items in the pilot study.

The findings demonstrate the refinement and enhancement of the MAT-D(VS) scale in the actual study compared to the pilot study. These improvements contribute to the scale's validity and reliability in measuring visual-spatial aptitude. Further adjustments based on the identified issues will ensure the continued effectiveness of the scale in future applications.

Reliability Analysis of the MAT-D(VS)

The study employed three indices to assess the reliability of the MAT-D(VS) scale. Internal consistency, measured by Cronbach's alpha (α), indicated a good degree of consistency among items (α=0.77), falling within the recommended range of 0.70 to 0.90. Person reliability, akin to Cronbach's alpha, revealed a confidence level of 0.78 in the ability estimates provided by the sample. Additionally, item reliability, indicating the consistency of item difficulty across different samples, yielded a high value of 0.97. These indices collectively demonstrate the reliability of the scale items and the appropriateness of the sample used for development.

Item Fit Analysis of the MAT-D(VS)

The initial Quest analysis showed favorable fit of most items, except for SO4, which had an infit mean-square value of 1.32. After removing outlying cases, all items demonstrated acceptable fit with infit mean-square values between 0.79

and 1.26. These values align with recommendations by Wright and Linacre (1994), indicating a good fit within the wider range applicable to non-high-stakes assessments.

Furthermore, the Rasch analysis provided insight into item difficulty levels, showing that the test accurately matched the abilities of test takers. The Wright map revealed a distribution of item difficulty similar to the distribution of candidate abilities, with both distributions approximating normality. Thus, the test appears fair and accurately measures the candidate's abilities.

Examination of Item Distractors

The study further focused on analyzing the quality of item distractors in the MAT-D(VS). Nineteen items were retained with no changes, indicating improvement from the pilot study. Distractor analysis, as recommended by Haladyna (2004), supported the refinement of items. Amendments were made to distractors based on student performance.

For instance, in FGP3, distractor A proved confusing to high-ability students due to its similarity to the stimulus. Similarly, in SO3, distractor C led to confusion among high-ability students. Future amendments may involve removing or altering these distractors to improve clarity and effectiveness.

The use of distractor analysis aligns with the test development process outlined by DeVellis (2003), enabling precise refinements at the distractor level. This ensures the development of a reliable and accurate measure capable of distinguishing between low- and high-performing examinees.

Thus, the reliability and item fit analyses demonstrate the robustness of the MAT-D(VS) scale, while the examination of item distractors highlights the importance of refining items for clarity and effectiveness in measurement. These findings contribute to the continued improvement and validity of the scale for assessing visual-spatial aptitude.

Moreover, the findings of this study support the findings of existing literature on visual-spatial aptitude, particularly Carroll's three-stratum theory, and contribute to addressing gaps in understanding and measuring this construct. The identification of suitable constructs and adherence to established guidelines underscore the study's significance in advancing research in this area.

Aligning with Carroll's three-stratum theory of intelligence, which posits Stratum I and Stratum II abilities, the established six-factor structure of visual-spatial aptitude supports and demonstrates correlations among specific abilities indicative of a more general underlying ability. Carroll (2003) suggests that despite correlated abilities, Stratum I abilities should be linearly independent, a concept supported by the study's results.

Moreover, Hegarty and Waller (2005) and Linn and Peterson (1985, as cited in Yilmaz, 2009) advocate for the examination of spatial ability at the level of distinct abilities. The six-factor structure identified in this study provides evidence for a set of abilities suitable for measuring visual-spatial aptitude. Johnson and Bouchard (2005) note the absence of a specific set of abilities for assessing visual-spatial ability, emphasizing the significance of identifying appropriate constructs. The study's chosen constructs, supported by existing literature as sub-constructs of visual-spatial aptitude (Magno, 2009), serve as a foundation for future assessments in this area. Additionally, the model fit evidence from confirmatory factor analysis supports the suitability of these constructs for scale design, allowing for the identification of latent strengths and weaknesses, as suggested by Anastasi and Urbina (1997).

The use of factor analysis likewise mirrors Carroll's (1993) initial work, where this technique was employed to explore the structure of human abilities. While few researchers have undertaken efforts comparable to Carroll's, the current study contributes to understanding the structure of visual-spatial aptitude, aligning with established techniques in theory support (Bickley et al., 1995). Furthermore, the identification of the six-factor structure offers clarity to the field, addressing concerns raised by D'Oliveira (2004) regarding the structure of visual-spatial aptitude. The support for these constructs and the generated operational definitions guide future research, aiding in standardizing essential terminologies in the field of psychological testing.

The study's adherence to DeVellis' (2003) guidelines for scale development proves beneficial in guiding the development process of the MAT-D(VS). The iterative process demonstrates improvements in scale psychometric properties from the pilot to the main study. While further amendments are warranted, particularly concerning object assembly, the efficacy of DeVellis' guidelines in various scale development contexts is supported.

CONCLUSIONS

The study probes into the theoretical implications of its findings, particularly with Carroll's three-stratum theory of cognitive abilities. Through confirmatory factor analysis, the study confirms the six-factor structure of visual-spatial aptitude, aligning with Carroll's framework. This empirical support not only affirms the existence of Stratum I and Stratum II abilities but also extends Carroll's theory beyond its original Western context to an Asian setting. As highlighted by

Pinna, Conti, and Porcheddu (2021), Gestalt psychology's insights into contrast polarity further reinforce the theoretical underpinnings of the study, emphasizing the importance of perceptual organization.

Furthermore, the study contributes to the field by refining operational definitions grounded in extant literature. By updating and validating these definitions, the study paves the way for future research to build upon a standardized conceptual framework. This standardization facilitates the development of new scales and assessment tools, aligning with recommendations for greater coherence within the field (D'Oliveira, 2004). Such advancements not only enhance the robustness of Carroll's three-factor theory but also provide a comprehensive framework for understanding cognitive abilities in diverse contexts.

From a practical perspective, the study lays the groundwork for the development of the MAT-D(VS) scale, showcasing favorable psychometric properties. This tool holds promise for enhancing career guidance processes by providing counselors with accurate insights into students' visual-spatial competencies. By identifying both strengths and weaknesses, the scale empowers counselors to offer tailored guidance, steering students towards compatible career paths (Krumboltz & Vidalakis, 2000). Furthermore, the scale fosters self-reflection among students, facilitating informed career decisions aligned with their abilities (Savickas et al., 2009).

Looking ahead, future research directions involve refining the MAT-D(VS) scale through iterative revisions and comprehensive validation procedures. Amendments to item distractors, generation of new items, and norming procedures are essential steps to ensure the scale's reliability and validity. Moreover, the scale's compatibility with other cognitive assessments and its fairness across diverse demographic groups warrant further investigation. Future studies that would emphasize the significance of cross-cultural research in understanding the universality of psychological constructs such as visual-spatial abilities can be worth pursuing (Tomaszewski Farias et al.,2024; Xu, 2024). Through rigorous refinement and validation, the MAT-D(VS) scale upholds to be a valuable tool for both research and practical applications in career guidance.

In conclusion, the study contributes to theoretical advancements by extending Carroll's theory to an Asian context and refining operational definitions within the field. The development of the MAT-D(VS) scale practically holds promise for enhancing career guidance processes and facilitating informed decision-making among students. Moving forward, continued refinement and validation efforts will solidify the scale's utility and applicability in diverse contexts, further enriching the field of cognitive assessment and career guidance.

LITERATURE CITED

Adams, R. J. & Khoo, S-T. (1996). *Quest: The interactive test analysis system*. Victoria, Australia: The Australian Council for Educational Research Ltd.

Anastasi, A. & Urbina, S. (1997). *Psychological Testing* (7th Ed.). New Jersey: Prentice-Hall, Inc.

Baker, F. B. (2001). *The basics of Item Response Theory* (2nd Ed.). United States: ERIC Clearinghouse on Assessment and Evaluation.

Bartlett, K. (2023). A critical examination of spatial skills assessment: Validity, bias, and technology (Order No. 30685514). Available from ProQuest Dissertations & Theses Global. (2890698651). Retrieved from https://www.proquest.com/dissertations-theses/critical-examination-spatial-skills-assessment/docview/2890698651/se-2

Beauducel, A. & Herzberg, P. Y. (2006). On the performance of Maximum Likelihood versus Means and Variance Adjusted Weighted Least Squares estimation in CFA. *Structural Equation Modelling, 13*(2), 186-203. doi: 10.1207/s15328007sem1302

Bickley, P. G., Keith, T. Z., & Wolfle, L. M. (1995). The three-stratum theory of cognitive abilities: Test of the structure of intelligence across the life span. *Intelligence, 20*(3), 309-328. doi: 10.1016/0160-2896(95)90013-6

Blum, D., Holling, H., Galibert, M. S., & Forthmann, B. (2016). Task difficulty prediction of figural analogies. *Intelligence, 56,* 72-81. doi: 10.1016/j.intell.2016.03.001

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

Bordens, K. S. & Abbott, B. B. (2011). *Research designs and methods: A process approach* (8th Ed.). New York: McGraw-Hill.

Butner, J. E., Gagnon, K. T., Geuss, M. N., Lessard, D. A., & Story, T. N. (2015). Utilizing topology to generate and test theories of change. *Psychological Methods, 20*(1), 1-25. doi: 10.1037/a0037802

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge: Cambridge University Press.

Catts, H. W., Fey, M., Zhang, X., & Tomblin, J. B. (2001). Estimating the risk of future reading difficulties in kindergarten children: A research-based model and its clinical implementation. *Language, Speech, and Hearing Services in Schools, 32*(1), 38-50.

Cherry, K. (2023, March 16). What Is an Aptitude Test? Intelligence Tests vs. Aptitude Tests. Verywell Mind. Retrieved from https://www.verywellmind.com/what-is-an-aptitude-test-2794806

Clare, G. (2023, March 23). 10 Activities to Develop Visual-Spatial Skills in Kids. WonderBaby.org. Retrieved from https://www.wonderbaby.org/articles/visual-spatial-skills

Cohen, R. J. & Swerdlik, M. E. (2018). *Psychological testing and assessment: An introduction to tests & measurement* (9th Ed.). United States of America: The McGraw-Hill Companies, Inc.

D'Oliveira, T. C. (2004). Dynamic spatial ability: An exploratory analysis and a confirmatory study. *The International Journal of Aviation Psychology, 14*(1), 19-38. doi: 10.1207/s15327108ijap1401_2

de Freitas, E., & McCarthy, M. (2014). (Dis)Orientation and spatial sense: Topological thinking in the middle grades. *PNA. Revista de Investigación en Didáctica de la Matemática, 9*(1), 41-51., 41-51. doi: 10.1080/026432900380463

Denker, M., Schütte, C., Kersting, M., Weppert, D., & Stegt, S. J. (2023, January). How can applicants' reactions to scholastic aptitude tests be improved? A closer look at specific and general tests. In *Frontiers in Education* (Vol. 7, p. 931841). Frontiers.

DeVellis, R. F. (2003). *Scale development: Theory and applications* (2nd Ed.). California: Sage Publications, Inc.

DiBattista, D. & Kurzawa, L. (2011). Examination of the quality of multiple-choice items on classroom tests. *The Canadian Journal for the Scholarship of Teaching and Learning, 2*(2), Article 4. doi: 10.5206/cjsotl-rcacea.2011.2.4

Englehard, G., Jr. (2013). *Invariant measurement: Using Rasch models in the social, behavioural and health sciences*. New York: Routledge.

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2012). *How to design and evaluate research in education* (8th Ed.). New York: McGraw-Hill.

Furr, R. M. (2011). *Scale construction and psychometrics for social and personality psychology*. London: SAGE Publications Ltd.

Geer, E. A., Barroso, C., Conlon, R. A., Dasher, J. M., & Ganley, C. M. (2024). A meta-analytic review of the relation between spatial anxiety and spatial skills. *Psychological Bulletin.* Advance online publication. https://doi.org/10.1037/bul0000420

Geer, E. A., Quinn, J. M., & Ganley, C. M. (2019). Relations between spatial skills and math performance in elementary school children: A longitudinal investigation. *Developmental Psychology, 55*(3), 637–652. https://doi.org/10.1037/dev0000649

Godoy, F. & Rodríguez, A. (2004). Defining and comparing content measures of topological relations. *GeoInformatica, 8*(4), 347-371. doi: 10.1023/B:GEIN.0000040831.81391.1d

Gonthier, C., Harma, K., & Gavornikova-Baligand, Z. (2024). Development of reasoning performance in Raven's matrices is grounded in the development of effective strategy use. *Journal of Experimental Psychology: General,* 153(3), 689–705. https://doi.org/10.1037/xge0001518

Gottfredson, L. S. (2003). The challenge and promise of cognitive career assessment. *Journal of Career Assessment, 11*(2), 115-135. doi: 10.1177/1069072702250415

Hair, J. F. Jr., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th Ed.). New Jersey: Pearson Education, Inc.

Haladyna, T. M. (2004). *Developing and validating multiple-choice test items* (3rd Ed.). New Jersey: Lawrence Erlbaum Associates, Inc.

Hassock, L., & Hill, C. (2022). Employability and Employment: The Role of Higher Education in a Rapidly Changing World. In: Ng, B. (eds) Higher Education and Job Employability. *Knowledge Studies in Higher Education,10*. Springer, Cham. https://doi.org/10.1007/978-3-031-05716-8_8

Healy, M. (2023, June 4). Why careers and employability learning matters in higher education. Michael Healy, PhD. Retrieved from https://mojohealy.com/post/why_ career_education_matters/

Hegarty, M. & Waller, D. A. (2005). Individual differences in spatial abilities. In P. Shah & A. Miyake (Eds.), *The Cambridge handbook of visuospatial thinking* (pp. 121-169). New York: Cambridge University Press.

Hu, L.-t. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modelling, 6*(1), 1-55. doi: 10.1080/10705519909540118

Irvin, P. S., Alonzo, J., Lai, C.-F., Park, B. J., & Tindal, G. (2012). Analyzing the reliability of the easyCBM Reading Comprehension measures: Grade 7 (Technical Report #1206). Oregon: *Behavioural Research and Teaching*.

Ishikawa, T. (2021). Spatial thinking, cognitive mapping, and spatial awareness. *Cognitive Processing, 22*(Suppl 1), 89–96. https://doi.org/10.1007/s10339-021-01046-1

Ivie, J. L. & Embretson, S. E. (2010). Cognitive process modeling of spatial ability: The assembling objects task. *Intelligence,* 38(3), 324-335. doi: 10.1016/j.intell.2010.02.002

Johnson, W., & Bouchard, T. J., Jr. (2005). The structure of human intelligence: It is verbal, perceptual, and image rotation (VPR), not fluid and crystallized. *Intelligence, 33*(4), 393-416. doi: 10.1016/j.intell.2004.12.002

Kell, H. J., Lubinski, D., & Benbow, C. P. (2013). Who rises to the top? Early indicators. *Psychological Science, 24*(5), 648-659. doi: 10.1177/0956797612457784

Krumboltz, J. D. & Vidalakis, N. K. (2000). Expanding learning opportunities using career assessments. *Journal of Career Assessment, 8*(4), 315-327.

Linacre, J. M. (2002). What do infit and outfit, mean-square and standardized mean? *Rasch Measurement Transactions, 16*(2), 878. Retrieved from http://www.rasch.org/rmt/rmt162f.html

Magno, C. (2009). Taxonomy of aptitude test items: A guide for item writers. *The International Journal of Educational and Psychological Assessment*. 2, 39-53.

Mamauag, M. F. M., Tai, Y. S., Arosh, S. G., Teo, S. W., & Ooi, H. J. (2016, July). Establishing the psychometric qualities of the Multiple Aptitude Test using the Rasch model. Paper presented at the 31st International Congress of Psychology, Yokohama, Japan. Abstract retrieved from http://onlinelibrary.wiley.com/ doi/10.1002/ijop.12345/epdf

Mercer | Mettl.(2021). *The Absolute Handbook for Understanding Aptitude Tests.* Retrieved from https://resources.mettl.com/wp-content/uploads/2021/06/MM_The-Absolute-Handbook-for-Understanding-Aptitude-Tests.pdf

Newcombe, N. S., & Shipley, T. F. (2014). Thinking about spatial thinking: New typology, new assessments. In *Studying visual and spatial reasoning for design creativity* (pp. 179-192). Dordrecht: Springer Netherlands.

Pinna, B., Conti, L., & Porcheddu, D. (2021). On the role of contrast polarity in perceptual organization: A Gestalt approach. *Psychology of Consciousness: Theory, Research, and Practice, 8*(4), 367–396. https://doi.org/10.1037/cns0000300

Qualtrics [Computer software]. (2017). Retrieved from https://www.qualtrics.com/homepage/

Şahin, A., & Anıl, D. (2017). The effects of test length and sample size on item parameters in item response theory. *Educational Sciences: Theory & Practice*, 17, 321–335. http://dx.doi.org/10.12738/estp.2017.1.0270

Savickas, M. L., Nota, L., Rossier, J., Dauwalder, J.-P., Duarte, M. E., Guichard, J., Soresi, S., Van Esbroeck, R., & van Vianen, A. E. M. (2009). Life designing: A paradigm for career construction in the 21st century. *Journal of Vocational Behavior, 75*(3), 239-250. doi: 10.1016/j.jvb.2009.04.004

Sedgwick, P. (2014). Statistical Question Cross sectional studies: advantages and disadvantages. *BMJ-British Medical Journal, 348.* doi: http://dx.doi.org/10.1136/bmj.g2276

Siroky, K., & Di Leonardi, B. C. (2015). Refine test items for accurate measurement: Six valuable tips. *Journal for Nurses in Professional Development, 31*(1), 2-8. doi: 10.1097/NND.0000000000000123

Tomaszewski Farias, S., De Leon, F. S., Gavett, B. E., Fletcher, E., Meyer, O. L., Whitmer, R. A., DeCarli, C., & Mungas, D. (2024). Associations between personality and psychological characteristics and cognitive outcomes among older adults. *Psychology and Aging.* Advanced online publication. https://doi.org/10.1037/pag0000792

Wright, B. D. & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions, 8*(3), 370. Retrieved from http://www.rasch.org/rmt/rmt83b.html

Xu, H. (2024). In the name of interests: A joint consideration of interest development and consequence. *Journal of Counseling Psychology.* Advanced online publication. https://doi.org/10.1037/cou0000723

Yilmaz, H. B. (2009). On the development and measurement of spatial ability. *International Electronic Journal of Elementary Education, 1*(2), 83-96.

Yu, C-H. (2017). A simple guide to Item Response Theory (IRT) and Rasch modeling. Retrieved from http://www.creative-wisdom.com/computer/sas/IRT.pdf

Zhang, X. & Lin, D. (2015). Pathways to arithmetic: The role of visual-spatial and language skills in written arithmetic, arithmetic word problems, and nonsymbolic arithmetic. *Contemporary Educational Psychology, 41,* 188-197. doi: 10.1016/j.cedpsych.2015.01.005

Zhu, C., Leung, C. O-Y., Lagoudaki, E., Velho, M., Segura-Caballero, N., Jolles, D., Duffy, G., Maresch, G., Pagkratidou, M., & Klapwijk, R. (2023). Fostering spatial ability development in and for authentic STEM learning. *Frontiers in Education,* 8, 1138607. doi: 10.3389/feduc.2023.1138607. Retrieved from https://www.frontiersin.org/articles/10.3389/feduc.2023.1138607/full

ACKNOWLEDGEMENTS

The researchers extend their heartfelt thanks to the study participants for their time and insights, which greatly helped in the completion of this research. Sincere appreciation to the Faculty of Psychology and Behavioral Sciences of HELP University, Kuala Lumpur, Malaysia. Deepest gratitude to CareerSense of HELP University to where this project was conceived.

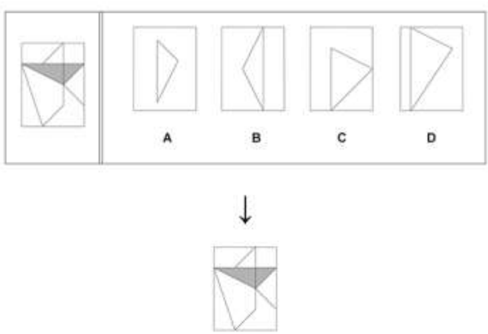
APPENDIX A
Information Sheets for the Subscales of MAT-D(VS)

Figure-Ground Perception

Visual-Spatial Test

Figure-Ground Perception

This sub-test measures your ability to use figural cues to distinguish the boundaries of a figure from the background. In the example, the figures in options A to D may be hidden in the leftmost figure. You are to identify the correct option, i.e. the figure that is hidden in the stimulus. The correct answer is A.



Please note that the borders around the options in this sample item are for size comparison purposes. Please also note that the correct answer can be mirrored, in different orientations, or both.

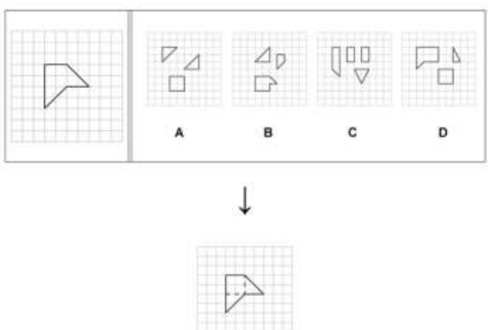
9

Object Assembly

Visual-Spatial Test

Object Assembly

This sub-test measures your ability to reconstruct disassembled items, similar to playing a jigsaw puzzle. The leftmost figure will be a whole object. You are to decide which set of disassembled figures could make up the whole object. The pieces may be rotated, mirrored, and/or displaced. In the example below, Figure A is the correct answer because all the pieces make up the stimulus.



Please note that due to space constraints, all options are scaled accordingly.

12

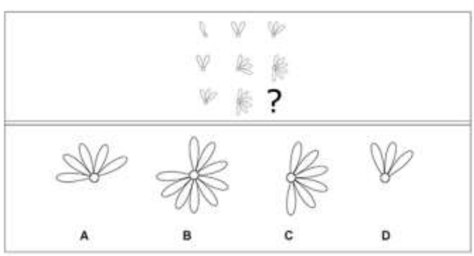
Progressive Series

Visual-Spatial Test

Progressive Series

This sub-test measures your ability to identify the underying pattern of progression across a number of similar figures in the top row. Based on this information, you should infer the missing entry required to complete the progression. In the example given below, flower petals are always added in a clockwise manner. The number of flower petals also increases based on this formula:

Second flower = 2 x (number of petals in first flower in the same row)
Third flower = 3 x (number of petals in first flower in the same row)



For the third row, the number of petals on the last flower follows the formula:

3 x (3 petals on first flower in the same row) = 9 flower petals

Hence, the answer is B.

Please note that due to space constraints, all options are scaled accordingly.

15

Spatial Orientation

Visual-Spatial Test

Spatial Orientation

This sub-test measures your ability to understand spatial relationships that exist in regard to an object, and to identify how an object would look as it is viewed from different angles. In this sub-test, given the leftmost figure, select another figure that is exactly the same but seen from a different orientation. In the sample below, the correct answer is A.

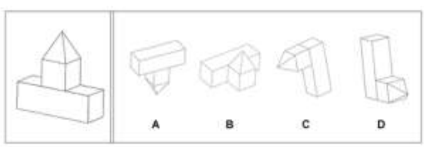


Figure A is the correct answer because it represents the leftmost figure when it is turned upside-down. Figures B, C, and D are incorrect as they involve separation of the parts of the figure, and placement of those constituent parts in varying positions. In other words, Figures B, C, and D are not the same as the leftmost figure as they have been modified.

Please note that due to space constraints, all options are scaled accordingly.

19

Spatial Visualisation

Visual-Spatial Test

Spatial Visualisation

This sub-test measures your ability to mentally manipulate an object and predict how it would look after being rotated, twisted, or mirrored. In this sub-test, the first two figures from the left form the sample set; the third and last figures form the test set. After identfying the manipulations applied to the first figure that changes it into the second figure, you are then required to figure out what happens when the third figure undergoes the same series of manipulations. In the sample below, the correct answer is A.

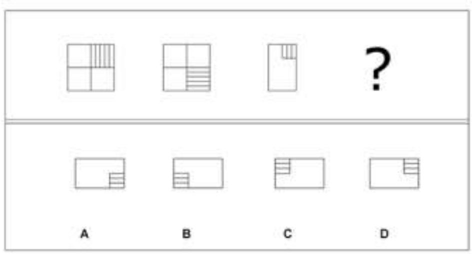


Figure A is the correct answer because it represents the third figure under the same manipulation as in the sample set; it has been rotated 90° clockwise. Figures B, C, D, are incorrect as they do not match the manipulation carried out in the sample set.

Plese note that there may be more than one manipulation carried out in the subsequent questions.Please also note that due to space constraints, all options are scaled accordingly.

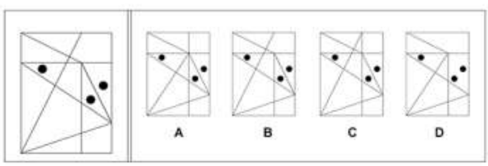
22

Visual Discrimination

Visual-Spatial Test

Visual Discrimination

This sub-test measures your ability to detect if two objects are similar or different. In this sub-test, you are required to select the figure that is exactly the same as the leftmost figure. In the example given below, the answer is B as the lines and dots in both figures are in the exact same locations within their respective boxes.



Please note that due to space constraints, all options are scaled accordingly.

26