Global Cognitive Ability Test

G-CAT

Technical Manual and User Guide

Publisher Information

Seliant Aps
Gammel Hasserisvej 123
9000 Aalborg
Denmark

Product Code: GCAT025

For orders and inquiries:
Seliant Aps, Customer Service
Email: contact@seliant.com
Website: www.seliant.com

G-CAT Technical Manual

This document is the official Technical Manual for the Global Cognitive Ability Test (G-CAT). Within these pages, you will find a comprehensive overview of the theoretical foundations that underpin the G-CAT, the research and validation studies that support its use, and detailed guidance on how to administer, and interpret the results.

The G-CAT is designed to measure core cognitive abilities that are critical to success across a wide range of occupational roles. By providing objective, reliable, and valid measures of these abilities, the G-CAT helps organizations make more informed, evidence-based decisions when selecting and developing talent.

In addition to theoretical insights, this manual includes practical considerations for integrating the G-CAT into your selection system, ensuring fairness and compliance with relevant legal and professional standards, and optimizing its predictive value in various organizational contexts. Whether you are an HR professional, a psychologist, or a stakeholder in the talent management process, the G-CAT Technical Manual serves as an essential resource for understanding and leveraging the full potential of this assessment.

Contents

What is the G-CAT

1.1 Description of the G-CAT

The Global Cognitive Ability Test (G-CAT) is a psychometric tool designed to help organizations make better decisions in their personnel selection processes. By assessing candidates' cognitive abilities, the G-CAT identifies individuals who are best suited for roles where strong mental skills contribute to improved job performance.

Cognitive ability—encompassing constructs such as numerical reasoning, abstract reasoning, and spatial visualization—has consistently been found to be one of the strongest predictors of job performance across a wide range of occupational roles (Sackett et al., 2017). By systematically assessing candidates' cognitive abilities through a standardized, reliable, and valid measure such as the G-CAT, employers can enhance their ability to identify individuals who are more likely to succeed in complex, dynamic, or information-rich work environments.

Unlike subjective methods of evaluation—such as unstructured interviews or informal references—this test ensures that all candidates are assessed under the same standardized conditions, minimizing biases and providing a more equitable platform for talent identification. The G-CAT's design emphasizes fairness, rigor, and scientific grounding. Its items have been validated through empirical research and optimized to reflect the cognitive demands commonly associated with job tasks in modern workplaces.

In practice, the results of this test can be integrated into broader selection systems, combined with other complementary assessments (e.g., personality inventories, structured interviews, skill-based tests) to form a holistic profile of each candidate's potential. By leveraging the predictive power of cognitive ability, organizations can improve the quality of hire, enhance training outcomes, and ultimately gain a competitive advantage through a more capable and adaptable workforce.

1.2 Applications of the G-CAT in Talent Management

The G-CAT is designed to support organizations in personnel selection, placement, and promotion decisions. By incorporating this assessment into a comprehensive talent management strategy, employers can objectively identify candidates with the cognitive problem-solving abilities essential for success across various roles. The test results serve multiple strategic purposes:

1.2.1 Hiring Decisions

At the initial selection stage, test scores provide a standardized, quantifiable measure of a candidate's cognitive skill set. This enables decision-makers to rank-order applicants with precision, ensuring that hiring outcomes are guided by evidence-based criteria rather than subjective impressions or biases.

1.2.2 Role Placement and Internal Mobility

Beyond initial hiring, the test is valuable for evaluating current employees for internal promotions, role changes, or high-potential development programs. By identifying employees who demonstrate the mental acuity required for more complex or strategically demanding positions, organizations can align talent more effectively with job demands.

1.2.3 Workforce Planning and Succession Management

The G-CAT also plays a crucial role in long-term workforce planning. By identifying employees with strong cognitive abilities, organizations can build a pipeline of talent ready to assume leadership roles or tackle new challenges as the company evolves.

1.3 Intended Population for Administration of the G-CAT

The G-CAT is designed for adult candidates or employees seeking positions or advancement within a wide range of occupational fields and organizational levels. While the assessment can be applied broadly, it is most appropriate for individuals who have completed at least a high school level of education or equivalent. The test is suitable for roles that vary in complexity—from entry-level positions requiring fundamental problem-solving skills, to professional and managerial roles demanding advanced reasoning abilities, analytical judgment, and adaptability. The utility of the G-CAT increases with the level of complexity and cognitive demands associated with the tasks of the job position in consideration.

The intended population includes, but is not limited to, individuals applying for roles in industries such as business, finance, engineering, healthcare, technology, and government services. Because cognitive ability has been shown to predict performance across a spectrum of job types and levels, this test can be effectively employed across diverse sectors. It is also adaptable for use with both domestic and international populations, provided that test materials are appropriately translated, adapted, and validated for the linguistic and cultural context in which they are administered.

1.4 Cognitive Abilities Measured by the G-CAT

The G-CAT evaluates cognitive abilities that are essential for problem-solving, decision-making, and other mentally demanding tasks. These abilities are assessed through carefully designed non-verbal tasks to ensure fairness across diverse populations. The cognitive abilities measured in the G-Cat are the following.

1.4.2 Numeric Ability

Numeric ability involves understanding and working with numerical information. It evaluates the ability to:

- Analyze patterns and relationships between numbers.
- Solve quantitative problems.
- Interpret and manipulate data effectively.

This ability is crucial for roles requiring logical thinking, data interpretation, and accurate calculations, such as in finance, engineering, and management positions.

1.4.3 Spatial Ability

Spatial ability assesses the capacity to visualize and manipulate objects in a three-dimensional space. It measures the ability to:

- Recognize spatial patterns and relationships.
- Mentally rotate and transform shapes.
- Solve problems involving physical configurations.

This skill is vital for roles in fields like architecture, design, and technology, where spatial understanding enhances productivity and creativity.

1.4.4 Abstract Ability

Abstract ability evaluates logical thinking and problem-solving skills without relying on language or prior knowledge. It focuses on:

- Identifying patterns and logical rules.
- Drawing conclusions based on incomplete or complex information.
- Adapting to new problems and challenges.

Abstract ability is a strong indicator of general intelligence and adaptability, making it a valuable measure for innovation-driven fields and positions requiring logical thinking.

1.4.5 General Cognitive Ability (GCA)

This ability emerges as the combination of specific reasoning abilities such as spatial, numeric, and abstract ability. It is often represented as “g” in the relevant scientific literature. It represents a person’s broad mental capacity to reason, solve problems, learn, adapt, and perform complex cognitive tasks. Job candidates with high levels of GCA are more likely to:

- Learn quickly and adapt to new challenges.
- Solve complex problems efficiently.
- Contribute relevant ideas to reach organizational goals.

General Cognitive Ability is an excellent predictor of job performance across various tasks and job occupations. It is also a very strong predictor of learning ability and training performance, meaning that candidates with higher levels of GCA are able to learn complex subjects that may be harder for candidates with lower levels. In that sense, GCA is a very valuable factor to take into consideration when making decisions related to personnel selection.

1.5 Components of the Test

The G-CAT consists of three main Components: **instructions**, **test items**, and **test report**. Each part plays a vital role in ensuring the test's effectiveness and fairness while providing actionable insights. This chapter provides an overview of these components and their detailed descriptions.

1.5.2 Instructions

The first part of the G-CAT consists of clear and concise instructions provided to test-takers before they begin the assessment. These instructions are designed to minimize ambiguity and confusion, and be accessible to individuals from diverse backgrounds. This helps test-takers familiarize themselves with the format, reducing anxiety and optimizing their performance.

1.5.3 Test Items

The G-CAT assesses general cognitive ability through three subtests, each targeting a specific reasoning skill. The subtests collectively evaluate fluid intelligence—the ability to think logically and solve novel problems. Each subtest employs non-verbal formats to ensure fairness and accessibility.

1.5.3.1 Spatial Ability Items

Spatial Ability measures the ability to visualize and manipulate objects in space. This subtest includes 10 mental rotation items. Test-takers evaluate a reference object and determine which of several rotated variations matches it.

Spatial Ability Item Example: A 3D object is displayed alongside rotated variations. The task is to indicate if the compared objects are the same or different.

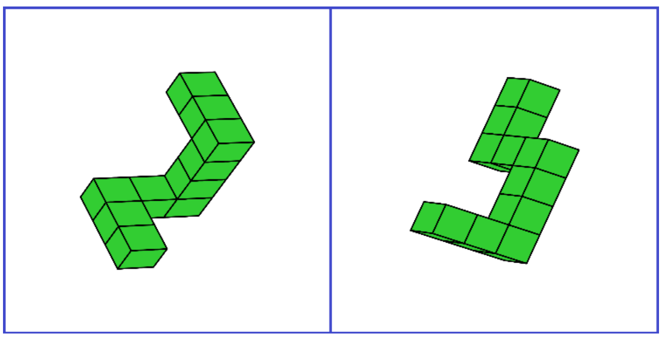

Figure 1: Sample objects used in Spatial Ability items.

Key characteristics include:

- **Visual imagery**: Geometric shapes are used to ensure cultural neutrality.
- **Spatial transformations**: Tasks require mental manipulation of objects along one or more axes.

1.5.3.2 Numeric Ability Items

Numeric Ability reflects the ability to work with numbers, recognize patterns, and solve quantitative problems. This subtest comprises of 10 items presented in a series-type format. Test-takers analyze sequences of numbers or symbols, identify the underlying pattern, and select the correct next item in the series.

Key characteristics include:

- **Non-verbal presentation**: Patterns are visually displayed, eliminating reliance on language.
- **Logical structure**: Tasks involve arithmetic operations or geometric progressions.

Numeric Ability Item Example: A series of numbers is shown. The task is to indicate what is the next number that should be in the series:

1, 2, 4, 8… ?

Figure 2: Sample number series used in Numeric Ability items.

1.5.3.3 Abstract Ability Items

Abstract Ability evaluates the ability to identify relationships, patterns, and logical rules among visual elements. The subtest comprises 10 matrix reasoning items where test-takers deduce the rule governing a grid-like pattern and select the option that completes the matrix.

Key characteristics include:

- **Non-verbal stimuli**: Patterns use shapes, colors, and positions instead of language-based content.
- **Rule-based reasoning**: Test-takers analyze changes in size, shape, and arrangement.
- **Increasing complexity:** Items progress from simple, single-rule patterns to more complex matrices involving multiple rules and transformations.
- **Cultural neutrality:** Designed to minimize cultural or linguistic biases, ensuring fairness for test-takers from diverse backgrounds.

Abstract Ability Item Example: A 3x3 matrix grid is displayed with one missing cell. Test-takers select the option that completes the logical sequence from a set of possible responses.

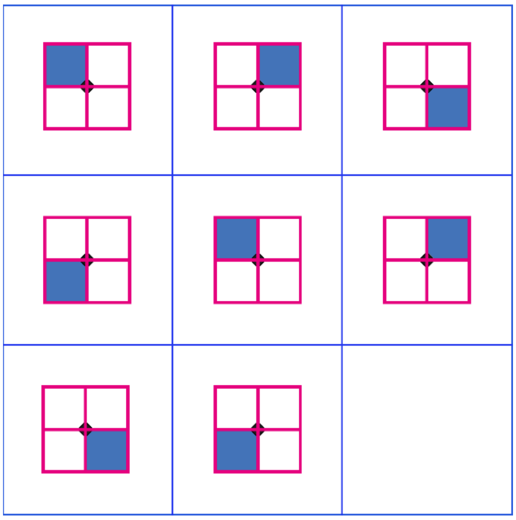

Figure 3: Sample matrix used in Abstract Ability items.

1.5.4 Test Report

After completing the G-CAT, test administrators receive a comprehensive report summarizing the performance of the test-taker. The report includes:

- **General Cognitive Ability (GCA) Score**: A composite score reflecting overall cognitive ability.
- **Subtest Scores**: Separate scores for Numeric, Spatial, and Abstract Abilities, providing detailed insights into specific cognitive domains.
- **Percentile Rankings**: Performance comparisons against a normative sample.
- **Graphical Representations**: Visual aids like bar graphs and percentile curves to enhance interpretability.

The report is designed to be user-friendly and actionable, offering valuable insights into recruitment, talent development, and workforce planning.

1.6 Scoring System

The scoring system of the G-CAT is designed to objectively evaluate test-takers' cognitive abilities and provide meaningful interpretations of their performance. This comprehensive system combines raw score calculation, and norm-referenced scaling to ensure fairness, accuracy, and actionable insights. The reference norms for the G-CAT are derived from the validation sample used in the pilot study for the development of the test, which represents a diverse and global population of various countries and languages.

1.6.1 Raw Score Calculation

The G-CAT employs distinct raw scoring methods tailored to each subtest, ensuring precision and relevance in evaluating specific cognitive abilities.

1.6.1.1 Spatial Ability Subtest

The Spatial Ability Subtest consists of 10 items, each containing 4 sub-items. The scoring method is as follows:

- If **none** of the 4 sub-items are answered correctly: The score for the item is **0**.
- If only **1** of the 4 sub-items is answered correctly: The score for the item remains **0**, reducing the impact of guessing.
- If **2** of the 4 sub-items are answered correctly: The score for the item is **0.5**.
- If **3** of the 4 sub-items are answered correctly: The score for the item is **0.75**.
- If all **4** sub-items are answered correctly: The score for the item is **1**.

Calculating the Total Spatial Ability Score: The total score for the Spatial Ability Subtest is calculated by summing the scores of all 10 items. To ensure a discrete, non-decimal score, the following rounding rules are applied:

- Scores ending in **.25** are rounded **down**.
- Scores ending in **.5** or **.75** are rounded **up**.

Score Range: Individual items: **0** to **1**, Total Spatial Ability Subtest: **0** to **10**.

This method provides an objective evaluation of spatial reasoning performance while preventing inflated scores from random guessing.

1.6.1.2 Numeric Ability Subtest

The Numeric Ability Subtest consists of 10 items designed to measure quantitative problem-solving skills. The scoring is straightforward:

- Each correct answer is awarded **1** point.
- Incorrect or unanswered items receive **0** points.

Score Range: Individual items: **0** to **1**, Total Numeric Ability Subtest: **0** to **10**.

1.6.1.3 Abstract Ability Subtest

The Abstract Ability Subtest also includes 10 items focused on identifying patterns and logical structures. Scoring follows the same method:

- Each correct answer is awarded **1** point.
- Incorrect or unanswered items receive **0** points.

Score Range: Individual items: **0** to **1,** Total Abstract Ability Subtest: **0** to **10**.

1.6.1.4 General Cognitive Ability (GCA) Score

The overall GCA Score is calculated by summing the raw scores of all three subtests:

GCA Score = Numeric Ability Score + Spatial Ability Score + Abstract Ability Score

Score Range: Total GCA Score: **0** to **30**.

This scoring method provides a granular view of performance in specific cognitive domains while offering a holistic measure of overall cognitive ability.

1.6.2 Transformation to Percentile Scores

To enable meaningful interpretation and comparison, raw scores are transformed into percentile scores. These scores indicate the percentage of test-takers in the normative sample who scored equally or below the individual. For instance, a percentile rank of 85 means the test-taker outperformed 85% of the population. Percentile scores closer to 99 are higher and scores closer to 1 are lower.

Purpose of Percentile Scores

- **Comparative Assessment**: Places an individual's score within the context of the normative population, providing insights into their relative standing.
- **Standardized Interpretation**: Converts raw scores into metrics that enable equivalence with scores of other tests that measure the same psychological construct.
- **Fair Benchmarking**: Enables equitable comparisons across diverse populations.

1.6.3 Reference Norms

The norms are based on a globally representative validation sample of 1,232 participants, ensuring that scores are contextualized against a diverse and inclusive population.

1.7 Timing of the Test

The G-CAT incorporates precise timing protocols to evaluate cognitive abilities under realistic, time-pressured conditions. These features ensure the assessment is both efficient and reflective of real-world cognitive demands.

1.7.1 Time Allocation

Each subtest in the G-CAT is designed with specific time limits, totaling **15 minutes** for the complete test. This structured timing is based on empirical research to optimize test reliability, validity, and fairness.

Time Allocation by Subtest:

- **Spatial Ability Subtest**: 6 minutes.
- **Numeric Ability Subtest**: 5 minutes.
- **Abstract Ability Subtest**: 4 minutes.

The time limits are designed to balance two critical objectives:

- **Cognitive Efficiency**: Considers the test-taker's ability to process and respond to information quickly and accurately.
- **Realistic Simulation**: Mimic workplace scenarios where cognitive tasks must often be completed under time constraints.

1.7.2 Key Benefits of Timed Subtests

- Provides a consistent framework for comparing test-taker performance.
- Encourages focused and efficient problem-solving.
- Limits the impact of prolonged deliberation or external factors that could skew results.

1.8 Test-Taker Review of Items

The G-CAT includes an item review feature that allows test-takers to revisit and modify their answers within the allotted time for each subtest. This feature mirrors real-world decision-making processes, where individuals often need to reassess and refine their judgments under time pressure.

1.8.1 Review Mechanism

- **Within-Subtest Review**: Test-takers can navigate freely between items in a subtest as long as time remains.
- **Finalizing Responses**: Once the time for a subtest expires, answers are automatically submitted, and the test-taker proceeds to the next subtest.

1.8.2 Advantages of the Review Feature

- Reduces test anxiety by offering flexibility to correct errors.
- Encourages strategic time management and prioritization.
- Allows test-takers to leverage insights gained while completing later items within a subtest.

Development of the G-CAT

The development of the G-CAT followed a rigorous, systematic process to ensure its validity, reliability, fairness, and applicability across diverse global populations. Guided by professional standards such as the *Standards for Educational and Psychological Testing*, the G-CAT adheres to best practices in psychometric test design and development.

2.1 Key Phases in the Test Development Process

2.1.1 Defining the Objective of the Test

The objective of the test was defined as being able to assess general cognitive ability for personnel selection across various roles and industries by combining assessments of narrower cognitive abilities such as spatial, numeric, and abstract ability. The scores of the tests are meant to be considered when taking decisions about what candidates are the best to consider in personnel selection contexts.

2.1.2 Domain and Content Specification

Cognitive Domains: The test targets four core domains:

- **Spatial Ability**: Evaluated using mental rotation items.
- **Numeric Ability**: Assessed through number series items.
- **Abstract Ability**: Measured with matrix completion items.
- **General Cognitive Ability**: Derived as a higher-order factor integrating the three specific domains of Spatial, Numeric, and Abstract Reasoning.

Timing: Subtests are timed as follows:

- **Spatial Ability**: 6 minutes
- **Numeric Ability**: 5 minutes
- **Abstract Ability**: 4 minutes

These time limits were determined through statistical analysis to calibrate test difficulty and ensure proper differentiation of cognitive abilities.

2.1.3 Item Creation

Item Generation Framework: Items for the G-CAT were developed by operationalizing the underlying cognitive processes required for each cognitive ability. Automatic item generation frameworks were established for each subtest, enhancing the objectivity and scalability of item creation.

Advantages of Automatic Item Generation:

- Objectively align items with defined constructs and cognitive domains.
- Reduces likelihood of error in creating of items when compared to non-automatic item generation.

Pilot versions of each subtest, consisting of 20 items each, were created and administered to the validation sample for empirical testing. From these pilot subtests, 10 high-quality items for each subtest emerged as the final test items. The remaining items were discarded due to sub-optimal psychometric properties.

2.1.4 Pilot Testing

Pilot testing consists of administering a preliminary or "pilot" version of the test to a sample of test-takers that are considered representative of the population the test is intended to be used for. The purpose of pilot testing is to evaluate the psychometric properties of the items and collect data to refine scoring methods and establish required norms for the test.

Validation sample of the pilot test:

The sample used for the pilot test is called "validation sample", as it provides the data to validate the test and determine its psychometric properties. For the G-CAT, a diverse and representative sample of the global workforce, including variations in age, gender, education, and cultural background, was used. The sample size was 1232 participants.

Findings:

The pilot testing identified 10 items with adequate psychometric properties for each subtest. These items represent their intended psychological constructs and can be used to assess different levels of cognitive abilities.

2.1.5. Validation Studies

Validation studies provided robust evidence for the test's ability to measure cognitive abilities accurately and fairly. These studies yielded four types of validity evidence for the G-CAT:

- **Criterion-Related Validity**: The G-CAT scores were strongly correlated with the scores of a 9-item version of the Raven Progressive Matrices, a well-established measure of abstract ability and fluid intelligence.
- **Structural Validity**: Confirmatory Factor Analysis (CFA) confirmed the intended structure of the G-CAT, demonstrating that items correspond to their respective subtests and that the General Cognitive Ability score represents a higher-order factor integrating Spatial, Numeric, and Abstract abilities.

- **Subtest Convergence Validity**: High correlations among the Spatial, Numeric, and Abstract Total Scores aligned with theoretical expectations, supporting the validity of both the subtest and overall scores.
- **Group Differences Validity**: Analyses of mean scores across education levels, academic disciplines, majors, and job categories revealed expected patterns. For instance, individuals in STEM fields and higher education levels consistently scored higher across cognitive domains, supporting the test's construct validity.

2.1.6. Reliability Studies

Reliability metrics demonstrated the Internal Consistency of the G-CAT scores: Coefficients Alpha (≥ 0.70) and Omega (≥ 0.76) indicated strong internal consistency for all subtests and the overall General Cognitive Ability score, confirming the reliability of derived scores.

2.1.7. Fairness Studies

The fairness of the Global Cognitive Ability Test (G-CAT) was assessed by comparing test scores between English and non-English speakers. Analyses included spatial, numeric, abstract, and general ability scores, along with the Raven-9 test as a benchmark. Effect sizes (Eta Squared and Cohen’s d) indicated that language differences accounted for less than 1% of score variance, with negligible to very small differences between groups. Correlation matrices showed consistent relationships among cognitive domains, reinforcing that language background had minimal impact on performance.

The similarity in effect sizes between the G-CAT and Raven-9 supports the test’s validity across linguistic groups. Strong correlations between domain scores and the General Cognitive Ability Score further confirm its fairness. As a non-verbal assessment, the G-CAT minimizes language bias, making it a reliable tool for global personnel selection and ensuring fair evaluation across diverse backgrounds.

2.1.8. Norm Development

Normative data were established using the validation sample as a representation of the global workforce. These norms are used to rank and interpret test scores by transforming raw scores into percentile scores to rank the test-takers.

2.2 Rationale for Selecting the Specific Cognitive Abilities in the G-CAT

The G-CAT was designed to measure cognitive abilities most relevant to personnel selection in a fair, efficient, and culturally inclusive manner. This section explains why the test emphasizes spatial, numeric, and abstract abilities while excluding other commonly assessed abilities, such as verbal reasoning, memory, and processing speed.

2.2.1 Abilities Included in the G-CAT

2.2.1.1 Spatial Ability

This ability involves the capacity to understand, manipulate, and reason about visual shapes, forms, and the relationships between objects in space. Spatial ability is particularly important in fields that require interpreting blueprints, working with complex machinery, engaging in product design, or navigating physical environments.

Justification for inclusion in the G-CAT:

- **Building Block for Mental Representation**
  By mastering spatial concepts, individuals gain a basis for forming, manipulating, and comparing mental images in various cognitive tasks.
- **Cultural Neutrality**: Tasks assessing spatial ability rely on visual stimuli, minimizing linguistic or cultural biases.
- **Job Relevance**: Spatial ability is essential in fields such as engineering, architecture, design, and technical problem-solving.

2.2.1.2 Numeric Ability

This ability refers to an individual's capacity to understand, manipulate, and draw logical conclusions from numerical information. This includes basic arithmetic operations and solving quantitative problems that require the application of mathematical concepts. In many modern work settings—ranging from finance and engineering to operations and logistics—numeric ability is essential for informed decision-making.

Justification for inclusion in the G-CAT:

- **Essential Mental Ability:** Numeric ability is a fundamental cognitive skill that underpins many decision making, as it represents the capacity of dealing with the quantitative aspects and patterns of different problems and situations.
- **Cultural Neutrality**: Numbers and quantitative logic are culturally universal, ensuring fairness in a global testing context.
- **Job Relevance**: Numeric ability is critical for roles requiring data interpretation, financial analysis, and logical decision-making.

2.2.1.3 Abstract Ability

This ability is the capacity to identify patterns, relationships, and underlying principles that are not immediately obvious or tied to already known information. It requires working flexibly with symbols, concepts, and rules to solve problems for which no straightforward, learned solution is available. Abstract ability is especially valuable in innovation-driven roles or dynamic environments, as it supports employees in formulating strategies, detecting trends, and generating novel ideas.

Justification for inclusion in the G-CAT:

- **Essential component of Mental Abilities**: Abstract ability is a direct measure of "fluid" intelligence, which reflects a person's capacity to learn and adapt to new challenges.
- **Cultural Neutrality**: Non-verbal matrix reasoning tasks minimize language dependence, making the test accessible to diverse populations.
- **Generalizability**: Abstract ability is a versatile skill relevant across numerous job types and industries.

2.2.2 Abilities Excluded from the G-CAT

2.2.2.1 Verbal Reasoning Ability

This is the ability to understand, analyze, and reason using language-based information. It encompasses interpreting written passages, identifying relationships between concepts, drawing logical conclusions from text, and evaluating arguments. This cognitive ability requires a strong grasp of vocabulary, grammar, and syntax, as well as the capacity to comprehend nuanced meanings, detect implicit assumptions, and recognize logical inconsistencies.

Justification for Exclusion: Cultural and Linguistic Bias

Verbal reasoning tests are often influenced by a test-taker's language proficiency, educational background, and cultural context. Including this ability would disadvantage non-native speakers and individuals from diverse linguistic backgrounds.

2.2.2.2 Memory Abilities

Memory abilities refer to an individual's capacity to encode, store, and retrieve information over short or long periods. This includes short-term memory (holding information briefly), working memory (actively using information during tasks), and long-term memory (retaining information over time). Memory is essential for learning, problem-solving, and daily functioning, enabling individuals to recall facts, follow instructions, and apply past experiences to new situations.

Justification for Exclusion: Cheating Risk in Online Testing

Memory tasks are highly susceptible to cheating in online environments. Test-takers could use external aids like writing down information, taking screenshots, or consulting unauthorized materials, compromising the test's validity. Unlike reasoning tasks that require real-time thinking, memory tasks can be easily manipulated with such tools. This makes it difficult to ensure scores reflect true cognitive ability, posing a risk to the fairness and reliability of the G-CAT.

2.2.2.3 Processing Speed

Processing speed refers to how quickly and accurately an individual can perceive, interpret, and respond to simple cognitive tasks. Processing speed influences overall cognitive functioning, affecting problem-solving, adaptability, and multitasking. Faster processing speed enhances performance in academic, professional, and everyday situations, contributing to greater productivity and decision-making efficiency.

Justification for Exclusion: Technical Variability

Processing speed tasks are sensitive to technological differences that can unfairly affect scores. Variations in internet speed, device performance, screen refresh rates, and input methods (e.g., touchscreen vs. mouse) can create inconsistent testing conditions. Slow connections or lagging devices may lower scores, while faster equipment could offer an advantage. These technical inconsistencies make it difficult to ensure results reflect true cognitive abilities, compromising the fairness and reliability of the G-CAT in an online setting.

2.2.3 Summary of selection of abilities for the G-CAT

The selection of spatial, numeric, and abstract abilities in the G-CAT underscores the test's commitment to fairness, cultural inclusivity, and practical relevance. By focusing on abilities that minimize linguistic and cultural biases while accurately capturing critical aspects of cognitive functioning, the G-CAT aspires to provide a balanced and reliable measure for diverse populations. Excluding more language-dependent, memory-focused, and speed-based tasks further safeguards test integrity in an online setting. Ultimately, this approach supports consistent, meaningful assessments that empower employers and individuals to make informed decisions grounded the cognitive ability potential of candidates.

2.3 Characteristics of Validation Sample

The validation sample serves as a critical component of the G-CAT development process. It is the foundation to conduct the required statistical and psychometric analyses to ensure that the test and its items are robust, reliable, and generalizable across diverse populations. The validation sample was also designed to align with the overall goals of the G-CAT test by ensuring the inclusion of diverse demographic and professional characteristics, which enhances the generalizability and applicability of the test outcomes. The data provided by the validation sample was used to provide the necessary evidence of validity, reliability, and fairness of the test.

2.3.1 Recruitment and Data Collection

Participants for the validation sample were recruited through an online platform designed specifically for the administration of a pilot form of the test. Recruitment aimed to gather data from individuals that tend to participate in personnel selection procedures. Participants were incentivized by receiving feedback on their performance on various test items. They also provided informed consent to participate in the pilot study.

2.3.2 Administration *of the G-CAT Pilot Form*

The validation sample participants completed the G-CAT pilot form, which consisted of three subtests, each containing 20 pilot items intended to measure spatial, numeric, and abstract abilities. Additionally, a 9-item form of the Raven's Progressive Matrices test was included for purposes of criterion validation. Key aspects of the test administration included:

- **Time Limits**: Each subtest was administered with a strict five-minute time limit, as per the test design. The time limit of each subtest was changed in the formal version of the G-CAT to calibrate their difficulty levels.
- **Instructions**: Clear, concise instructions were provided to all participants to minimize confusion and ensure uniformity.
- **Device Compatibility**: The test was optimized for various devices, including desktops, laptops, tablets, and smartphones.
- **Data Security**: Measures were implemented to protect participant data, including anonymization of responses.
- **GDPR Compliance**: The pilot study was designed to comply with GDPR standards. No personal data was collected during the study. All participant responses were anonymized to ensure privacy. Additionally, participants had the right to request the removal of their anonymized data in accordance with GDPR requirements.

2.3.3 Sampling Method

The validation sample represents a convenience sample. Participants were recruited via an online platform, and access to the test was unrestricted, allowing individuals from various backgrounds to voluntarily participate. Because participants self-selected to participate in the study, the researchers had limited control over the composition of the sample. While this approach ensured diversity, it also introduced potential biases related to self-selection.

After data cleaning procedures, such as screening for inattentive responses and cases with insufficient data, the sample size for the pilot study was reduced from 3784 to 1232. The diversity and representativeness of the data can be observed in the information presented below.

2.3.4 Statistics of the Validation Sample

This section provides a comprehensive overview of the statistical characteristics of participants in the validation sample. The data has been organized into five main categories to facilitate clarity and ease of interpretation: **(1)** Demographic data (age, sex, ethnicity), **(2)** Language data (native language, English proficiency level), **(3)** Regional data (country, continent), **(4)** Employment data (employment status, job category), and **(5)** Education (education level, academic major, academic discipline).

For some of these characteristics, participants were allowed to avoid providing answers. The statistics are shown in tables and charts. For ease of visualization in some charts, some categories with low response frequencies are bundled together into a "Miscellaneous" category. By examining these categories, readers can gain a deeper understanding of the sample's composition and diversity.

2.3.5 Demographic Data

This section provides a broad overview of the demographic data collected for the validation sample. By examining these foundational attributes, readers can gauge the demographic diversity and representativeness of the participant group, setting the context for interpreting the analyses presented in this manual.

2.3.5.1 Age

The ages of the participants go from 18 to 69 years old. The mean age is 29.28 (SD = 10.44), with a median of 26 and a mode of 22. The 25th percentile is 22 and the 75th percentile is 34, reflecting a broad representation of generational cohorts (Table 1, Figure 4).

2.3.5.2 Sex

Male participants account for 56.41% of the validation sample, female participants make up 39.37%, and 4.22% prefer not to disclose their sex. This distribution indicates a moderate ratio between male and female respondents, with a smaller subset choosing non-disclosure (Table 2, Figure 5).

2.3.5.3 Ethnicity

White participants form the largest ethnic group at 25.81%, followed by 13.15% preferring not to disclose their ethnicity. Southeast Asian (9.17%), South Asian (6.98%), British (5.84%), and East Asian (5.84%) form other prominent segments. Additional groups, such as Latino (white) (2.35%), Central Asian (2.35%), and Arab (2.11%), further broaden the distribution. The remaining categories each represent less than two percent of the sample, reflecting a wide-ranging mix of ethnic backgrounds (Table 3, Figure 6).

2.3.6 Language

Participants in the validation sample represent a variety of linguistic backgrounds, spanning multiple first languages and varying uses of English. This diversity underscores the global nature of the study, ensuring a broad range of language experiences within the sample.

2.3.6.1 Native Language

English is the most frequently reported native language at 42.61%, followed by Tagalog (5.60%), Hindi (4.22%), and Indonesian (4.06%). Additional widely spoken languages include Portuguese, Arabic, Spanish, and French, each accounting for smaller portions of the sample. Many other languages, such as Tongan, Bravanese, Betawi, and dozens more, are represented by small yet meaningful percentages, reflecting a broad linguistic diversity within the participant group (Table 4, Figure 7).

2.3.6.2 English Proficiency Level

In response to the self-report question "How well do you speak English?", nearly half of the participants (49.19%) describe their proficiency as "Fluid," while 38.64% consider themselves to speak it "Very Well." Meanwhile, 9.74% rate their English as "Not Very Well," 1.79% prefer not to answer, and 0.65% describe their proficiency as "Poorly" (Table 5, Figure 8)

2.3.7 Regional Data

The validation sample encompasses participants from a variety of countries and continents, reflecting broad geographic diversity. This distribution captures a range of cultural, economic, and social contexts, ensuring wide coverage of regional backgrounds within the study.

2.3.7.1 Country

The largest share of participants is from the United States (20.54%), followed by India (10.15%), the United Kingdom (9.09%), the Philippines (8.36%), and Australia (7.06%). Smaller proportions come from a broad spectrum of other nations, illustrating the wide geographic range of the sample (Table 6, Figure 9).

2.3.7.2 Continent

Asia is the most represented continent at 32.47%, followed by North America (24.84%) and Europe (24.59%). Smaller yet meaningful numbers come from Oceania (7.95%), Africa (6.74%), and South America (3.41%), reflecting the global coverage of the sample (Table 7, Figure 10).

2.3.8 Employment

The validation sample features a broad range of employment status and occupational backgrounds, reflecting diverse industry sectors. This variety highlights the heterogeneity of the participant pool and ensures representation of multiple work contexts.

2.3.8.1 Employment Status

Employed participants form the largest group at 43.83%, followed by 21.67% who are currently students. Meanwhile, 14.94% report being not employed and 11.12% are not employed but seeking work. A smaller portion (6.41%) prefer not to answer, while 1.06% are retired and 0.97% identify as homemakers. These figures showcase a diverse range of labor force participation and life circumstances within the sample (Table 8, Figure 11).

2.3.8.2 Job Categories

Participants' job categories span multiple professional fields. Significant segments include Engineering (11.85%), Banking and Finance (7.31%), Management (7.14%), and Information Technology (6.98%). Smaller yet meaningful portions emerge in areas like Education, Health Care, and Retail, among others. The "Other" category represents jobs that are not included in the main specific categories, being the most frequent response at 22.89%. This distribution underscores the wide-ranging occupations within the sample, reflecting varied expertise and professional backgrounds (Table 9, Figure 12)

2.3.9 Education

Participants' educational backgrounds include a range of degree levels, academic majors, and disciplines, reflecting diverse fields of study and expertise. By capturing both the nature and depth of formal education, the data highlights variations in academic pathways among the validation sample. This overview provides context for understanding how different educational achievements may relate to other participant characteristics and outcomes.

2.3.9.1 Education Level

The largest portion of participants (30.93%) report holding a four-year college or university degree, while 15.34% are currently pursuing such a degree, and 14.94% have completed high school as their highest level of education. Additionally, 12.50% hold a graduate school degree, and 5.11% have some college experience but did not graduate. Smaller segments include those who chose "Other Education" or "Prefer Not To Say" (both 4.55%), those with less than 12 years of education (3.73%), and individuals currently in graduate school (3.08%). A minority hold an associate degree (2.92%), a doctorate degree (1.62%), or are currently working toward a doctorate (0.73%). This range of educational backgrounds reflects the diverse academic profiles among participants (Table 10, Figure 13).

2.3.9.2 Academic Major

Computer Science (9.25%) emerges as the most commonly reported specific major, followed by Business Administration (6.82%). Smaller yet noteworthy shares focus on fields like Education, Mechanical Engineering, Arts, and Social Sciences. Meanwhile, more specialized majors—from Aeronautical Engineering to Botany—underscore the breadth of academic interests within the sample. After these named majors, 8.12% of participants report having no major, potentially reflecting those still exploring academic pathways, while the largest proportion (27.03%) prefer not to disclose their major (Table 11, Figure 14).

2.3.9.3 Academic Discipline

Business is the most commonly reported academic discipline (14.37%), followed by Engineering (12.74%), Computer Sciences (9.50%), Social Sciences (4.87%), and Health (4.06%). Smaller yet meaningful segments include Arts, Education, Natural Sciences, Mathematics, Communications, Public Service, Language and Literature Studies, Design, Aviation, Cultural Studies, and Agriculture, reflecting a diverse array of fields. After these named disciplines, 8.36% of respondents selected “Other,” and 27.03% opted not to disclose their academic discipline (Table 12, Figure 15).

2.3.10 Limitations

- **Convenience Sampling**: As a convenience sample, the findings may not generalize to populations that are not internet-savvy or do not typically participate in online assessments.
- **Self-Selection Bias**: Participants who opted into the study may differ systematically from those who did not, which could influence the representativeness of the sample.
- **Language Proficiency**: English fluency was assessed using a self-report question: "How well do you speak English?" The subjective nature of this measure might not capture actual proficiency levels.
- **Language Representation**
  The validation sample is disproportionately composed of native English speakers, who account for a significant share of the participants. This overrepresentation may limit the generalizability of findings to populations where English is not the primary language, potentially skewing results in favor of native speakers.
- **Regional Representation**
  Participants from North America and Asia dominate the sample, while regions such as South America and Africa are underrepresented. This imbalance reduces the applicability of findings to populations from less-represented regions, limiting the global validity of the test.
- **Platform Access**: Recruitment via an online platform might exclude individuals without reliable internet access, potentially underrepresenting certain demographics.

2.3.11 Tables of the Statistics of the Validation Sample

Table 1: Statistics of Age data.

Statistic	Age
Mean	29.28
Median	26
Mode	22
Standard Deviation	10.44
Minimum	18
Maximum	69
25th Percentile	22
75th Percentile	34

Table 2: Statistics of Sex data.

Sex	N	%
Male	695	56.41%
Female	485	39.37%
Prefer not to say	52	4.22%

Table 3: Statistics of Ethnicity data.

Ethnicity	N	%	Ethnicity	N	%
White	318	25.81%	Han	12	0.97%
Prefer not to say	162	13.15%	West African	12	0.97%
Southeast Asian	113	9.17%	South African	11	0.89%
South Asian	86	6.98%	Middle eastern	8	0.65%
British	72	5.84%	Central African	7	0.57%
East Asian	72	5.84%	Aboriginal Australian	6	0.49%
Other Ethnicity	52	4.22%	Native American	6	0.49%
West European	37	3.00%	Turkic	6	0.49%
Central Asian	29	2.35%	Pacific islander	5	0.41%
Latino (white)	29	2.35%	Persian	5	0.41%
Arab	26	2.11%	Jewish (other)	5	0.41%
Latino (non-white)	25	2.03%	Jewish (Ashkenazi)	4	0.32%
Germanic	21	1.70%	Jewish (Sephardic)	3	0.24%
Slavic	20	1.62%	Caribbean	3	0.24%
Irish	19	1.54%	Joseon	2	0.16%
East African	19	1.54%	North African	2	0.16%
African American	19	1.54%	Yamato	1	0.08%
Nordic	14	1.14%	North Asian	1	0.08%

Table 4: Statistics of Native Language data.

Native Language	N	%	Native Language	N	%	Native Language	N	%
English	525	42.61%	Bulgarian	6	0.49%	Samoan	2	0.16%
Tagalog	69	5.60%	Ukrainian	6	0.49%	Mongolian	2	0.16%
Hindi	52	4.22%	Croatian	6	0.49%	Maltese	2	0.16%
Indonesian	50	4.06%	Greek	6	0.49%	Shona	2	0.16%
Portuguese	29	2.35%	Serbian	6	0.49%	Lithuanian	2	0.16%
Arabic	25	2.03%	Japanese	5	0.41%	Slovak	2	0.16%
Dutch	22	1.79%	Italian	5	0.41%	Berber	2	0.16%
Other Language	22	1.79%	Marathi	5	0.41%	Norwegian	2	0.16%
German	19	1.54%	Acholi	4	0.32%	Cree	1	0.08%
Tamil	17	1.38%	Bosnian	4	0.32%	Mirpuri	1	0.08%
Vietnamese	17	1.38%	Sinhalese	4	0.32%	Georgian	1	0.08%
Prefer not to say	17	1.38%	Visayan	4	0.32%	Kikuyu	1	0.08%
Danish	16	1.30%	Farsi	4	0.32%	Akan	1	0.08%
Bengali	16	1.30%	Amharic	4	0.32%	Hausa	1	0.08%
Spanish	16	1.30%	Polish	4	0.32%	Nepali	1	0.08%
Urdu	15	1.22%	Hungarian	4	0.32%	Tongan	1	0.08%
Malayalam	13	1.06%	Turkish	4	0.32%	Pidgin English	1	0.08%
Mandarin	13	1.06%	Javanese	4	0.32%	Uzbek	1	0.08%
French	13	1.06%	Bravanese	4	0.32%	Cambodian	1	0.08%
Malay	12	0.97%	Punjabi	3	0.24%	Ewe	1	0.08%
Swedish	12	0.97%	Estonian	3	0.24%	Basque	1	0.08%
Gujarati	12	0.97%	Finnish	3	0.24%	Chuukese	1	0.08%
Cantonese	11	0.89%	Azerbaijani	3	0.24%	Kashmiri	1	0.08%
Thai	11	0.89%	Twi	3	0.24%	Dinka	1	0.08%
Russian	9	0.73%	Yoruba	3	0.24%	Icelandic	1	0.08%
Romanian	9	0.73%	Macedonian	3	0.24%	French Canadian	1	0.08%
Telugu	9	0.73%	Albanian	3	0.24%	Fante	1	0.08%
Swahili	8	0.65%	Chin	3	0.24%	Betawi	1	0.08%
Afrikaans	7	0.57%	Igbo	3	0.24%	Ashante	1	0.08%
Czech	7	0.57%	Latvian	2	0.16%	Kazakh	1	0.08%

Table 5: Statistics of English Proficiency Level data.

English Proficiency Level	N	%
Fluid	606	49.19%
Very Well	476	38.64%
Not Very Well	120	9.74%
Prefer not to say	22	1.79%
Poorly	8	0.65%

Table 6: Statistics of Country data.

Country	N	%	Country	N	%
United States	253	20.54%	Serbia	3	0.24%
India	125	10.15%	Morocco	3	0.24%
United Kingdom	112	9.09%	Bulgaria	3	0.24%
Philippines	103	8.36%	Turkey	3	0.24%
Australia	87	7.06%	Portugal	2	0.16%
Indonesia	51	4.14%	Norway	2	0.16%
Canada	47	3.81%	Tunisia	2	0.16%
Brazil	29	2.35%	Nepal	2	0.16%
Netherlands	18	1.46%	Latvia	2	0.16%
Malaysia	18	1.46%	Mongolia	2	0.16%
Denmark	17	1.38%	Azerbaijan	2	0.16%
South Africa	16	1.30%	Colombia	2	0.16%
Vietnam	16	1.30%	Bosnia and Herzegovina	2	0.16%
Sweden	15	1.22%	Estonia	2	0.16%
Germany	14	1.14%	Ethiopia	2	0.16%
Nigeria	13	1.06%	Hungary	2	0.16%
Ireland	12	0.97%	Israel	2	0.16%
Pakistan	9	0.73%	Bahrain	2	0.16%
Egypt	9	0.73%	Lebanon	2	0.16%
France	9	0.73%	Armenia	2	0.16%
Belgium	9	0.73%	Anguilla	2	0.16%
Thailand	9	0.73%	Andorra	2	0.16%
Switzerland	8	0.65%	Uganda	1	0.08%
Spain	8	0.65%	Bahamas	1	0.08%
Finland	8	0.65%	Cambodia	1	0.08%
Kenya	8	0.65%	United Arab Emirates	1	0.08%
Argentina	7	0.57%	Trinidad and Tobago	1	0.08%
Greece	7	0.57%	Belarus	1	0.08%
Romania	7	0.57%	Antigua and Barbuda	1	0.08%
Bangladesh	6	0.49%	Uzbekistan	1	0.08%
Oman	5	0.41%	Bermuda	1	0.08%
Hong Kong	5	0.41%	Taiwan	1	0.08%
Angola	5	0.41%	Botswana	1	0.08%
Czech Republic	5	0.41%	Burundi	1	0.08%
Algeria	5	0.41%	Slovakia	1	0.08%
Sri Lanka	5	0.41%	South Korea	1	0.08%
Austria	5	0.41%	Panama	1	0.08%
China	5	0.41%	Moldova	1	0.08%
Zambia	4	0.32%	Mexico	1	0.08%
Poland	4	0.32%	Mauritius	1	0.08%
Ukraine	4	0.32%	Malta	1	0.08%
Albania	4	0.32%	Madagascar	1	0.08%
Russia	4	0.32%	Lithuania	1	0.08%
New Zealand	4	0.32%	Kazakhstan	1	0.08%
Croatia	4	0.32%	Papua New Guinea	1	0.08%

This table continues on the next page.

Table 6 (continuation): Statistics of Country data.

Country	N	%	Country	N	%
Slovenia	4	0.32%	Chad	1	0.08%
Japan	4	0.32%	Peru	1	0.08%
Tanzania	4	0.32%	Iraq	1	0.08%
Singapore	4	0.32%	Georgia	1	0.08%
North Macedonia	3	0.24%	Saudi Arabia	1	0.08%
Afghanistan	3	0.24%	El Salvador	1	0.08%
Italy	3	0.24%	Ecuador	1	0.08%
Iran	3	0.24%	Mozambique	1	0.08%
Ghana	3	0.24%	Zimbabwe	1	0.08%
Puerto Rico	3	0.24%	-	-	-

Table 7: Statistics of Continent data.

Continent	N	%
Asia	400	32.47%
North America	306	24.84%
Europe	303	24.59%
Oceania	98	7.95%
Africa	83	6.74%
South America	42	3.41%

Table 8: Statistics of Employment Status data.

Employment Status	N	%
Employed	540	43.83%
Currently a student	267	21.67%
Not employed	184	14.94%
Not employed seeking work	137	11.12%
Prefer not to say	79	6.41%
Retired	13	1.06%
Homemaker	12	0.97%

Table 9: Statistics of Job Category data.

Job Category	N	%
Other Job Category	282	22.89
Engineering	146	11.85%
Banking And Finance	90	7.31%
Management	88	7.14%
Information Technology	86	6.98%
Admin And Clerical	80	6.49%
Sales And Marketing	72	5.84%
Health Care	65	5.28%
Education - Teaching	48	3.90%

This table continues in the next page.

Table 9 (continuation): Statistics of Job Category data.

Job Category	N	%
Retail	40	3.25%
Science And Biotech	39	3.17%
Customer Service	35	2.84%
Human Resources	31	2.52%
Law Enforcement and Legal	30	2.44%
Transportation	30	2.44%
Automotive	24	1.95%
Construction	22	1.79%
Pharmaceutical	12	0.97%
Manufacturing	12	0.97%

Table 10: Statistics of Education Level data.

Education Level	N	%
College/University Degree (4 Years)	381	30.93%
Currently In College/University	189	15.34%
High School Graduate	184	14.94%
Graduate School Degree	154	12.50%
Some College/University but Did Not Graduate	63	5.11%
Prefer not to say	56	4.55%
Other Education	56	4.55%
Less Than 12 Years of Education	46	3.73%
Currently In Graduate School	38	3.08%
Associate Degree (2 Years)	36	2.92%
Doctorate Degree	20	1.62%
Currently Working Towards a Doctorate Degree	9	0.73%

Table 11: Statistics of Academic Major data.

Academic Major	N	%	Academic Major	N	%
Prefer Not To Say	333	27.03%	Telecommunications Engineering	7	0.57%
Computer Science	114	9.25%	Architecture	7	0.57%
No Major	100	8.12%	Economics	7	0.57%
Business Administration	84	6.82%	Law	5	0.41%
Other Major	63	5.11%	Interior Design	5	0.41%
Education	37	3.00%	Health Services Administration	5	0.41%
Mechanical Engineering	32	2.60%	Commerce	5	0.41%
Electrical Engineering	31	2.52%	Physical Therapy	5	0.41%
Finance	31	2.52%	Human Resources Administration	4	0.32%
Arts	30	2.44%	Logistics	4	0.32%
Accounting	30	2.44%	Naval Engineering	4	0.32%
Media Studies	23	1.87%	Chemistry	3	0.24%
Civil Engineering	22	1.79%	Nutrition	3	0.24%
Social Sciences	18	1.46%	Social Work	3	0.24%

This table continues on the next page.

Table 9 (continuation): Statistics of Academic Major data.

Academic Major	N	%	Academic Major	N	%
Languages	17	1.38%	Physics	3	0.24%
Psychology	17	1.38%	Veterinary Medicine	2	0.16%
Industrial Engineering	16	1.30%	Systems Engineering	2	0.16%
Biology	15	1.22%	Botany	2	0.16%
Mathematics	14	1.14%	Bioethics	2	0.16%
Medicine	13	1.06%	Pharmacology	2	0.16%
Chemical Engineering	13	1.06%	Military Science	1	0.08%
Marketing	12	0.97%	Animal Sciences	1	0.08%
Aeronautical Engineering	12	0.97%	European Studies	1	0.08%
Law Enforcement	12	0.97%	Fashion	1	0.08%
Humanities	11	0.89%	Culinary Arts	1	0.08%
Applied Mathematics	9	0.73%	Neuroscience	1	0.08%
Environmental Science	9	0.73%	Medical Assistance	1	0.08%
Aviation	9	0.73%	Astronomy	1	0.08%
Nursing	8	0.65%	Geological Engineering	1	0.08%
Biomedical Engineering	8	0.65%	-	-	-

Table 12: Statistics of Academic Discipline.

Academic Discipline	N	%
Prefer Not to Say	333	27.03%
Business	177	14.37%
Engineering	157	12.74%
Computer Sciences	117	9.50%
Other Academic Discipline	103	8.36%
Social Sciences	60	4.87%
Health	50	4.06%
Arts	40	3.25%
Education	38	3.08%
Natural Sciences	35	2.84%
Mathematics	27	2.19%
Communications	24	1.95%
Public Service	21	1.70%
Language And Literature Studies	17	1.38%
Design	15	1.22%
Aviation	9	0.73%
Cultural Studies	5	0.41%
Agriculture	4	0.32%

2.3.12 Visualizations of the Statistics of the Validation Sample

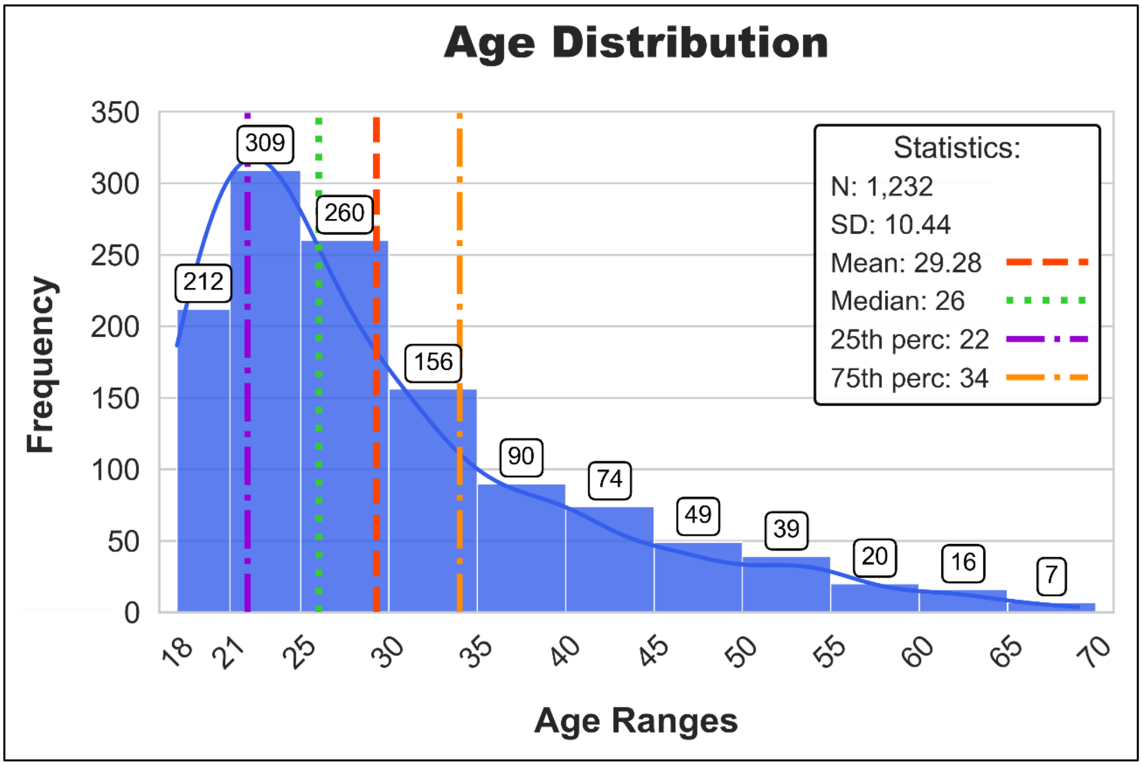

Figure 4: Histogram of Age data

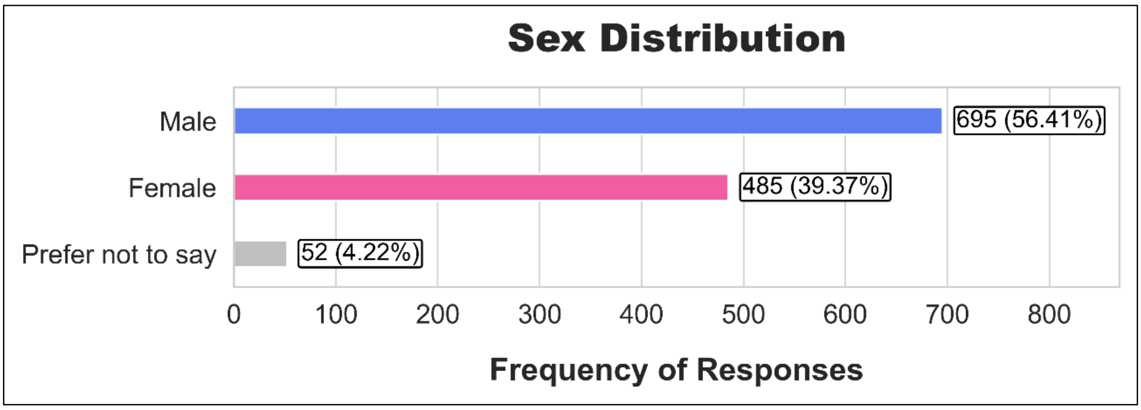

Figure 5: Bar chart of Sex data.

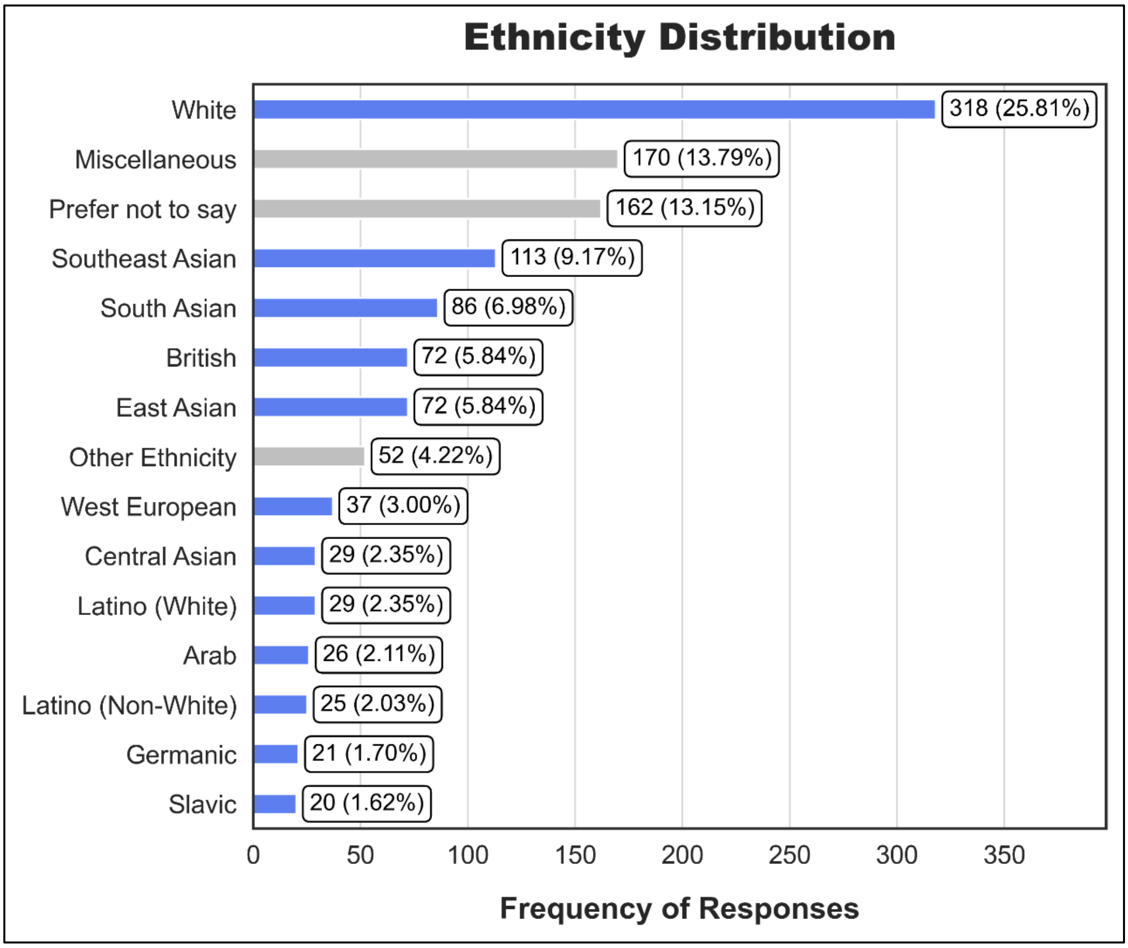

Figure 6: Bar chart of Ethnicity data.

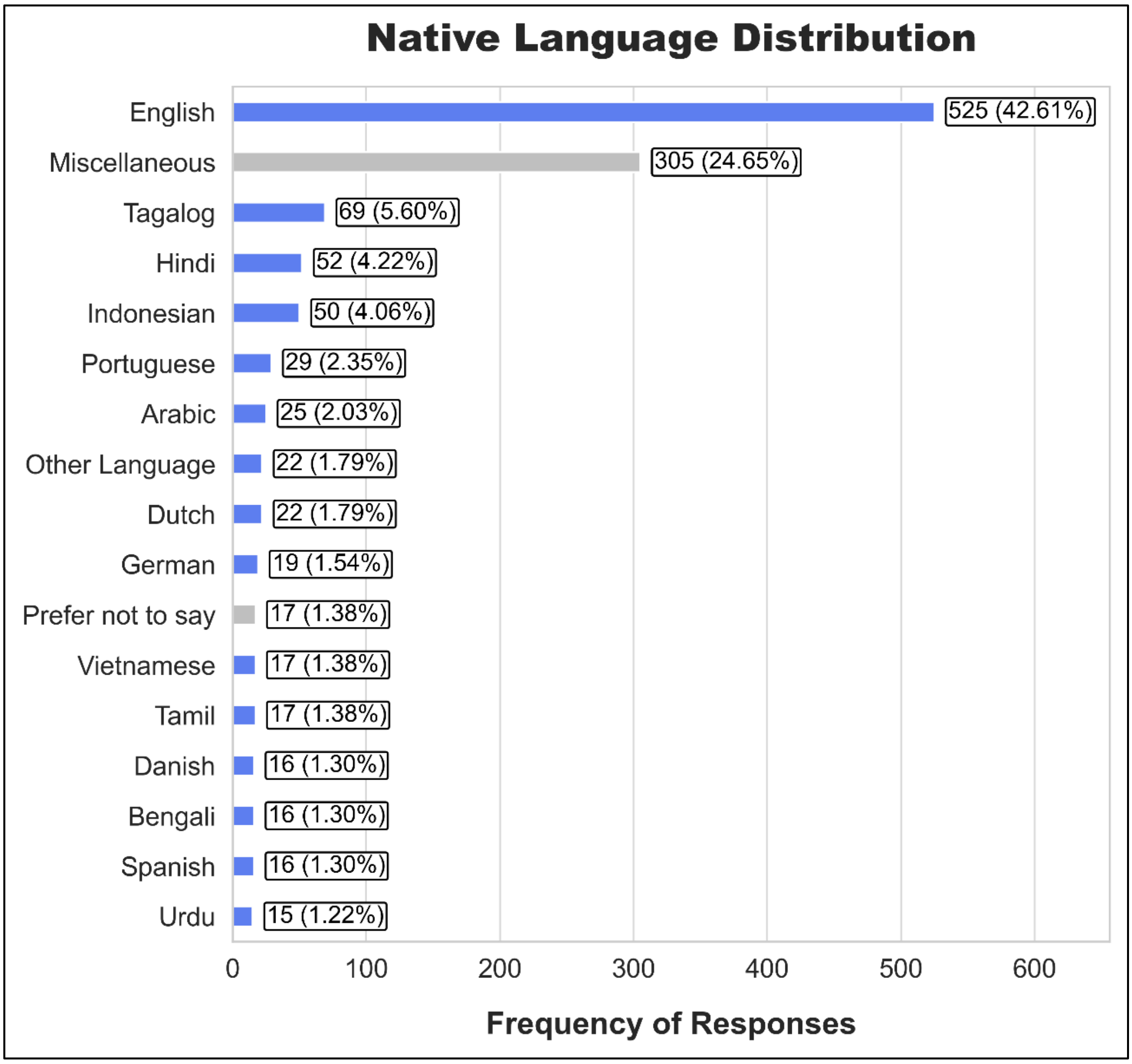

Figure 7: Bar chart of Native Language data.

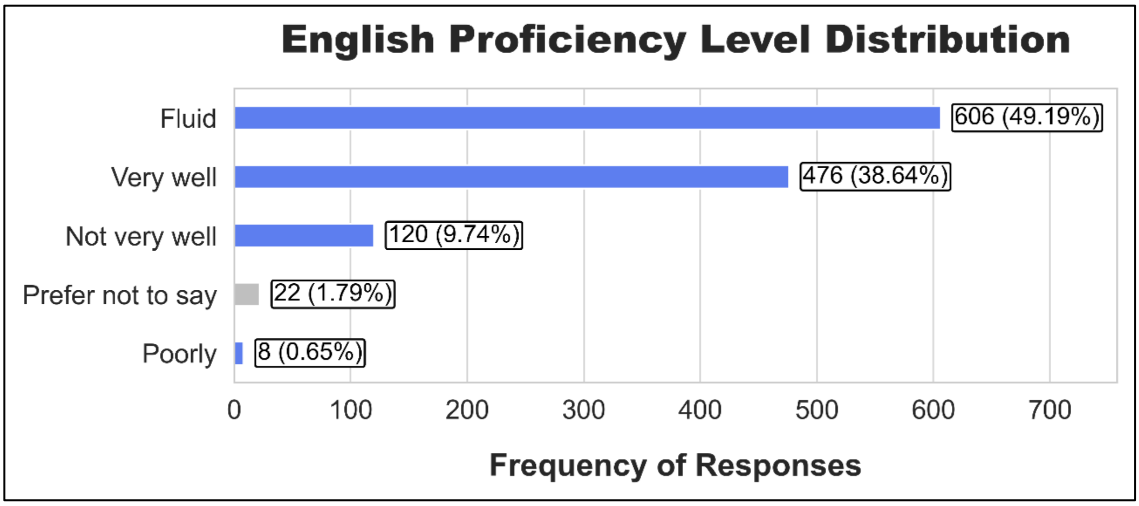

Figure 8: Bar chart of English Proficiency Level data.

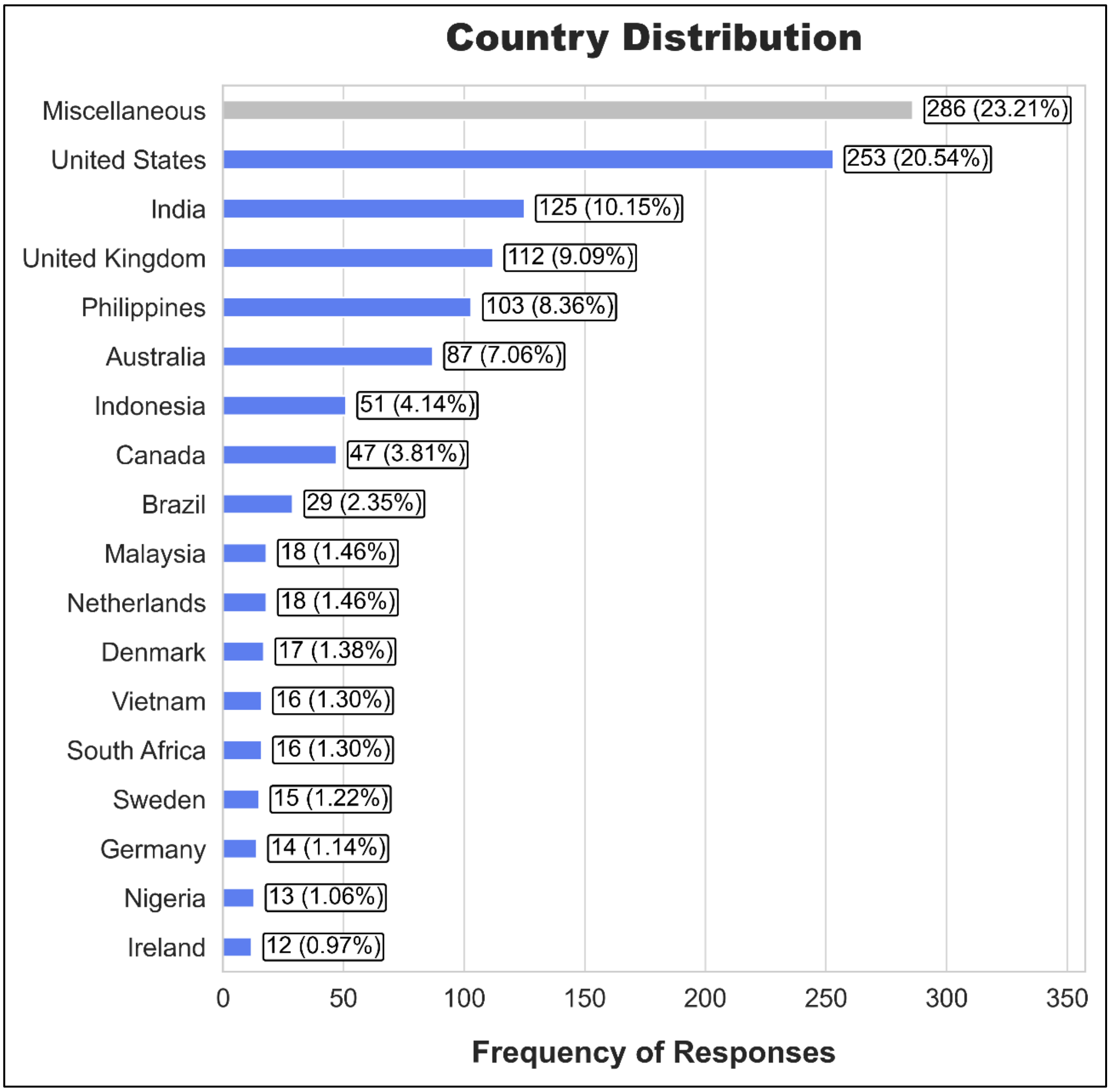

Figure 9: Bar Chart for Country data

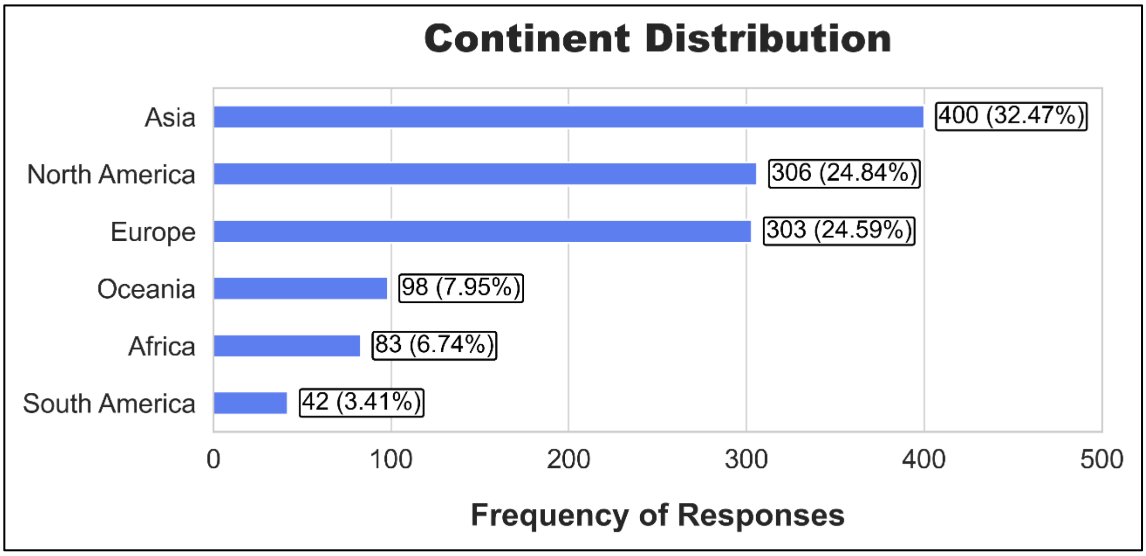

Figure 10: Bar chart of Continent data.

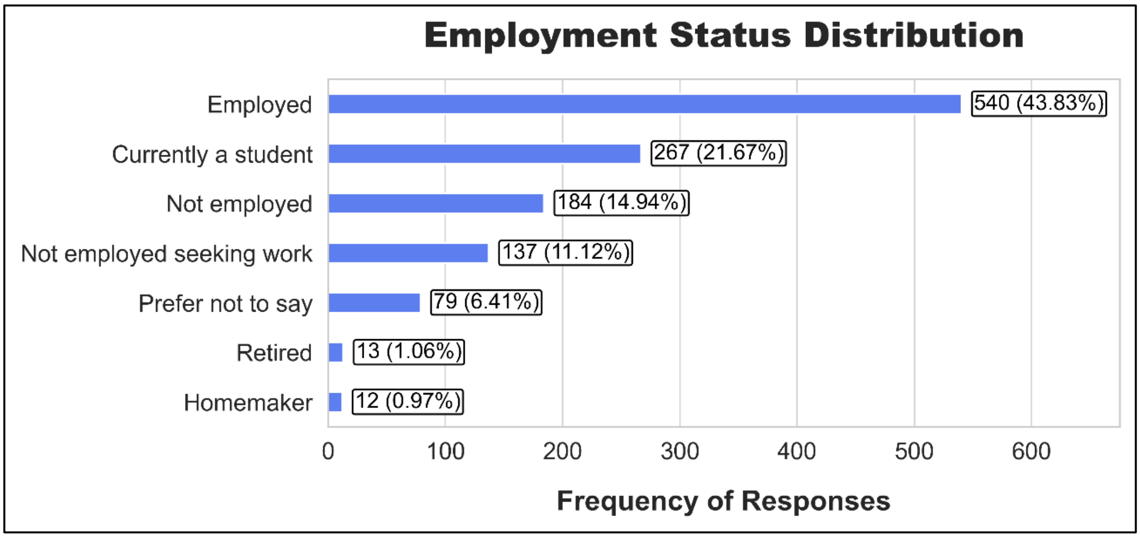

Figure 11: Bar chart of Employment Status data.

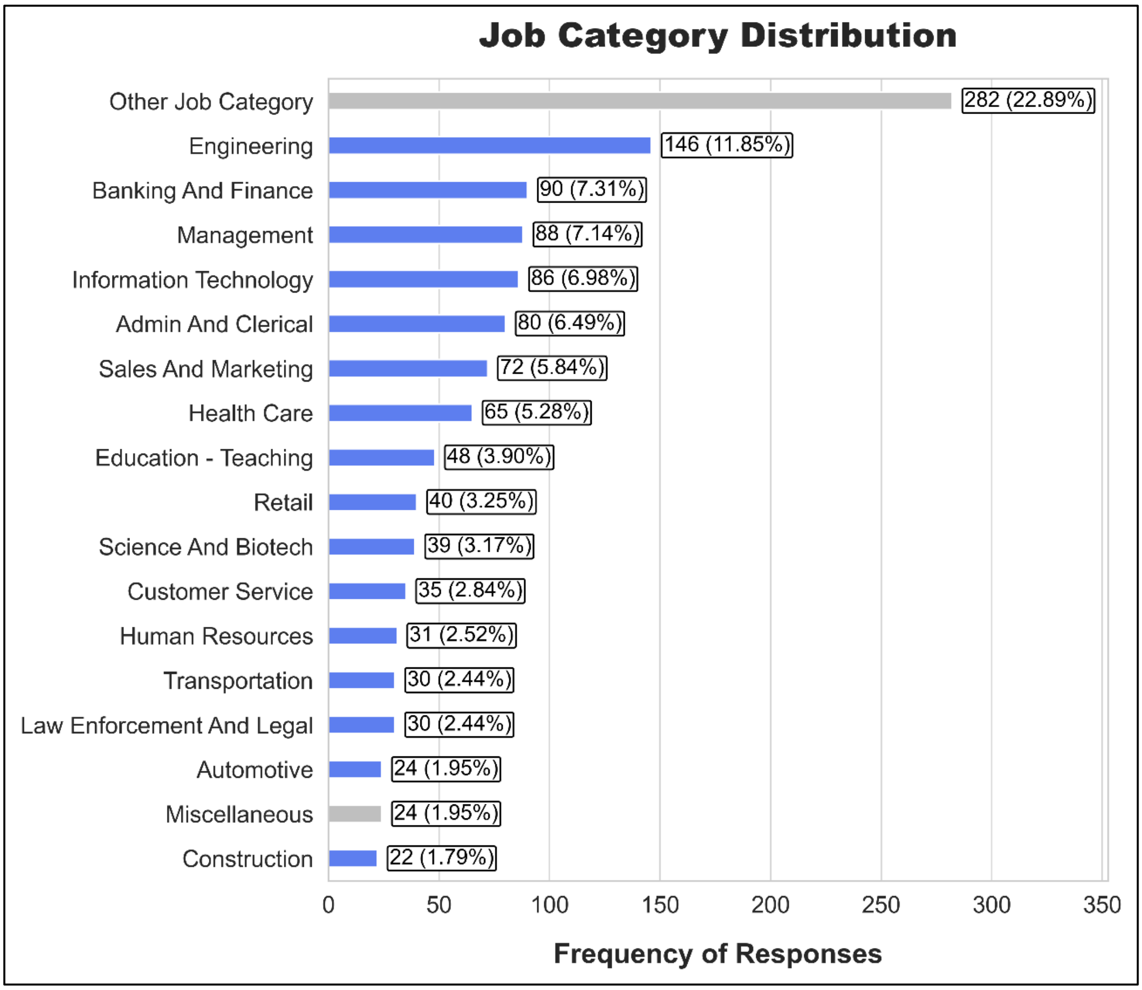

Figure 12: Bar chart of Job Category Status data.

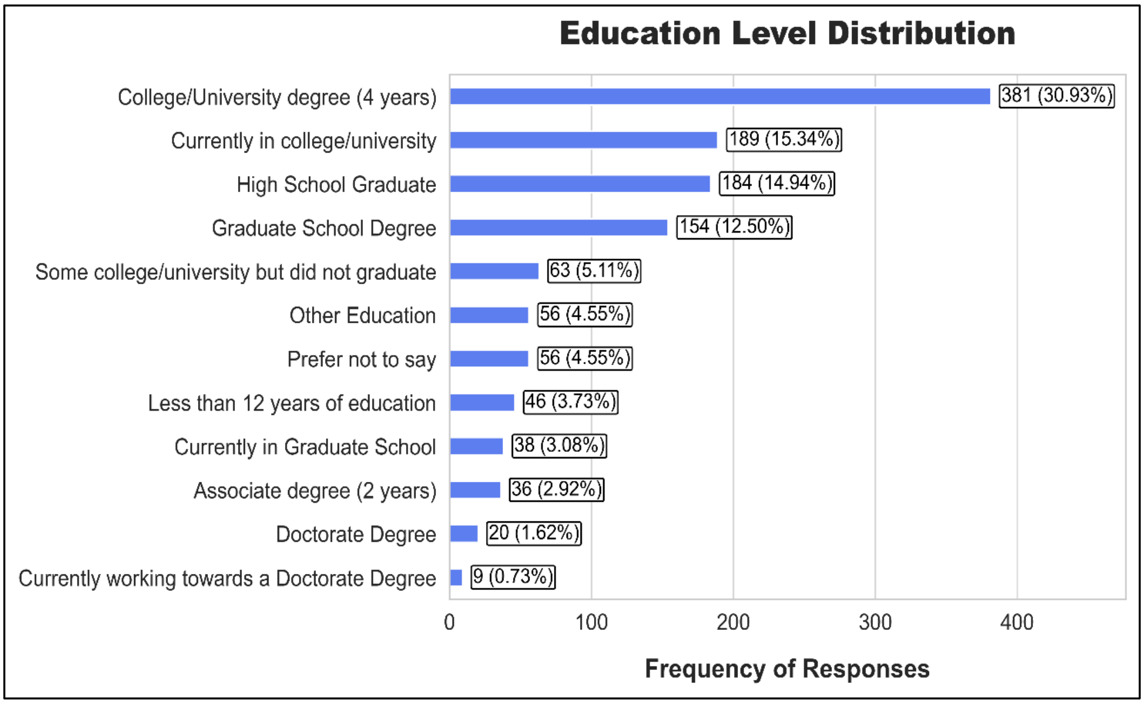

Figure 13: Bar chart of Education Level data.

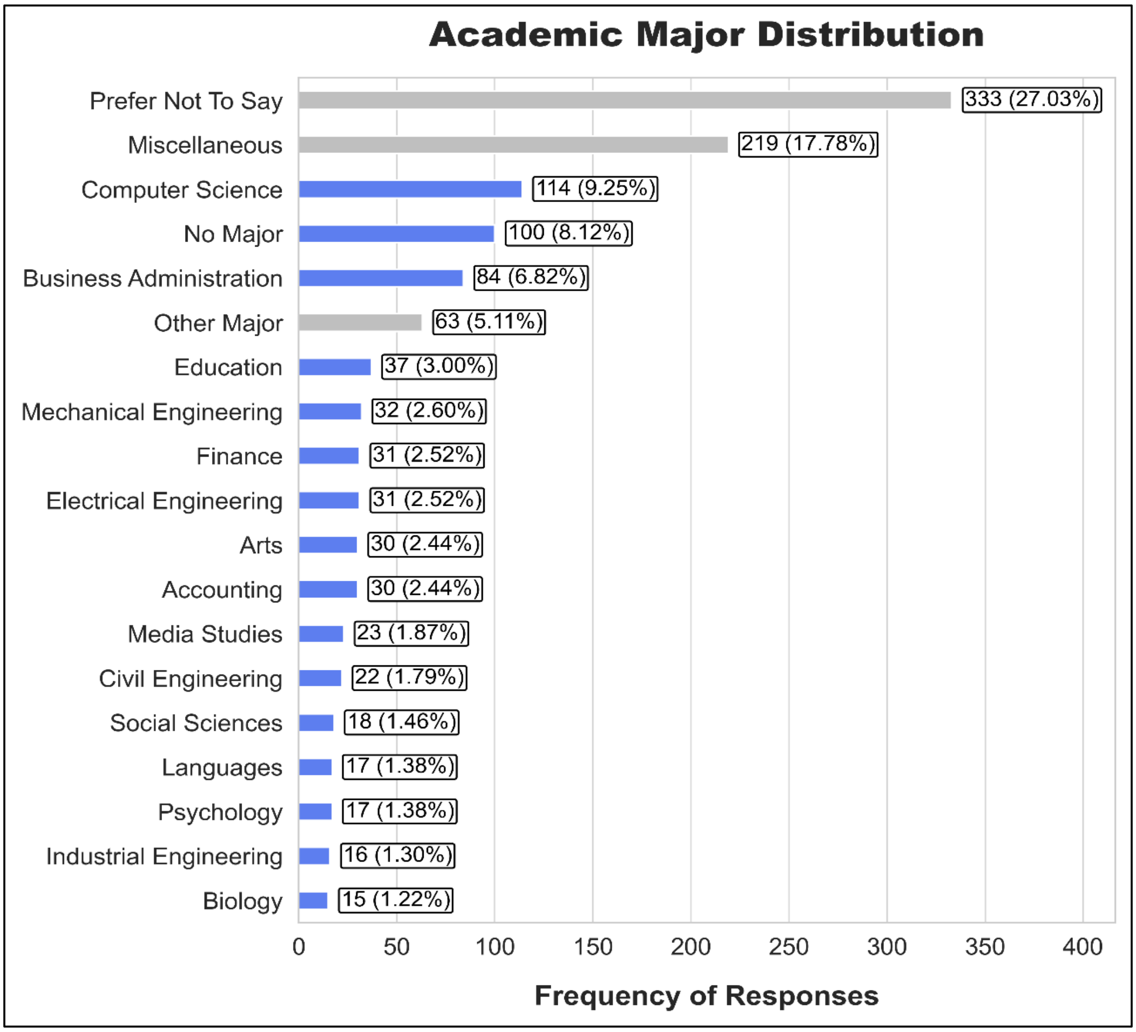

Figure 14: Bar chart of Academic Major data.

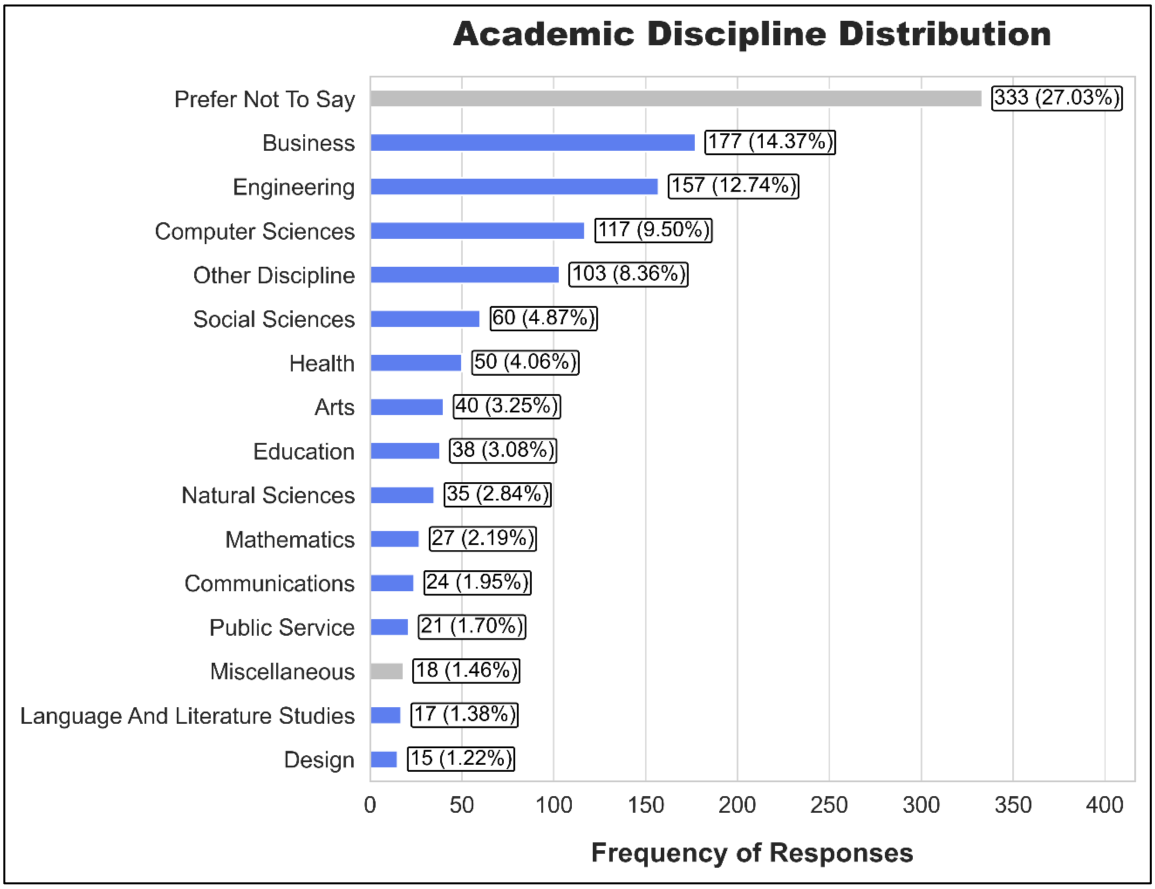

Figure 15: Bar chart of Academic Discipline data.

Psychometric Properties of the G-CAT: Validity, Reliability, and Fairness

This chapter describes the psychometric properties of the G-CAT, including its Validity, Reliability, and Fairness. Validity analyses provide evidence that the scores of the test truly represent the cognitive ability constructs it intends to measure. The reliability analyses ensure the consistency and precision of the test scores, while Fairness analyses indicate that the G-CAT can be applied to diverse global populations of different languages.

3.1 Validity of the G-CAT

The Validity of a test refers to the extent to which the it measures what it intends to measure and the degree to which its scores support meaningful and appropriate interpretations. It is not a single property but a collection of evidence that demonstrates the test's ability to correctly assess cognitive abilities.

Validity is established through multiple approaches. For the G-CAT, the methods for determining validity are Criterion Validity, Structural Validity, Subtest Convergence Validity, and Group Differences Validity.

3.1.1 Criterion Validity

Criterion validity examines the extent to which a test correlates with an external measure that serves as a standard or benchmark. In the context of the G-CAT, criterion validity was evaluated by analyzing the correlations between the test's main scores—General Cognitive Ability (GCA), Spatial, Numeric, and Abstract abilities—and a 9-item version (Bilker et al., 2012) of a widely recognized external criterion: The Raven Progressive Matrices.

By comparing the G-CAT scores to the Raven-9 score, we can determine the extent to which the G-CAT aligns with this established measure, thus providing evidence for its criterion validity.

3.1.1.1 Characteristics of the Raven-9 Scale

The Raven-9 is a shortened version of the Raven's Standard Progressive Matrices (RSPM), developed to assess abstract ability and fluid intelligence. By selecting a representative subset of 9 items from the original 60-item scale, the Raven-9 preserves the psychometric rigor of the full RSPM while offering a time-efficient alternative for cognitive assessment.

Predictive Accuracy

The Raven-9 demonstrates exceptional predictive accuracy for the full RSPM score, with correlations exceeding 0.98. This strong alignment ensures that the shortened version effectively captures the cognitive abilities measured by the full scale.

Internal Consistency

The Raven-9 maintains high reliability indices, with Intraclass Correlation Coefficients (ICC) ranging from 0.882 to 0.921 in validation samples. These values are comparable to those of the full RSPM, confirming that the Raven-9 retains internal consistency despite its reduced length.

Administration Efficiency

Requiring approximately 85% less time to administer compared to the full RSPM, the Raven-9 is ideal for large-scale assessments and time-limited studies, offering significant practical advantages while maintaining robust psychometric properties.

Dimensionality

Similar to the full RSPM, the Raven-9 assesses a unidimensional construct—abstract reasoning ability—making it an appropriate tool for evaluating a core component of fluid intelligence.

Fairness

The nonverbal nature of the Raven-9 reduces cultural and language biases, ensuring its fairness in cross-cultural settings. This makes the test widely applicable across diverse populations.

Breadth of Difficulty

The 9 items in the Raven-9 were carefully selected to provide a representative spread across difficulty levels, preserving the scale's ability to differentiate between individuals with varying levels of cognitive ability.

3.1.1.2 Correlation Analysis between the G-CAT and the Raven-9

To assess criterion validity, Pearson correlations were computed between the Raven 9 total score and the G-CAT scores, including the Spatial, Numeric, Abstract, and General Cognitive Ability scores. The correlation matrix below presents the results:

Table 13: Correlations between the G-CAT and Raven-9 scores.

Test Scores	Correlation Coefficients with the Raven-9
G-CAT General Cognitive Ability	0.620
G-CAT Spatial Ability	0.476
G-CAT Numeric Ability	0.423
G-CAT Abstract Ability	0.588

Note: All correlation coefficients have p-values of *< .001.*

Key Findings

- **General Cognitive Ability**: The strongest correlation was observed between the General Cognitive Ability (GCA) score and the Raven-9 score ($r = 0.620$, $p < .001$). This robust relationship indicates that the GCA score is highly aligned with the external measure of cognitive ability provided by the Raven-9.

- **Abstract Ability**: Among the subtest scores, the Abstract scale score exhibited the highest correlation with the Raven-9 total score ($r = 0.588$, $p < .001$). This finding aligns with the abstract reasoning nature of both measures.
- **Spatial Ability**: The Spatial scale score demonstrated a moderate correlation with the Raven 9 total score ($r = 0.476$, $p < .001$), suggesting that spatial reasoning abilities contribute to overall cognitive performance but may be less aligned with the Raven-9 compared to abstract reasoning.
- **Numeric Ability**: The Numeric scale score showed the lowest correlation with the Raven 9 total score ($r = 0.423$, $p < .001$). This result highlights that while numeric reasoning is a component of general cognitive ability, it may capture distinct aspects not fully assessed by the Raven-9.

The findings demonstrate that the G-CAT is a valid measure of cognitive ability, as evidenced by its strong alignment with the Raven-9. The GCA score, in particular, shows a strong correlation, making it a valid indicator of overall cognitive ability. The subtest scores further provide valuable insights into specific cognitive domains, enhancing the test's utility as a detailed cognitive assessment.

3.1.2 Structural Validity

Structural Validity is usually known as "Construct Validity" in the literature. The name "Structural Validity" is used instead for the following reasons:

- Since all types of validity evidence can be considered as Construct Validity (Slaney, 2017), it is more appropriate to use a different name for the type of validity that is obtained by using a specific approach.
- The analysis method used in this instance is Confirmatory Factor Analysis (CFA), which allows to assess if the structure of the test, represented by a model that groups each test item into a respective subtest, corresponds to the item response patters observed in the data from the validation study.

In that sense, Structural Validity represents the appropriateness of using specific items to measure the desired psychological constructs. This results in evidence for the following assumptions regarding the test:

3.1.2.1 Appropriateness of scoring rules of test

The CFA verifies that each specific item belongs to a specific subtest and not another subtest. In that sense, the items can be grouped into separate categories that represent their underlying psychological constructs. Thus, it can be asserted that each individual item represents their intended psychological constructs and therefore the scores of each individual item can be added together to obtain the scores for each subtest.

3.1.2.2 Replication of the known structure of cognitive abilities

The CFA validates the assumption that the structure of cognitive abilities can be better explained by what is called a Bi-Factor model. In this model, all test items represent a general cognitive ability factor (g), while also representing specific ability factors (e.g., spatial, numeric, and abstract abilities).

This model aligns with the multifaceted nature of cognitive abilities, as research consistently shows that while a general cognitive factor (g) accounts for substantial variance in performance across different tasks, distinct abilities also contribute uniquely (Carroll, 1993). This provides evidence that the test properly measures the theoretical structure or dimensions of cognitive ability it is designed to assess.

To evaluate the structural validity of the G-CAT, various structural models were compared by using CFA on the item data of the validation sample, followed by a model comparison using an ANOVA analysis. This was meant to assess if the data from the validation sample reflects the expected model (Bi-Factor), or other model alternatives.

3.1.2.3 Structural Models Tested

The table below summarizes the key aspects of these models, including their assumptions, scoring rules, and implications:

Table 14: Structural models used in the CFA.

Model	**Assumptions**	**Scoring Rules**	**Implications of Scoring Rules**
Independent Model	Assumes no relationship among items. This model serves as a baseline to compare other models.	Not suitable for scoring rules, as it assumes no relationships among items. Scores would be random or unrelated.	Lacks interpretive value and is not recommended for scoring.
Unidimensional Model	Assumes all items measure a single general factor, such as overall cognitive ability.	Use a single total score by summing all item responses, reflecting a single general factor (e.g., general cognitive ability).	Indicates that all the item scores should be used for a single GCA score only.
Correlated Factors Model	Assumes multiple factors (e.g., spatial, numeric, and abstract abilities) are distinct but related.	Compute separate scores for each factor (e.g., spatial, numeric, abstract abilities) and report them independently. No GCA score can be derived.	Indicates that each subtest item can be used for their respective subtest score but they can't be used to calculate a GCA score.
Bifactor Model	Assumes that a general factor influences all items while specific factors influence subsets of items.	Compute a general factor score (general cognitive ability) and separate specific factor scores for each domain (e.g., spatial, numeric, abstract).	Offers the most nuanced scoring, separating general ability from domain-specific abilities, making it ideal for detailed diagnostics.

These models were compared by using CFA techniques to verify how well each model fits with the data from the validation sample.

3.1.2.4 Confirmatory Factor Analysis (CFA)

Confirmatory Factor Analysis (CFA) is a statistical technique used to test whether the data fits a hypothesized measurement model. In CFA, researchers specify the expected relationships among variables, often based on theory, and then test how well the observed data conforms to these expectations.

Characteristics of the CFA

Data Type: The test items are treated as ordered categorical data to account for the ordinal nature of the spatial item scores and the dichotomous nature of the numeric and abstract item scores.

Fit Indices Reported: Fit indices are numerical summaries that assess how well a statistical model represents the observed data from the validation sample. They help determine whether the relationships among items align with the proposed theoretical structure. The fit indices reported in this chapter include:

- **Chi-Square (χ^2)**: Assesses the difference between observed and predicted covariance matrices. Smaller values indicate a better fit.
- **CFI (Comparative Fit Index)**: Evaluates model fit relative to a null model, with values above 0.95 indicating excellent fit.
- **TLI (Tucker-Lewis Index)**: Indicates incremental model improvement, with values above 0.95 signifying good fit.
- **RMSEA (Root Mean Square Error of Approximation)**: Measures the error of approximation in the population, with values below 0.06 considered acceptable.
- **SRMR (Standardized Root Mean Square Residual)**: Evaluates the average discrepancy between observed and predicted correlations, with values below 0.08 indicating good fit.

Estimator: The estimator is a statistical method used to approximate model parameters by minimizing the differences between the observed and predicted data. In this analysis, we used the Robust Diagonal Weighted Least Squares (DWLS) estimator, which is suitable for handling ordinal data and non-normal distributions.

Scaling Correction Factor: Adjustments were applied to chi-square statistics to correct for non-normality, enhancing the robustness of model fit assessments. In cases where data exhibit skewness or kurtosis, traditional chi-square values can be inflated, leading to incorrect conclusions about model adequacy. To address this, a scaling correction factor—such as the Satorra-Bentler correction or similar robust estimators—was implemented. This correction adjusts the chi-square statistic and its associated p-values, ensuring more reliable interpretations of model fit by accounting for deviations from multivariate normality. As a result, the likelihood of Type I errors (false positives) is reduced, providing a more accurate evaluation of the model's performance.

3.1.2.5 Comparison of structural models

The table below summarizes the fit indices for the five models.

Table 15: Fit indices of the models used in the CFA.

Index of Fit	Independent Model	Unidimensional Model	Correlated Factors Model	Bifactor Model
Chi-Square (χ^2)	11,435.94	2,198.62	888.534	1,318.91
Degrees of Freedom (df)	405	405	402	375
P-Value (χ^2)	< 0.001	< 0.001	< 0.001	< 0.001
Scaling Correction Factor	1.712	1.079	1.082	0.935
CFI	0.44	0.91	0.968	0.971
TLI	0.374	0.902	0.963	0.966
RMSEA	0.184	0.08	0.045	0.045
90% CI RMSEA (Lower, Upper)	(0.182, 0.186)	(0.076, 0.084)	(0.043, 0.048)	(0.043, 0.048)
SRMR	0.221	0.089	0.056	0.057

3.1.2.6 Key Findings on Structural Models

- The Bifactor Model demonstrates the best overall fit, with the highest CFI and TLI and the lowest RMSEA and SRMR.
- The Independent Model performs poorly across all fit indices, confirming that items are related.
- The Unidimensional Model, while reasonable, fits less well than the Bifactor Model, indicating that a single general factor is insufficient to explain item variance.

3.1.2.7 Model Comparisons Using ANOVA

The models were also compared using scaled chi-square difference tests. These tests evaluate whether a more complex model provides a significantly better fit to the data than a simpler nested model. The specific comparisons made are guided by theoretical and statistical considerations:

- **Independent vs. Unidimensional**: This comparison evaluates whether assuming a single underlying factor improves the fit over assuming no relationships among items.
- **Unidimensional vs. Correlated Factors**: This comparison tests whether modeling multiple interrelated factors provides a better fit than a single general factor.
- **Correlated Factors vs. Bifactor**: Assesses whether adding a general factor alongside specific factors leads to a significantly better fit than just interrelated factors.

Table 16: Comparison of the models used in the CFA.

Comparison	Chi-Square Difference	Degrees of Freedom Difference	P-Value
Independent vs. Unidimensional	9,237.30	0	< 0.001
Unidimensional vs. Correlated	405.6	3	< 0.001
Correlated vs. Bifactor	209.6	27	< 0.001

3.1.2.8 Interpretation of ANOVA Results

- **Independent vs. Other Models**: The Independent Model has significantly worse fit than all other models ($p < 0.001$).
- **Unidimensional vs. Correlated Factors**: The Correlated Factors Model fits significantly better than the Unidimensional Model, as indicated by a lower chi-square value and improved fit indices.
- **Bifactor vs. Other Models**: The Bifactor Model demonstrates the best fit overall, with significantly lower chi-square and superior fit indices.

3.1.2.9 Summary of Structural Validity Findings

The analysis demonstrates that the Bifactor Model provides the best representation of the test's structure, confirming its structural validity. This model supports the presence of a general cognitive ability factor alongside domain-specific factors, such as spatial, numeric, and abstract abilities. The superior fit indices and ANOVA results justify the selection of the Bifactor Model as the foundation for scoring and interpreting test results.

The implications for scoring are significant: the Bifactor Model allows for both a comprehensive general cognitive ability score and specific factor scores for each domain. This dual scoring approach enables a more nuanced understanding of an individual's cognitive profile, facilitating targeted interpretations and applications of the test results in occupational settings.

The findings provide a robust basis for using the Bifactor Model to ensure the test scores are meaningful, reliable, and aligned with the theoretical constructs the test is designed to measure.

3.1.3 Subtest Convergence Validity

Subtest convergence validity examines the extent to which the relationships between subtest scores align with theoretical expectations and prior literature. For this test, the convergence between the Spatial subtest score, Numeric subtest score, Abstract subtest score, and the General Cognitive Ability Score was analyzed using a correlation matrix.

This chapter presents evidence supporting the subtest convergence validity of the test by examining the intercorrelations of these scores. The results highlight meaningful relationships among the subtests and confirm that the test measures distinct but related cognitive abilities.

Table 17: Correlation matrix of the G-CAT scores.

Scores	General Cognitive Ability	Spatial Ability	Numeric Ability	Abstract Ability
General Cognitive Ability	—			
Spatial Ability	0.784	—		
Numeric Ability	0.777	0.426	—	
Abstract Ability	0.859	0.539	0.474	—

Note: All correlation coefficients have p-values of < *.001.*

The correlations between the Spatial, Numeric, and Abstract Total Scores demonstrate relationships consistent with theoretical expectations and prior research on the cognitive abilities literature (Carroll, 1993), which suggests that these domains, while distinct, share meaningful relationships due to their shared association with general cognitive ability.

The results indicate that each subtest measures a distinct aspect of cognitive ability while maintaining meaningful relationships with the others. This supports the use of subtest scores for targeted interpretations and domain-specific assessments.

3.1.4 Group Differences Validity

Understanding group differences in the G-CAT scores is crucial for assessing its validity in regard to how cognitive ability manifests itself in different groups of people. By observing differences in test scores between different groups and comparing the observed scores to real-world expectations regarding mental abilities, it can be shown that the G-CAT is able to capture meaningful differences in cognitive abilities that exist between those groups.

This chapter synthesizes findings across four group types: Education Levels, Academic Disciplines, Academic Majors, and Job Categories. By examining how the G-CAT scores vary across these groups, we demonstrate the test's construct validity—the ability of the test scores to represent accurate measurements of cognitive abilities.

3.1.4.1 Cognitive Ability Differences Across Education Levels

People with different education levels tend to exhibit varying levels of cognitive abilities. Generally, individuals with higher levels of education demonstrate stronger performance across cognitive domains. This relationship is rooted in the increasing cognitive demands associated with higher educational attainment, such as abstract reasoning, and problem-solving skills.

Participants from the validation sample were categorized according to their reported education level. The scores were then analyzed to understand trends in cognitive performance across education levels. These levels were ranked based on the type of degree obtained, ranging from "Less than 12 years of education" (rank 1) to "Doctorate Degree" (rank 10) The differences in mean scores for Spatial, Numeric, Abstract, and General Cognitive Ability across education levels are shown in the table below:

Table 18: Mean G-CAT scores of each category in Education Level.

Education Level	Spatial	Numeric	Abstract	General	N	Rank
Less than 12 years of education	6.72	3.63	6.22	16.57	46	1
High School Graduate	7.49	4.08	6.76	18.33	184	2
Some college/university but did not grad	7.59	4.38	6.52	18.49	63	3
Associate degree (2 years)	7	4.19	6.25	17.44	36	4
Currently in college/university	7.69	4.62	7.16	19.47	189	5
College/University degree (4 years)	7.48	4.75	7.35	19.58	381	6
Currently in Graduate School	7.34	4.89	7.68	19.92	38	7
Graduate School Degree	7.49	4.65	7.3	19.44	154	8
Currently working towards Doctorate	7.33	4.56	8.44	20.33	9	9
Doctorate Degree	7.95	5.55	8.1	21.6	20	10

To visually represent these differences, line charts were created to illustrate trends across cognitive domains by education level.

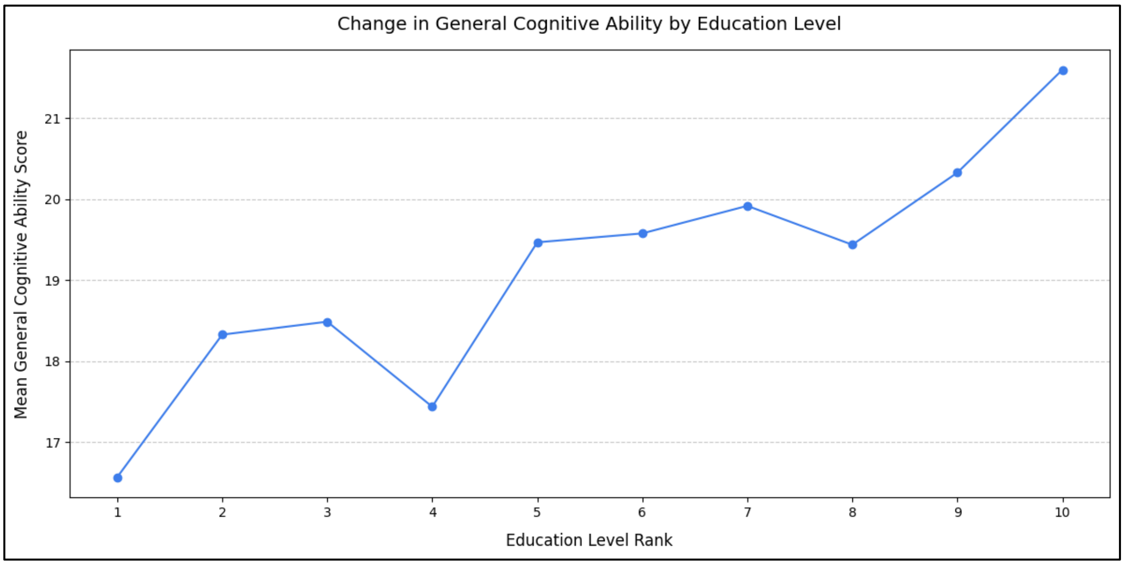

Figure 16: Line chart of mean GCA scores by Education Level.

As seen in the table and line charts, cognitive ability scores increase consistently with higher education levels. Doctorate Degree holders scored the highest (21.60), followed by those pursuing Doctorate Degrees (20.33), while individuals with less than 12 years of education scored the lowest (16.57). These results indicate the test's sensitivity to differences in educational attainment, capturing meaningful variations in cognitive performance.

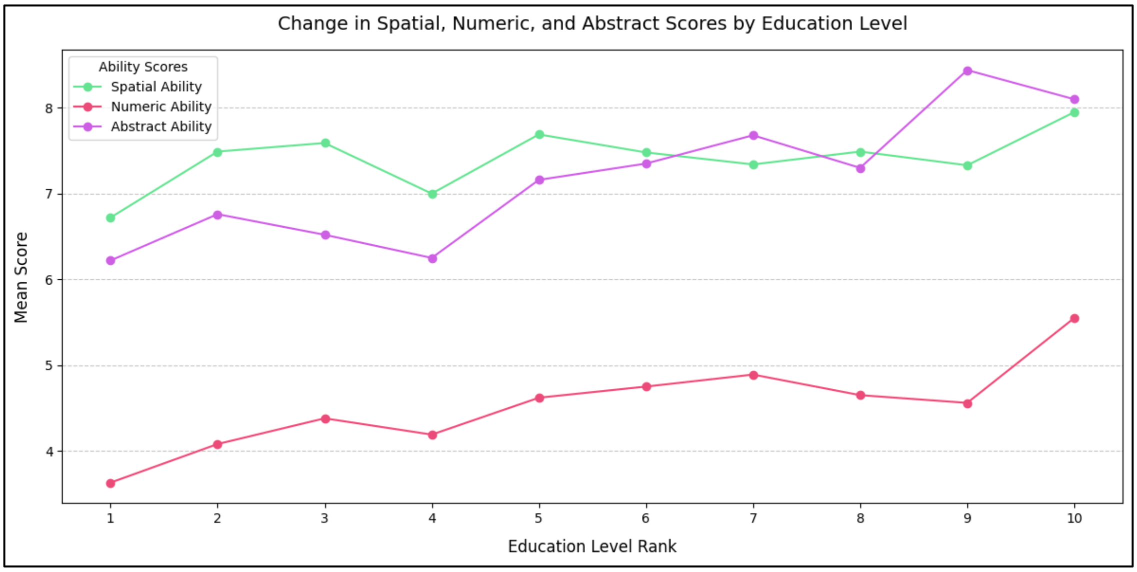

Figure 17: Line chart of mean Spatial, Numeric, and Abstract scores by Education Level.

Additionally, the relationship between education levels and cognitive abilities was analyzed using the Spearman correlation coefficient. This analysis provided insights into the strength of these associations across different cognitive domains.

Table 19: Spearman correlations between Education Level and mean G-CAT scores.

Score	**Spearman Correlation with Education Level**
Mean General Cognitive Ability	0.89***
Mean Spatial Total Score	0.31
Mean Numeric Total Score	0.83***
Mean Abstract Total Score	0.90***

Note: *** indicates correlation coefficients with p-values of < *.001.*

The correlation results confirm a strong positive relationship between education level and cognitive ability, specifically in the domains of General, Numeric, and Abstract Ability. Although the correlation of Spatial ability is weaker, this finding aligns with expectations that higher education levels are associated with greater cognitive ability, further supporting the test's validity.

Key insights of differences among Education Levels

- **General Cognitive Ability**: Doctorate Degree holders (21.60) scored the highest, followed by those pursuing Doctorate Degrees (20.33). Individuals with less than 12 years of education scored the lowest (16.57).

- **Spatial Ability**: Scores are relatively stable across levels, with Doctorate Degree holders (7.95) leading, while those with less than 12 years of education scored 6.72.
- **Numeric Ability**: A strong positive trend is observed, with Doctorate Degree holders scoring 5.55 compared to 3.63 for those with less than 12 years of education.
- **Abstract Ability**: This domain demonstrates the strongest correlation with education level, with Doctorate Degree holders scoring 8.10.

3.1.4.2 Cognitive Ability Differences Across Academic Disciplines

People with different academic disciplines tend to have varying levels of cognitive abilities. Disciplines related to Mathematics, Engineering, and Design typically require advanced abstract reasoning and problem-solving skills, which is reflected in their cognitive ability scores.

Participants from the validation sample were categorized according to their reported academic discipline. These disciplines were ranked based on their General Cognitive Ability scores, and differences in mean scores across Spatial, Numeric, Abstract, and General Cognitive Ability domains were analyzed. These disciplines were ranked based on their General Cognitive Ability scores. Only categories with sample sizes of at least 10 were considered for the analysis. The differences in mean scores for Spatial, Numeric, Abstract, and General Cognitive Ability across academic disciplines are shown in the table below:

Table 20: G-CAT mean scores for each category of Academic Disciplines.

Academic Discipline	Spatial	Numeric	Abstract	General	N	Rank
Education	6.45	3.87	6.03	16.34	38	1
Public service	7.24	3.86	6.86	17.95	21	2
Health	7.20	4.26	6.50	17.96	50	3
Social sciences	7.25	4.17	6.60	18.02	60	4
Communications	7.08	4.50	6.79	18.38	24	5
Business	7.27	4.62	7.08	18.97	177	6
Language and literature studies	7.65	4.35	7.47	19.47	17	7
Arts	8.03	4.15	7.45	19.63	40	8
Computer sciences	7.74	5.04	7.50	20.28	117	9
Natural sciences	7.91	5.00	7.83	20.74	35	10
Engineering	7.93	5.39	8.00	21.32	157	11
Design	8.40	5.20	8.13	21.73	15	12
Mathematics	8.59	5.81	7.89	22.30	27	13

To visually represent these differences, line charts were created to illustrate trends across cognitive domains by academic discipline.

As seen in the table and line charts, disciplines related to Mathematics and Engineering tend to have higher cognitive ability scores, which aligns with the premise that these fields demand higher levels of cognitive ability to address complex abstract tasks. For example, Mathematics exhibited the highest General Cognitive Ability scores (22.30), with strong performance in Spatial (8.59) and Numeric (5.81) abilities, while also having a high score in Abstract ability (7.89). These results demonstrate the G-CAT's ability to capture meaningful cognitive differences across academic disciplines.

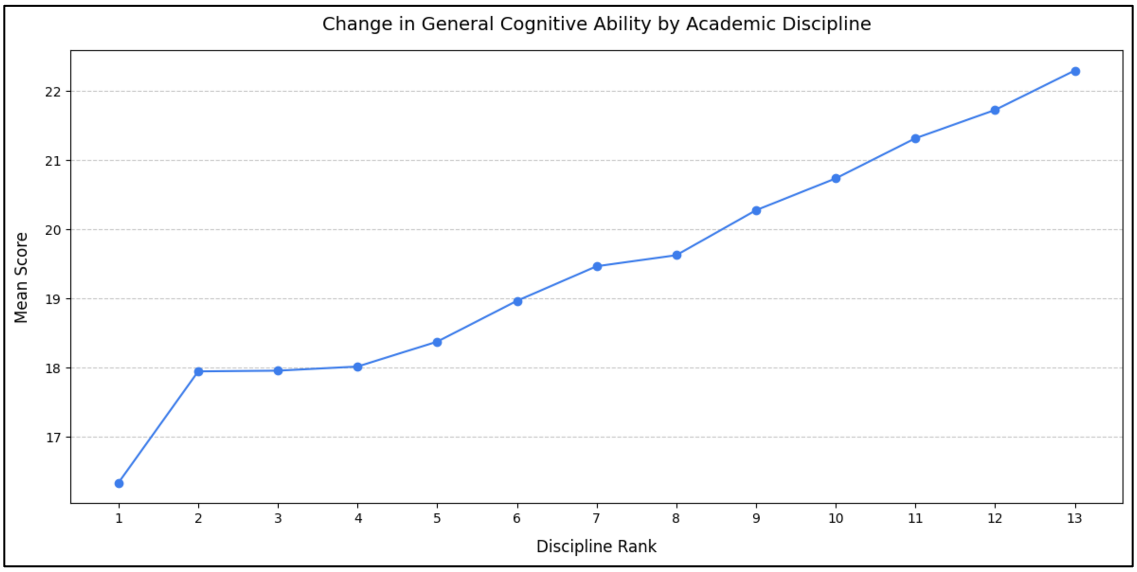

Figure 18: Line chart of mean GCA scores by Academic Discipline.

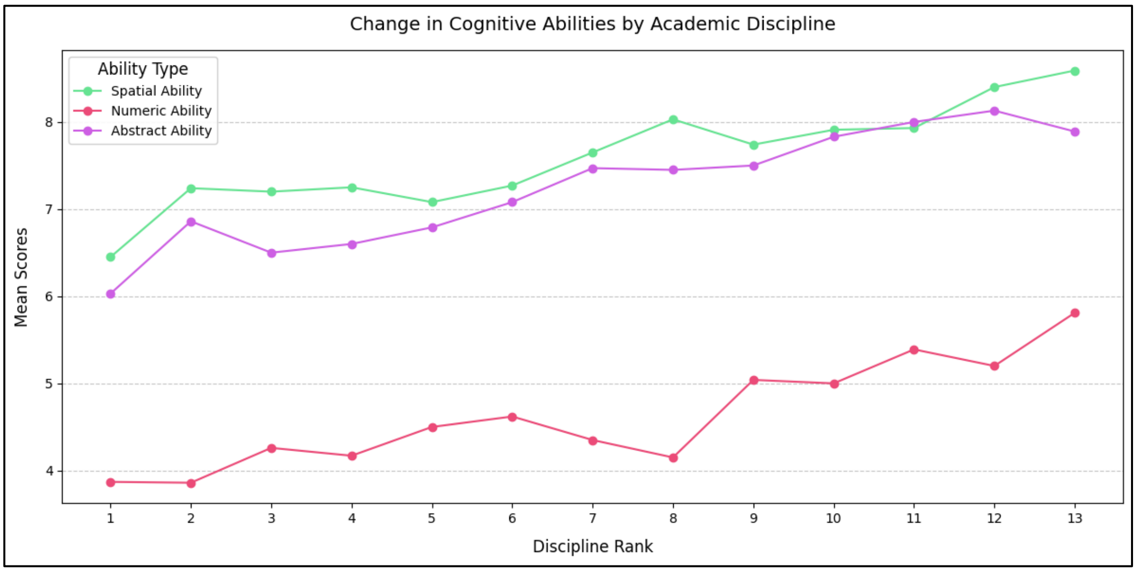

Figure 19: Line chart of mean Spatial, Numeric, and Abstract scores by Academic Disciplines.

Key insights of differences among Academic Disciplines

- **General Cognitive Ability**: Disciplines such as Mathematics (22.30), Design (21.73), and Engineering (21.32) exhibit the highest scores, reflecting the cognitive demands of these fields. Conversely, Education (16.34) and Public Service (17.95) have comparatively lower scores.

- **Spatial Ability**: Mathematics (8.59) and Design (8.40) demonstrate elevated spatial reasoning skills, while Education (6.45) scores the lowest.
- **Numeric Ability**: Mathematics (5.81) and Engineering (5.39) lead in numeric scores, emphasizing the quantitative rigor required in these disciplines.
- **Abstract Ability**: Design (8.13) and Engineering (8.00) rank highest, underscoring the importance of abstract reasoning in these fields.

3.1.4.3 Cognitive Ability Differences Across Academic Majors

People with different academic majors tend to have varying levels of cognitive abilities. Majors related to engineering, mathematics, and science often require advanced analytical and problem-solving skills, which is reflected in their cognitive ability scores.

Participants from the validation sample were categorized according to their reported academic major. These majors were ranked based on their General Cognitive Ability scores, and differences in mean scores across Spatial, Numeric, Abstract, and General Cognitive Ability domains were analyzed. Only categories with sample sizes of at least 10 were considered.

Table 21: G-CAT mean scores for each category of Academic Major.

Academic Major	Spatial	Numeric	Abstract	General	N	Rank
Education	6.46	3.89	6.05	16.41	37	1
Humanities	7.18	3.27	6.09	16.55	11	2
Social Sciences	7	4.61	6	17.61	18	3
Medicine	7	4.54	6.69	18.23	13	4
Media Studies	7.13	4.52	7	18.65	23	5
Other Major	7.29	4.22	7.17	18.68	63	6
Business Administration	7.31	4.48	6.95	18.74	84	7
Marketing	7.42	4.25	7.08	18.75	12	8
Finance	7.19	4.81	7.06	19.06	31	9
Accounting	7.4	4.63	7.13	19.17	30	10
Languages	7.65	4.35	7.47	19.47	17	11
Psychology	7.59	4.53	7.41	19.53	17	12
Law Enforcement	7.92	4.58	7.25	19.75	12	13
Computer Science	7.74	5.07	7.5	20.31	114	14
Chemical Engineering	7.46	5.08	8.08	20.62	13	15
Electrical Engineering	7.58	5.61	7.68	20.87	31	16
Arts	8.4	4.63	7.97	21	30	17
Industrial Engineering	8.56	4.56	8.25	21.38	16	18
Mechanical Engineering	7.53	5.75	8.09	21.38	32	19
Biology	8.4	5.07	7.93	21.4	15	20
Civil Engineering	8.32	5.41	7.91	21.64	22	21
Mathematics	8.5	5.43	7.86	21.79	14	22
Aeronautical Engineering	8.83	5.83	8.17	22.83	12	23

To visually represent these differences, line charts were created to illustrate trends across cognitive domains by academic major.

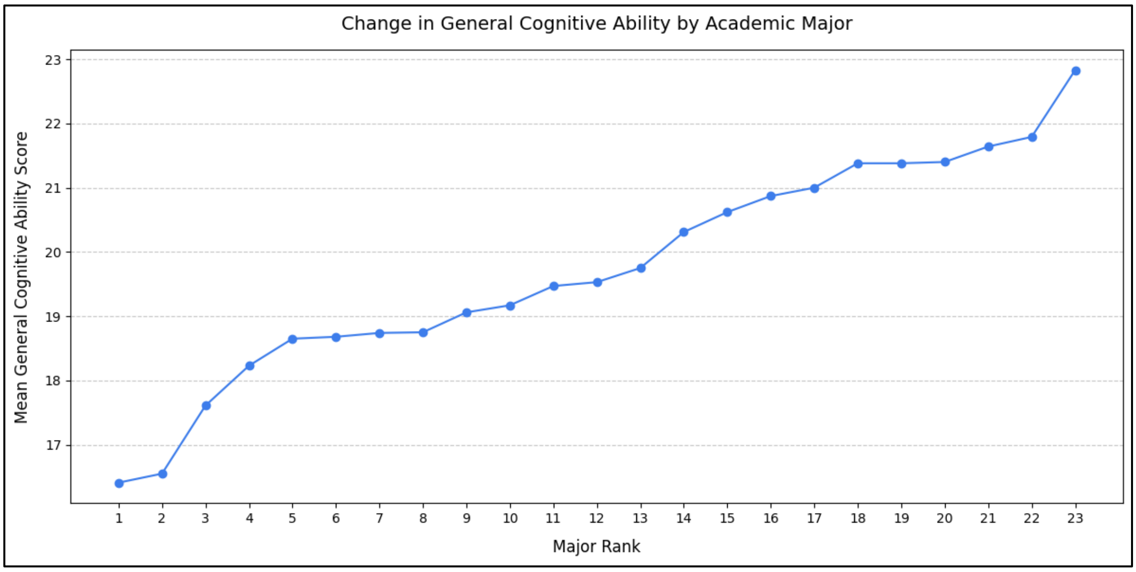

Figure 20: Line chart of mean GCA scores by Academic Major.

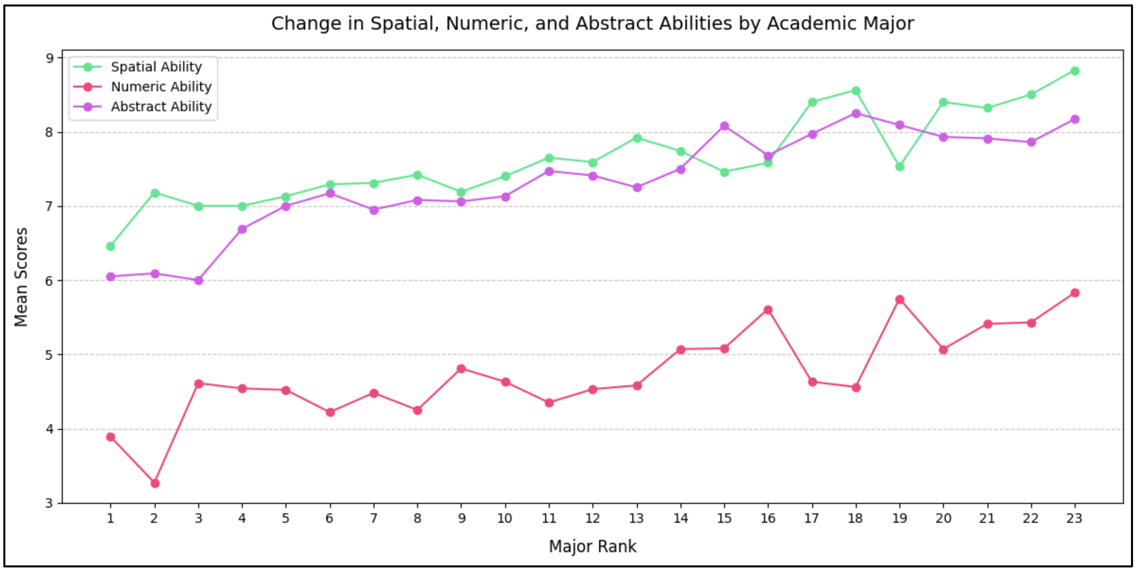

Figure 21: Line chart of mean Spatial, Numeric, and Abstract scores by Academic Major.

As seen in the table and line charts, majors such as Aeronautical Engineering and Mathematics rank the highest in cognitive ability scores, reflecting their advanced cognitive demands. In contrast, Nursing and Information Science rank lower, consistent with their practical and applied focus. These results support the premise that the G-CAT captures meaningful cognitive differences across academic majors.

Key insights of differences among Academic Majors

- **General Cognitive Ability**: Aeronautical Engineering (22.83) and Mathematics (21.79) rank the highest, reflecting advanced cognitive requirements. Education (16.41) scores the lowest.

- **Spatial Ability**: Aeronautical Engineering (8.83) and Industrial Engineering (8.56) demonstrate the highest spatial reasoning skills.
- **Numeric Ability**: Aeronautical Engineering (5.83) and Mechanical Engineering (5.75) lead, consistent with their quantitative emphasis.
- **Abstract Ability**: Industrial Engineering (8.25) and Aeronautical Engineering (8.17) rank highest, highlighting their problem-solving focus.

3.1.4.4 Cognitive Ability Differences Across Job Categories

People in different job categories tend to exhibit varying levels of cognitive abilities. Roles in science, engineering, and technology demand higher levels of cognitive ability, particularly in abstract reasoning and quantitative tasks.

Participants from the validation sample were categorized according to their reported job category. These categories were ranked based on their General Cognitive Ability scores, and differences in mean scores across Spatial, Numeric, Abstract, and General Cognitive Ability domains were analyzed.

Table 22: G-CAT mean scores for each Job Category.

Job Category	Spatial	Numeric	Abstract	General	N	Rank
Customer Service	6.60	4.06	6.03	16.69	35	1
Transportation	7.37	3.90	5.83	17.10	30	2
Human Resources	6.87	4.35	6.35	17.58	31	3
Sales and Marketing	7.22	3.90	6.53	17.65	72	4
Automotive	7.25	3.67	6.75	17.67	24	5
Admin and Clerical	7.05	4.01	6.68	17.74	80	6
Health Care	7.02	4.28	6.54	17.83	65	7
Retail	7.50	4.00	7.12	18.62	40	8
Law Enforcement and Legal	6.93	4.30	7.57	18.8	30	9
Construction	7.86	4.27	7.05	19.18	22	10
Education - Teaching	7.27	4.56	7.44	19.27	48	11
Management	7.44	4.66	7.24	19.34	88	12
Banking and Finance	7.48	4.67	7.36	19.50	90	13
Pharmaceutical	7.92	4.25	7.33	19.50	12	14
Information Technology	7.90	5.10	7.09	20.09	86	15
Manufacturing	8.17	4.33	7.75	20.25	12	16
Engineering	7.86	5.32	7.84	21.02	146	17
Science and Biotech	8.41	5.74	8.15	22.31	39	18

To visually represent these differences, line charts were created to illustrate trends across cognitive domains by job category.

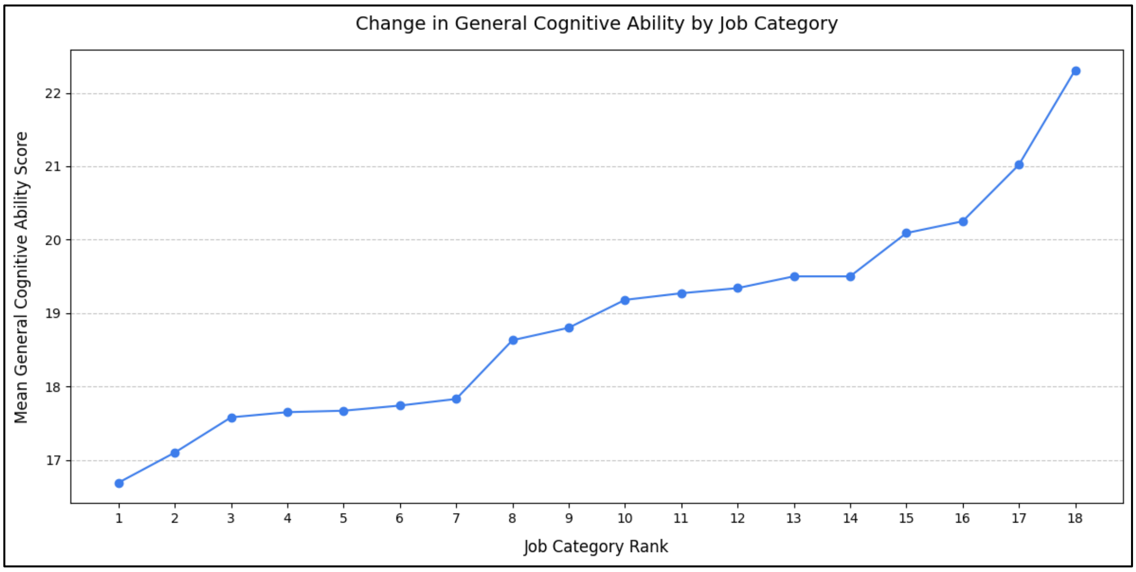

Figure 22: Line chart of mean GCA scores by Job Category.

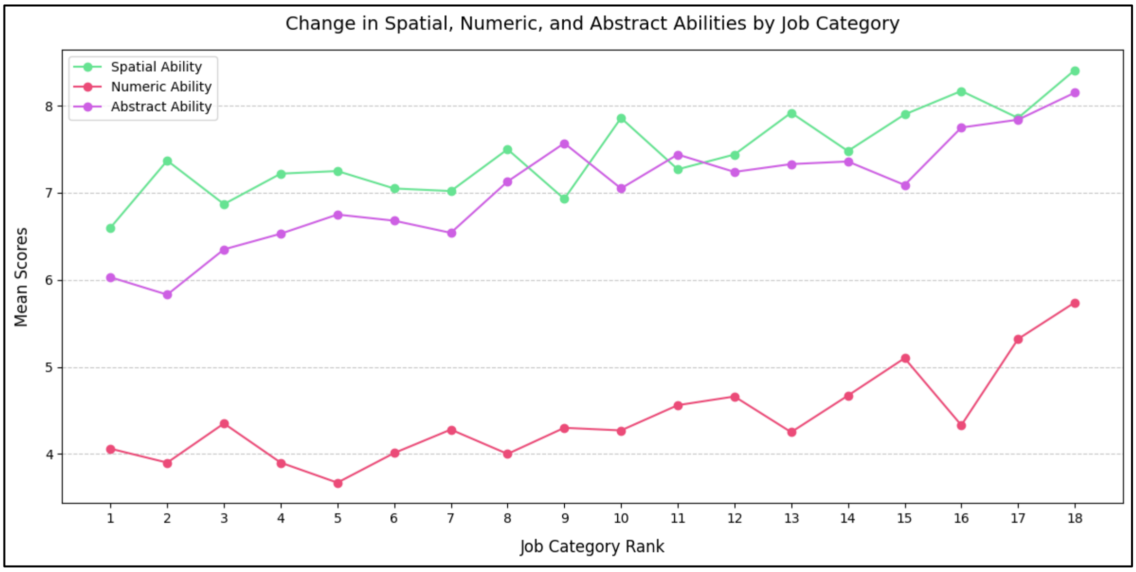

Figure 23: Line chart of mean Spatial, Numeric, and Abstract scores by Job Category.

As seen in the table and line charts, job categories related to Science and Technology demonstrate the highest cognitive ability scores, reflecting the complex problem-solving and technical skills required in these fields. For example, Science and Biotech exhibited the highest General Cognitive Ability scores (22.31), with strong performance in Spatial (8.41) and Numeric (5.74) abilities and the highest score for Abstract Ability (8.15). Conversely, categories such as Customer Service rank lower, consistent with their less cognitively intensive nature.

Key insights of differences among Job Categories

- **General Cognitive Ability**: Science and Biotech (22.31) and Engineering (21.02) rank the highest, indicating high cognitive demands. Customer Service (16.69) and Transportation (17.10) rank lowest.
- **Spatial Ability**: Science and Biotech (8.41) and Manufacturing (8.17) lead, reflecting spatial reasoning's importance in these fields.
- **Numeric Ability**: Science and Biotech (5.74) and Engineering (5.32) stand out, emphasizing their quantitative demands.
- **Abstract Ability**: The highest score is in Science and Biotech (8.15), aligning with the need for problem-solving in these roles.

3.1.5 Summary of Validity Analysis

The G-CAT demonstrates strong psychometric validity from multiple lines of evidence. First, its **Criterion Validity** is supported by significant correlations with the 9-item version of the Raven's Progressive Matrices (Raven-9), indicating that the G-CAT's General Cognitive Ability (GCA) score aligns well with an established measure of cognitive ability. **Structural Validity** findings from Confirmatory Factor Analysis (CFA) show that a **Bifactor Model**—reflecting both a common general factor (g) and specific subtest factors (Spatial, Numeric, Abstract)—fits the data best. Furthermore, **Subtest Convergence Validity** confirms that these domain-specific scores, while distinct, are meaningfully interrelated and correlate strongly with the overall GCA score.

Finally, **Group Differences Validity** analyses reveal that the G-CAT effectively captures known variations in cognitive ability across key demographic and professional categories. Higher educational attainment, mathematically or scientifically intensive academic majors, and specialized job fields (e.g., Engineering, Biotech) consistently show elevated test scores, aligning with well-documented expectations in cognitive research. These convergent findings underscore the G-CAT's overall robustness and support its use as a reliable, theoretically grounded measure of cognitive ability.

3.2 Reliability of the G-CAT

Reliability refers to the consistency and stability of the scores of a psychological test, such as G-CAT, in producing consistent results over time or across different populations. In the context of psychometric assessments, reliability is essential because it ensures that the test accurately measures the constructs it is intended to assess, free from random error as much as possible. A reliable test provides stable and dependable scores, reflecting the true cognitive abilities of test-takers rather than fluctuations due to external factors. The reliability of the G-CAT was assessed by using Internal Consistency methods, which were used to obtain the Standard Error of Measurement of the scores and their Confidence Intervals.

3.2.1 Internal Consistency Estimates of the G-CAT

For the G-CAT, internal consistency reliability was assessed using Cronbach's alpha (α), and McDonald's omega total (ω^T) coefficients. The omega coefficient was used to determine the Standard Error of Measurement (SEM) of the scores, while the SEM were used to calculate their Confidence Intervals (CI). These metrics provide insight into the precision and dependability of the scores for each subtest and the overall test.

- **Coefficient Alpha (α)**: Measures the internal consistency of test items. Values ≥ 0.70 are generally considered acceptable, while values ≥ 0.80 are preferred. This is a classical method used in test development. Although the Omega Total coefficient is superior, coefficient Alpha was included due to its conventional use in test development.
- **Coefficient Omega Total (ω^T):** Coefficient Omega Total is an advanced reliability estimate that accounts for both the general factor and specific factors within a test, providing a more accurate assessment of internal consistency than traditional methods like Cronbach's alpha. By considering the hierarchical structure of test items, ω^T offers a nuanced measure of how well the test captures the intended constructs. In psychometric evaluations, ω^T values of 0.70 or higher are deemed acceptable, while values of 0.80 or above are considered excellent, indicating a high degree of reliability and consistency in test scores

Table 23 presents key reliability statistics for the Global Cognitive Ability Test (G-CAT) and its subcomponents (Spatial, Numeric, and Abstract Ability) along with the General Cognitive Ability Score. The table includes measures of internal consistency (α and ω^T), Standard Errors of Measurement (SEM), score ranges of 95% and 99% Confidence Intervals, and Confidence Intervals of specific score ranges: Ranges of ±2 for subtest scores, and ranges of ±3 and ±4 CI for the GCA score.

Table 23: Reliability metrics of the G-CAT.

Score	α	ω^T	SEM	95% CI	99% CI	±2 CI	±3 CI	±4 CI
Spatial Ability	0.78	0.82	0.7	1.38	1.82	99.54%	N/A	N/A
Numeric Ability	0.7	0.76	0.9	1.82	2.39	96.90%	N/A	N/A
Abstract Ability	0.76	0.8	1	1.99	2.61	95.10%	N/A	N/A
General Cognitive Ability	0.86	0.88	1.7	3.23	4.25	N/A	93.10%	98.48%

- Cronbach's alpha (α) and Omega Total (ω^T) measure the internal consistency of test scores.
- All subtests have α ≥ 0.70 and ω^T ≥ 0.76, indicating acceptable to strong reliability.
- General Cognitive Ability (α = 0.86, ω^T = 0.88) has the highest reliability, suggesting strong score consistency.
- Omega Total (ω^T) is consistently higher than α, indicating that the test structure accounts for multidimensionality better than traditional alpha.

3.2.2 Standard Error of Measurement of the G-CAT Scores

SEM quantifies the expected variation in observed scores due to measurement error. Lower SEM values indicate greater precision. The SEM of the G-CAT scores are as follow:

- Spatial Ability (SEM = 0.7) has the lowest measurement error, meaning its scores are the most precise among the subtests.
- Numeric Ability (SEM = 0.9) and Abstract Ability (SEM = 1.0) have slightly higher SEM values, suggesting slightly more variability in their scores compared to the Spatial Ability subtest.
- General Cognitive Ability (SEM = 1.7) has the highest SEM due to its wider score range (0–30), meaning individual scores may fluctuate more due to measurement error.

3.2.3 Score ranges of the 95% and 99% Confidence Intervals

Confidence intervals (CI) are a range of values used to estimate the true value of a population parameter, such as a mean or proportion, based on sample data. They provide a level of certainty, usually expressed as a percentage (e.g., 95%), indicating that the true value is likely to fall within the specified range. Wider intervals suggest more uncertainty, while narrower intervals indicate greater precision in the estimate.

It is an industry standard to report the expected score ranges of the 95% and 99% Confidence Intervals of the scores of a test. Those CIs represent the estimated range within which a test-taker's true score is expected to fall with 95% and 99% confidence, respectively.

In that sense, considering CIs of 95%, for the Spatial Ability subtest, the true score of the test-taker is expected to be equal to the reported scored, plus or minus 1.38. For the Numeric Ability subtest, the true score is expected to be equal to the reported score plus or minus 1.82. For the Abstract Ability subtest, the variations are plus or minus 1.99, and for the General Cognitive Ability score, the variations are plus or minus 3.23.

For the CIs of 99%, for the Spatial Ability subtest, the variations are plus or minus 1.82. For the Numeric Ability subtest, they are plus or minus 2.39. For the Abstract Ability subtest, they are plus or minus 2.61. And for the GCA score, they are plus or minus 4.25.

As it can be observed, score ranges for the 99% CI are larger than those in the 95% CI. That is because larger CIs require a wider margin of error to ensure that we have greater certainty that the interval contains the true score. In other words, moving from a 95% to a 99% CI means we want to be even more confident in our estimate; to achieve this, the band around our estimated score must be expanded, resulting in larger plus-or-minus values. It should also be noted that GCA score variations are larger than the subtest scores due to the GCA score range being 0-30 while the ranges of the subtest scores are 0-10.

While it is useful to know how much score variation can be expected in regard to specific CIs, these variations are often expressed in decimal-point precision rather than discrete scores. Therefore, it may be more practical to use specific discrete score ranges, acknowledging that the confidence intervals might differ from the standard 95% or 99%. This approach helps prevent over-interpretation of minor decimal fluctuations and aligns more closely with how test results are typically reported in real-world contexts.

3.2.4 Confidence Intervals of Specific Score Ranges

In this section, we present confidence intervals based on fixed score ranges, which can be more intuitive and practical than standard 95% or 99% confidence intervals. By focusing on discrete score bands (e.g., ±2 points), we gain a clearer understanding of how confident we can be that a test-taker's true score lies within those specific ranges. As can be observed in table 23, the CIs for the ranges of the scores of the G-CAT are:

- For the subtests (0–10 scale), ranges of ±2 points are used. Their CIs are:
    - 99.54% confidence for Spatial Ability.
    - 96.90% confidence for Numeric Ability.
    - 95.10% confidence for Abstract Ability.

This means that if we assume the true score is within ±2 points of the reported score, we can be highly confident in its accuracy.

- **For General Cognitive Ability (0–30 scale),** ranges of ±3 and ±4 and points are used, Their CIs are**:**
    - For the ±3 score range, the CI is 93.1%.
    - For the ±4 score range, the CI is 98.5%.

This means that if we assume the true score is within ±3 points, we have 93.1% confidence in its accuracy, and if we assume the true score is within ±4 points, we have nearly 98.5% confidence in its accuracy.

3.2.5 Summary of Reliability Analysis

The G-CAT demonstrates solid reliability, as evidenced by acceptable-to-strong internal consistency estimates and precise score measurements. Cronbach's alpha (α) and McDonald's omega total (ω^T) coefficients for all subtests (Spatial, Numeric, Abstract) and the overall General Cognitive Ability (GCA) score exceed 0.70, confirming stable and dependable scores. The GCA score, in particular, reaches alpha and omega values above 0.85, indicating the highest degree of consistency among all sections. Notably, omega coefficients consistently outpace the corresponding alpha values, reflecting the multifaceted nature of the G-CAT and underscoring the importance of using more advanced reliability indices that account for both general and specific factors.

Alongside these strong internal consistency metrics, the G-CAT exhibits low Standard Errors of Measurement (SEM), meaning relatively small score fluctuations due to measurement error. This precision supports the reliable interpretation of subtest and overall scores. Confidence Intervals (CIs) based on standard 95% or 99% thresholds—and on specific discrete ranges—further illustrate the test's robust measurement properties. Although the GCA scale (0–30) produces a slightly larger SEM and broader CIs than the 0–10 subtests, this merely reflects its wider possible score range. Taken together, these reliability indicators provide firm evidence that the G-CAT delivers stable scores, making it a dependable tool for assessing cognitive ability across various contexts.

3.3 Fairness of the G-CAT

Fairness in cognitive ability testing ensures that scores accurately reflect test-takers' true abilities, free from bias or undue advantage based on demographic or contextual factors. In the context of the G-CAT, fairness is a critical consideration, as the test is designed for diverse populations worldwide.

To evaluate the fairness of the G-CAT, Effect Size analyses and Correlational Analyses were conducted to assess whether the test scores differ significantly between test-takers whose primary language is English and those whose primary language is not English.

For these analyses, the four scores of the test (spatial, numeric, abstract, and general ability) were considered. Additionally, the scores of the Raven-9 were included to assess how that criterion test varies between English and non-English groups and compare those variations with the variations of the scores of the G-CAT.

3.3.1 Effect Size Analyses

Effect size metrics were used to quantify the magnitude of differences in scores between the English-Speaking and Non-English-Speaking groups to provide evidence regarding the test's fairness in regard to the native language of the test-takers.

The validation sample was divided into two groups based on the native language variable:

- **English**: Test-takers whose primary language is English (n=525).
- **Non-English**: Test-takers whose primary language is not English (n=707).

The score differences between groups were analyzed by using the Eta Squared and Cohen's d methods. Results of these analyses are shown in Table 24.

Table 24: Effect sizes of G-CAT scores between Language groups.

Score	Eta Squared	Cohen's d
Spatial	0.0014	0.0751
Numeric	0.0058	-0.1544
Abstract	0.0086	-0.1877
Raven	0.0044	-0.1341
General	0.0038	-0.1241

3.3.1.1 Eta Squared Analysis

This metric represents the proportion of variance in test scores that is explained by the difference between groups (e.g., English and Non-English). Eta Squared values range from 0 to 1, where:

- Values closer to 0 indicate that group membership explains very little of the variance.
- Values closer to 1 indicate that group membership explains a substantial proportion of the variance.
- In psychometrics, values below 0.01 are typically considered negligible.

In this instance, all values are below 0.01, indicating that less than 1% of the variance in test scores is explained by language group differences. This suggests that language differences have minimal impact on performance variance.

3.3.1.2 Cohen's d Analysis

This metric measures the standardized mean difference between two groups. It quantifies the size of the effect in terms of standard deviations and is useful for understanding the practical significance of differences. In this context, Cohen's d helps determine whether observed differences in test scores are meaningful or practically insignificant. The interpretations of different Cohen’s d metrics can be observed in table 25.

Table 25: Interpretation of Cohen’s d metrics.

Cohen’s d	Interpretation
d = < 0.2	Negligible effect
d = 0.2 - 0.5	Small effect
d = 0.5 - 0.8	Medium effect
d = ≥ 0.8	Large effect

For the G-CAT, effect sizes for all variables fall between -0.19 and 0.08, indicating very small differences between the English and Non-English groups. By convention, d scores lower than 0.2 indicate negligible effect sizes. In that sense, these results suggest that any observed differences in test scores are not practically significant.

3.3.2 Correlation Analysis

Correlation matrices between the test scores for each group were compared to assess the variations in correlational magnitude between the groups. The differences between the correlation coefficients in both groups can inform if the nature of the scores dramatically changes for each or not. The correlations of the English group can be observed in Table 26, while the correlations for the Non-English group can be observed in Table 27.

Table 26: Correlation matrix of G-CAT scores for the English Language group.

Score	Spatial Ability	Numeric Ability	Abstract Ability	Raven	General Ability
Spatial Ability	-				
Numeric Ability	0.41	-			
Abstract Ability	0.53	0.45	-		
Raven	0.42	0.37	0.51	-	
General Ability	0.78	0.77	0.85	0.54	-

All correlation coefficients have p-values of < 0.001.

Table 27: Correlation matrix of G-CAT scores for the Non-English Language group.

Score	Spatial Ability	Numeric Ability	Abstract Ability	Raven-9	General Ability
Spatial Ability	-				
Numeric Ability	0.45	-			
Abstract Ability	0.57	0.5	-		
Raven-9	0.55	0.48	0.67	-	
General Ability	0.8	0.78	0.87	0.7	-

All correlation coefficients have p-values of < 0.001.

The tables indicate that correlations among cognitive abilities are generally consistent between the English and Non-English groups, demonstrating similar patterns of relationships. High correlations between domain scores and the General Cognitive Ability Score (e.g., 0.78 to 0.87) support the validity of the test’s composite score across both language groups.

3.3.3 Summary of Fairness Analysis

The results indicate that the G-CAT demonstrates fairness across language groups. The minimal effect sizes (both Eta Squared and Cohen's d) and consistent correlation patterns provide evidence that differences in test performance are negligible, supporting the validity of administering the test to individuals from diverse linguistic backgrounds.

When comparing the Raven-9 score with the scores of the G-CAT, the observed variations between English and Non-English groups are similarly small. For example, the Cohen's d for the Raven Total Score (-0.1341) is consistent with the effect sizes observed for G-CAT scores like Spatial Total Score (0.0751) and General Cognitive Ability Score (-0.1241). This alignment highlights the validity of the G-CAT scores, as the Raven-9 serves as a recognized benchmark for assessing general cognitive ability. By demonstrating similar patterns of fairness, the inclusion of the Raven-9 in the analysis strengthens the evidence supporting the G-CAT's effectiveness across language groups.

This finding aligns with the design objective of the G-CAT to provide a non-verbal assessment of cognitive ability, minimizing the influence of language on test performance. The low impact of language differences enhances the test's suitability for global use in personnel selection contexts. These findings are particularly relevant for organizations seeking a fair and equitable tool to evaluate candidates from diverse cultural and linguistic backgrounds. The G-CAT's design ensures that test-takers are assessed on their cognitive abilities without undue influence from language proficiency, thereby supporting its use in multinational and multicultural hiring scenarios.

How to use the G-CAT

4.1 How to Administer the G-CAT to candidates

Please click on this link to access the document that provides the instructions on how to administer the Global Cognitive Ability Assessment https://www.sotserver.com/admin/resources/guide/admin_guide.pdf.

4.2 Candidate Preparation for Taking the Test

Ensuring an optimal testing environment is critical for maintaining the validity and reliability of the Global Cognitive Ability Test (G-CAT). The following guidelines outline best practices for preparing a testing environment that minimizes distractions, ensures fairness, and promotes consistent results across test-takers.

4.2.1 Physical Environment

- **Quiet Space:** The testing location should be free from noise, interruptions, and external distractions. This applies whether the test is conducted remotely or in a controlled physical setting.
- **Lighting:** Ensure proper lighting to avoid screen glare or eye strain during the test. Candidates should use a well-lit room where the screen is clearly visible.
- **Seating and Desk Setup:** Use an ergonomic chair and a stable desk or table. Position the device at eye level to reduce neck strain and promote comfort during the test.

4.2.2 Device and Internet Requirements

- **Hardware:** A computer, laptop, phone, or tablet with updated operating systems is recommended.
  Devices must have a functional keyboard and mouse or a touchscreen interface.
- **Internet Connectivity:** A stable and reliable internet connection is essential. A minimum speed of 5 Mbps is recommended for seamless performance. Candidates should test their connectivity in advance to identify and address potential issues.
- **Web Browser Compatibility:** Ensure the test platform works on the internet browser used by the test-taker. The Chrome and Edge web browsers are recommended.

4.3 Use of the G-CAT in Personnel Selection

The G-CAT is a robust assessment tool designed to support organizations in making informed, data-driven decisions during the personnel selection process. By evaluating critical cognitive abilities required for job tasks, the G-CAT enhances the identification, selection, and development of high-performing individuals.

4.3.1 Role of Cognitive Ability in Personnel Selection

Cognitive ability is a well-established predictor of job performance across various roles and industries. Key reasons for its use in selection processes include:

- **Predictive Validity:** High cognitive ability correlates strongly with performance, especially in roles requiring problem-solving, learning agility, and adaptability.
- **Transferable Skills:** Cognitive abilities are not role-specific, making them applicable across diverse industries and job functions.
- **Efficiency:** Cognitive assessments provide a quick and reliable measure of candidates' potential, saving time and resources during the hiring process.

The G-CAT leverages these benefits by providing a reliable measure of general cognitive ability and its components.

4.3.2 Applications in the Hiring Process

The G-CAT can be seamlessly integrated into various stages of the hiring process, including:

- **Initial Screening:** The test helps narrow down the candidate pool by identifying individuals with the cognitive capacity to meet job demands.
- **Assessment Centers:** It serves as an objective tool in combination with other assessments, such as work sample tests or structured interviews.
- **Final Decision-Making:** The G-CAT scores provide valuable input to compare shortlisted candidates and ensure alignment with job requirements.

These applications help organizations streamline recruitment and enhance the fairness and transparency of the selection process while making sure that evidence is collected for the potential value provided by the hired candidates.

4.3.3 Advantages of Using the G-CAT in Personnel Selection

Integrating the G-CAT into the hiring process offers several benefits:

- **Objectivity:** The test provides a standardized measure of candidates' mental abilities and its scores are free from bias introduced by the subjective evaluations of hiring managers.
- **Efficiency:** Automated scoring and instant reporting save time compared to non-digital selection methods.
- **Fairness:** Norm-referenced scaling ensures candidates are evaluated relative to comparable reference groups, minimizing bias.
- **Legal Compliance:** The G-CAT aligns with best practices in psychometric testing, ensuring compliance with employment laws and ethical standards. These advantages make the G-CAT an essential tool for organizations aiming to build a high-performing workforce.

- **Use of Non-Verbal Item Formats:** The absence of verbal reasoning minimizes linguistic and cultural biases, making the test suitable for global populations.
- **Accessibility:** The online administration mode supports remote testing, enabling access for candidates from varied geographic and socio-economic backgrounds.

The G-CAT equips organizations with a powerful tool to identify and select candidates with the cognitive capacity to excel in their roles, ultimately contributing to organizational success and growth. By integrating the G-CAT into the hiring process, organizations can enhance the quality of their workforce while promoting fairness, efficiency, inclusivity, and return on investment.

4.3.4 Understanding the G-CAT Scores for Personnel Selection

This section explains what the G-CAT test scores represent and how to interpret them to make data-driven decisions for personnel selection.

4.3.4.1 Raw Scores

The raw score is the simplest measure of test performance, they represent the number of questions answered correctly. While this is the most direct outcome, raw scores by themselves are not meaningful. Since these scores represent directly unobservable mental abilities, they cannot be interpreted in the same way physical measures like weight or distance are interpreted. In that sense, the interpretation of the scores comes from the *scaling* of candidates' scores in relation to the scores of other candidates.

4.3.4.2 Percentile Scores

Scaled scores convert raw scores into a standardized scale, which indicates what place the test-taker occupiers in relation to other test-takers. This process ensures a more consistent comparison between test-takers or across multiple test sessions. The most commonly used scaled score for purposes of psychological testing is the percentile rank score.

The percentile rank score indicates the proportion of test-takers whose performance is at or below a certain score. For example, a percentile rank score of 80 means a test-taker performed as well as or better than 80% of test-takers in the reference norm group.

Percentile vs. Percentage
A high percentage of correct answers does not necessarily translate to a high percentile rank. Percentile ranks compare you to other test-takers, while percentage refers purely to the fraction of items answered correctly.

4.3.5 Scale Scores produced by the G-CAT

The G-CAT produces four main scores in raw and percentile formats. These are three subtest scores and one total scale score.

Subtest scores
These are the scores for the subtests. The name "subtest" indicates smaller scales that when combined form the Total Scale of Cognitive Ability.

Spatial Ability Score
This score represents the capacity to visualize, manipulate, and understand spatial relationships among objects or shapes. This dimension measures skills such as mental rotation, pattern recognition, and visual problem-solving.

Numeric Ability Score
This score represents the capability to work with numbers, including arithmetic, quantitative analysis, and logical reasoning involving numerical data.

Abstract Ability Score
This score represents the ability to draw logical inferences from patterns, shapes, or concepts. It is less tied to language or numbers and more about identifying underlying structures.

Total Scale score
This score is the result of the combination of the Spatial, Numeric, and Abstract Reasoning Ability subtest scores. In that sense, it represents the "total scale" of the test, which serves as an overarching snapshot of mental ability across multiple domains.

General Cognitive Ability Score (GCA)
The GCA Score represents general intelligence (g-factor) and is often correlated with job performance, learning ability, and problem-solving skills across different professions. It provides a broad estimate of a person's cognitive potential beyond individual cognitive ability domains.

4.3.6 Interpretation of Scores

The G-CAT scores are interpreted against a Norm group reference—an established standard of test performance within specific populations or industries. This allows to transform raw scores to percentile scores based on the validation sample used for the development of the G-CAT.

The raw scores can be used to compare candidates within the same hiring process by ranking them according to that score. Candidates with higher scores are expected to perform better in the respective abilities compared to candidates with lower scores.

The percentile scores can be used to interpret what rank place a specific test-taker would take when compared to participants in the validation study. They can also be used to compare candidates within the same hiring process similar to the raw scores.

4.3.6.1 Confidence Intervals of the Scores

Confidence intervals (CIs) provide an estimate of the range within which a test-taker's true score is likely to fall, given their observed score. They are valuable for understanding the precision of scores and assessing measurement error. The score ranges and Confidence Intervals for each score of the test are indicated in the test report.

For instance, if a test-taker's observed general cognitive ability score is 24 with a 95% CI, and a score range of ± of 3, their true score is likely to be 95% of the time within a range of 21 and 27.

4.3.6.2 Total Test vs. Subtest scores

It is recommended that the General Cognitive Ability Score takes precedence over the subtest scores (Spatial, Numeric, Abstract) when making decisions for personnel selection. The reasons for this include:

Superior Predictive Validity

- **GCA as the Strongest Predictor**: Extensive meta-analyses show that GCA is one of the best predictors of job performance across virtually all occupations. It accounts for an individual's ability to learn, solve problems, and adapt to novel situations, which are critical in most roles.
- **"g" Factor Dominance**: GCA reflects the general intelligence factor (*g*), which underpins performance on all cognitive tasks. Even in jobs requiring specific skills (e.g., programming), *g* explains the majority of variance in performance.

Higher Reliability

- **Composite Scores Are More Stable**: GCA aggregates performance across multiple subtests, reducing measurement error. Subtest scores, based on fewer items, are statistically less reliable and more prone to fluctuation, making them less trustworthy for high-stakes decisions.

Practical Efficiency

- **Simplifies Decision-Making**: Overemphasizing subtests may complicate selection processes without adding significant value. GCA provides a parsimonious metric that streamlines evaluations while maintaining accuracy.

- **Avoids Overfitting**: Tailoring subtest weights to perceived job requirements risks overfitting to specific traits, which may not generalize well. GCA's broad applicability ensures robustness across diverse roles and industries.

Legal and Fairness Considerations

- **Defensibility**: GCA is backed by decades of cognitive ability validation research, making it easier to justify in legal challenges. Subtest weighting, unless rigorously validated for a specific role, may invite claims of arbitrariness or bias.
- **Reduces Bias Risks**: Focusing on subtests might inadvertently disadvantage groups that perform differently on specific measures, even when those differences are irrelevant to job success. GCA minimizes such risks by emphasizing the most universally relevant trait.

Adaptability and Learning Potential

- **Future-Proofing**: GCA predicts not only immediate job performance but also an individual's capacity to acquire new skills and adapt to changing demands. This is critical in dynamic work environments where role requirements evolve over time.

By understanding how G-CAT scores are structured, interpreted, and applied to the personnel selection process, you can make more strategic, data-driven decisions when evaluating candidates for roles that demand a high level of cognitive capability.

4.4 Ethical Considerations for the Use of G-CAT in Personnel Selection

When using G-CAT as a tool for personnel selection, it is critical to uphold the highest ethical standards. The impact of automated or semi-automated decisions on individuals and organizations demands careful thought and ongoing vigilance. Below are key ethical considerations for using G-CAT in a recruitment and selection context:

4.4.1 Fairness and Non-Discrimination

G-CAT must be used in ways that minimize bias and uphold equal opportunity. This includes:

- **Algorithmic Bias**: Regularly evaluate and refine models to prevent negative outcomes based on race, gender, age, disability, or other protected characteristics.
- **Calibration and Monitoring**: Continuously calibrate assessment frameworks across a diverse range of candidates to ensure that recommended decisions do not inadvertently disadvantage specific groups

4.4.2 Transparency and Informed Consent

Employers should inform applicants about the use of G-CAT in the selection process. This encompasses:

- **Clear Communication**: Explain how G-CAT functions, what data it collects, and how these data inform hiring decisions.
- **Candidate Rights**: Provide candidates with opportunities to ask questions and offer consent for automated screening or assessment.

4.4.3 Privacy and Data Protection

Handling sensitive information responsibly is paramount:

- **Data Security**: Ensure robust data protection measures (encryption, secure storage, and controlled access) to safeguard personal details.
- **Data Minimization**: Collect and retain only the data necessary for the legitimate functioning of G-CAT; discard or anonymize data when no longer needed.

4.4.4 Accountability and Human Oversight

Human judgment remains essential to avoid ethically problematic "black-box" outcomes:

- **Review of Automated Results**: Assign qualified reviewers to interpret G-CAT analyses and cross-check results.
- **Appeal Mechanisms**: Offer candidates a clear process to appeal or request clarification on G-CAT-related decisions.

4.4.5 Compliance with Legal and Regulatory Standards

Organizations must stay current with applicable labor, data protection, and anti-discrimination laws:

- **Regulatory Alignment**: Align G-CAT's usage with local, national, and international guidelines (e.g., GDPR, EEOC).
- **Policy Documentation**: Maintain policy documents that define processes for ethical G-CAT deployment and incorporate changes in legislation promptly.

By integrating these ethical considerations into everyday practice, organizations can leverage G-CAT to make informed, equitable hiring decisions that respect candidates' rights and uphold societal expectations for fair and transparent employment processes, while also mitigating risks that could emerge from lack of compliance with pertinent labor laws and regulations.

REFERENCES

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Bilker, W. B., Hansen, J. A., Brensinger, C. M., Richard, J., Gur, R. E., & Gur, R. C. (2012). Development of abbreviated nine-item forms of the Raven's Standard Progressive Matrices Test. *Assessment, 19*(3), 354–369.

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

Jensen, A. R. (1998). *The g factor: The science of mental ability*. Praeger.

Raven, J. C. (1938). *Progressive matrices: A perceptual test of intelligence*. H.K. Lewis.

Sackett, P. R., et al. (2017). Revisiting the design of selection systems in light of new findings about the validity of widely used predictors. *Industrial and Organizational Psychology, 10*(3), 291–301.

Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C., de Fruyt, F., & Rolland, J. P. (2003). A meta-analytic study of general mental ability validity for different occupations in the European Community. *Journal of Applied Psychology, 88*(6), 1068–1081.

Slaney, K. (2017). *Validating psychological constructs: Historical, philosophical, and practical dimensions*.

Society for Industrial and Organizational Psychology. (2018). *Principles for the validation and use of personnel selection procedures* (5th ed.).