Construction and Standardization of Psychology Aptitude Test for Incoming College Psychology Students

Keisha Charisse O. Digon[1], Arjay Y. Alvarado[2]

[1]Mental Health Unit, Corazon Locsin Montelibano Memorial Regional Hospital, Philippines
Email: kcdigon@gmail.com
[2]College of Arts and Sciences, Carlos Hilado Memorial State University, Philippines
Email: arjay.alvarado@chmsu.edu.ph

Abstract— *Psychology programs in higher education institutions rely on accurate assessment tools to gauge the aptitude of incoming students effectively. However, existing standardized tests often fail to address the unique skill sets and knowledge domains specific to psychology, emphasizing the necessity for a tailored Psychology Aptitude Test. This study aims to bridge this gap by developing and standardizing a Psychology Aptitude Test tailored for incoming college psychology students. Anchored on Item Response Theory (IRT), the study endeavors to create a test that comprehensively evaluates students' preparedness for the BS Psychology program. Objectives include assessing the Psychology Aptitude Test's validity, item difficulty and discrimination indices, and reliability. Findings reveal a sound measurement with strong content validity, balanced difficulty levels, and internal consistency. Recommendations for ongoing item review, expanded validation studies, and predictive validity assessment are provided to optimize the test's effectiveness in evaluating psychological aptitude within college environments.*

Keywords — *Psychology Aptitude Test, Item Response Theory, psychometrics, validity and reliability, college readiness*

I. INTRODUCTION

Psychology, as an academic discipline and profession, plays an important role in understanding human behavior, cognition, and emotions. With the increasing popularity of psychology programs in colleges and universities, a compelling need arises to ensure the effectiveness of assessment tools for incoming students. While standardized tests exist for various academic fields, constructing an aptitude test specifically tailored for psychology students remains an unmet challenge (Cohen & Swerdlik, 2018). Existing standardized tests often lack specificity to the unique skill sets and knowledge domains required in psychology, potentially leading to inaccurate assessments of students' aptitude and preparedness (Bridgeman & Wendler, 2016). This gap highlights the necessity for a Psychology Aptitude Test to comprehensively evaluate incoming college psychology students' readiness and suitability for the demands of their chosen field.

Research in the field of psychological assessment has primarily focused on the development and validation of tests for specific constructs such as intelligence, personality, and clinical disorders (Graham, 2016). While these assessments provide valuable insights into individual differences, they do not fully capture the breadth of skills and competencies relevant to success in psychology education and practice. Furthermore, the existing literature on aptitude testing in psychology predominantly revolves around postgraduate or professional levels, overlooking the critical transition phase of incoming college students (Kuncel et al., 2013). Consequently, there is a notable gap in the literature regarding the construction and standardization of a Psychology Aptitude Test explicitly tailored for incoming college psychology students. Addressing this gap is essential for enhancing the accuracy and validity of student assessments and ultimately promoting the quality of psychology education.

Developing and implementing a Psychology Aptitude Test holds significant benefits for multiple stakeholders within

Digon and Alvarado

Journal of Humanities and Education Development (JHED)
J. Humanities Educ. Dev.- 7(2)-2025

the academic and professional domains. The test results may be utilized to make informed decisions regarding the selection of incoming psychology students, ensuring that admitted individuals possess the necessary foundational knowledge and skills for success in their academic endeavors. Additionally, students themselves may benefit from a more accurate assessment of their aptitude, guiding their academic and career aspirations and facilitating their personal and professional development in the field of psychology. Overall, the construction and standardization of a Psychology Aptitude Test has the potential to enhance the quality and rigor of psychology education, benefiting students, educators, and the broader psychological community alike.

Objectives of the Study

1. What is the validity of the Psychology Aptitude Test?

2. What is the difficulty index of each item of the Psychology Aptitude Test?

3. What is the discrimination index of each item of the Psychology Aptitude Test?

4. What is the reliability of the Psychology Aptitude Test?

5. What is the test norm of the Psychology Aptitude Test?

Framework

This study is anchored on the Item Response Theory (IRT) proposed by Lord and Novick (1968). This provides a strong framework for constructing and standardizing the Psychology Aptitude Test for incoming college psychology students. IRT models the relationship between an individual's abilities and how they respond to test items, yielding an advanced approach for creating test items. By utilizing IRT principles, the researchers can ensure that the items included in the Psychology Aptitude Test accurately measure students' psychology-related abilities across various difficulty levels. This ensures that the test provides reliable and valid assessments, crucial for evaluating students' aptitude and readiness for psychology education at the college level.

In the context of this study, IRT guides the item development process, allowing for the creation of test items that effectively discriminate between students with different levels of aptitude in psychology. Furthermore, IRT provides guidance for the calibration of test items, ensuring that they accurately measure students' abilities while maintaining consistency and reliability across different test administrations. By anchoring the study on IRT, the researchers aim to develop a Psychology Aptitude Test that provides precise and informative measurements of incoming college psychology students' aptitude for the field.

Scope and Limitations

The scope of this study encompassed the construction and standardization of a Psychology Aptitude Test for incoming college psychology students, including item development, validation procedures, pilot testing, and psychometric analysis. This aimed to provide a comprehensive assessment tool that accurately measures students' knowledge, skills, and abilities in Psychology.

II. MATERIALS AND METHODS

Research Design

Instrumentation design, particularly in the context of developing an aptitude test, involves creating a set of questions and tasks that accurately assess the knowledge, skills, and abilities of individuals in a specific field. This process includes designing questions that cover a wide range of topics within the subject area, varying in difficulty to effectively evaluate the candidates' proficiency. The design of an aptitude test for instrumentation engineers aims to test their understanding of concepts related to instrumentation, transducers, control systems, and other relevant areas (libguides.mit.edu, n.d.).

Furthermore, the design of an aptitude test for instrumentation engineers should ensure that the test is valid, reliable, and free of bias. Validity ensures that the test measures what it intends to measure, while reliability ensures consistent results when administered multiple times. Additionally, the design of an aptitude test should be based on a conceptual framework that aligns with the learning objectives and outcomes of the test.

Research Procedure

This research was advanced to develop an aptitude test that aims to serve as a reliable instrument systematically designed to assess various facets of students' cognitive abilities, analytical reasoning, and psychological knowledge. In doing so, the researchers followed this procedure:

Firstly, the researchers initiated the study by conducting a comprehensive review of relevant literature. Secondly, they developed test items based on the specifications outlined in the Introduction to Psychology course, a fundamental component of the BS Psychology program. Thirdly, the validation of these items was undertaken by five subject matter experts. Following the validation process, a request to carry out the study was submitted to the office of the Vice-President for Academic Affairs. Upon approval, the researchers administered the instrument to three-hundred sixty-eight applicants for the

Digon and Alvarado

Journal of Humanities and Education Development (JHED)
J. Humanities Educ. Dev.- 7(2)-2025

BS Psychology program at Carlos Hilado Memorial State University. Subsequently, the data collected from the pilot test was compiled and analyzed utilizing appropriate statistical methods.

Statistical Analysis

For problem number one, which aims to determine the validity of the Psychology Aptitude Test, the Content Validity Ratio developed by Lawshe was used.

For problem number two, which aims to determine the difficulty index of each item of the Psychology Aptitude Test, Crocker's (1986) method was followed

For problem number three, which aims to determine the discrimination index of each item of the Psychology Aptitude Test, Ebel's (1979) method was followed.

For problem number four, which aims to determine the reliability of the Psychology Aptitude Test, KR20 was used.

III. RESULTS AND DISCUSSION

Statement of the Problem No. 1

In the first phase, the researchers developed a 70-item test and were subjected to content validation. Five subject matter experts who are psychology teachers validated the test by following Lawshe's method (1975). After the validation process, only sixty (60) items yield at least a .99 content validity ratio (required CVR to be retained using five validators).

Statement of the Problem No. 2 and 3

Table 1 provides a detailed breakdown of how individual items performed in the psychology aptitude test, specifically looking at their difficulty and ability to differentiate between high- and low-scoring students. The distribution of difficulty levels was well-balanced, with most items categorized as moderately difficult, aligning with previous studies (Candiasa et al., 2018; Raza et al., 2022; Gupta, 2010; Propp, 2005; Lloyd, 1991). However, similar research shows that some items required revision due to moderate difficulty but weak discrimination. In contrast, others were removed entirely because they failed to effectively distinguish between students, particularly those with negative discrimination indices. This pattern suggests that the test was designed to cover a range of difficulty levels, ensuring a well-rounded evaluation of students' aptitude.

Looking at the discrimination index, the results varied. Certain items demonstrated strong discrimination, echoing findings from Lim et al. (2015). For instance, while categorized as very difficult, items like 32, 36, and 55 were still able to differentiate between students with higher

and lower aptitudes effectively. This suggests that even the most challenging items in the test play an important role in identifying differences in ability. However, items flagged for revision or removal indicate areas where the test could be improved to maintain its accuracy and effectiveness in assessing student aptitude.

To enhance the overall reliability and precision of the test, only the items with the highest discrimination indices—those classified as "Very Good"—were retained. These high-performing items were particularly effective at distinguishing between students with different levels of aptitude and thus contributed significantly to the test's accuracy. As a result, the original 60-item test was reduced to 30 carefully selected items.

The discrimination index was calculated by comparing how often students in the top 27% (higher scorers) and bottom 27% (lower scorers) answered each item correctly. This method helps ensure that the test accurately identifies variations in students' abilities, making it a reliable tool for assessing psychological aptitude.

Table 1. Difficulty and Discrimination Indices

Item No.	Difficulty Index		Discrimination Index	
1	0.49	Moderate	0.26	To be revised
2	0.75	Easy	-0.08	To be discarded
3	0.26	Difficult	0.38	**Very Good Item**
4	0.43	Moderate	0.29	To be revised
5	0.15	Difficult	0.51	**Very Good Item**
6	0.68	Moderate	0.08	To be discarded
7	0.63	Moderate	0.14	To be discarded
8	0.80	Easy	-0.03	To be discarded
9	0.38	Moderate	0.40	**Very Good Item**
10	0.82	Easy	-0.09	To be discarded
11	0.75	Easy	-0.03	To be discarded
12	0.43	Moderate	0.29	To be revised
13	0.86	Very Easy	-0.18	To be discarded

Digon and Alvarado

Journal of Humanities and Education Development (JHED)
J. Humanities Educ. Dev.- 7(2)-2025

No.	Value	Difficulty	Value	Remark
14	0.30	Moderate	0.41	**Very Good Item**
15	0.15	Difficult	0.55	**Very Good Item**
16	0.69	Moderate	0.04	To be discarded
17	0.58	Moderate	0.19	To be revised
18	0.76	Easy	0.03	To be discarded
19	0.48	Moderate	0.19	To be revised
20	0.32	Moderate	0.41	**Very Good Item**
21	0.16	Difficult	0.49	**Very Good Item**
22	0.86	Very Easy	-0.14	To be discarded
23	0.19	Difficult	0.48	**Very Good Item**
24	0.90	Very Easy	-0.23	To be revised
25	0.80	Easy	-0.08	To be discarded
26	0.21	Difficult	0.50	**Very Good Item**
27	0.17	Difficult	0.49	**Very Good Item**
28	0.65	Moderate	0.12	To be discarded
29	0.77	Easy	-0.07	To be discarded
30	0.47	Moderate	0.24	To be revised
31	0.85	Easy	-0.15	To be discarded
32	0.07	Very Difficult	0.57	**Very Good Item**
33	0.64	Moderate	0.13	To be discarded
34	0.32	Moderate	0.38	**Very Good Item**
35	0.60	Moderate	0.22	To be revised
36	0.11	Very Difficult	0.59	**Very Good Item**
37	0.16	Difficult	0.54	**Very Good Item**
38	0.71	Easy	-0.02	To be discarded
39	0.40	Moderate	0.32	**Very Good Item**
40	0.64	Moderate	0.18	To be revised
41	0.27	Difficult	0.44	**Very Good Item**
42	0.21	Difficult	0.49	**Very Good Item**
43	0.49	Moderate	0.23	To be revised
44	0.61	Moderate	0.14	To be discarded
45	0.34	Moderate	0.35	**Very Good Item**
46	0.43	Moderate	0.20	To be revised
47	0.31	Moderate	0.45	**Very Good Item**
48	0.44	Moderate	0.28	To be revised
49	0.32	Moderate	0.40	**Very Good Item**
50	0.40	Moderate	0.39	**Very Good Item**
51	0.54	Moderate	0.19	To be revised
52	0.30	Moderate	0.41	**Very Good Item**
53	0.29	Difficult	0.30	**Very Good Item**
54	0.24	Difficult	0.44	**Very Good Item**
55	0.12	Very Difficult	0.47	**Very Good Item**
56	0.33	Moderate	0.32	**Very Good Item**
57	0.32	Moderate	0.36	**Very Good Item**
58	0.32	Moderate	0.43	**Very Good Item**
59	0.24	Difficult	0.47	**Very Good Item**
60	0.23	Difficult	0.45	**Very Good Item**

Statement of the Problem No. 4

Table 2 provides key metrics for the Psychology Aptitude Test for incoming psychology students, showing that the test is both reliable and well-structured. With 363 students completing the test and a full 100% response rate, the data

Digon and Alvarado

Journal of Humanities and Education Development (JHED)
J. Humanities Educ. Dev.- 7(2)-2025

is complete and representative of the group. One of the test's strongest indicators is its internal consistency, reflected in a high Cronbach's Alpha (KR-20) score of 0.892. Since a reliability score above 0.70 is generally considered acceptable (Nunnally & Bernstein, 1994), this result suggests that the test consistently measures what it is designed to assess. A strong internal consistency like this ensures that students' scores are not random but instead reflect their actual aptitude for psychology (Anastasi & Urbina, 1997).

The test consists of 30 carefully selected items, allowing for a thorough evaluation of students' psychological aptitude while avoiding unnecessary redundancy. Research on test development emphasizes that balancing the number of items and their quality leads to more accurate assessments and prevents test fatigue (Thorndike & Thorndike-Christ, 2010).

Table 2. Reliability of Psychology Aptitude Test

Scale		n	%	Cronbach's Alpha (KR20)	N of Items
Psychology Aptitude Test	Cases Valid	363	100	.892	30
	Excluded	0	.0		
	Total	363	100		

Statement of the Problem No. 5

The Psychology Aptitude Test for incoming first-year college students was developed after refining the test through a careful selection process. Initially, it contained 60 items, but after analyzing how well each question measured what it was supposed to—by looking at difficulty levels, discrimination power, and overall reliability—the test was reduced to 30 stronger items. This refined version was then given to a new group of 387 students to establish a standard way of interpreting scores. The results were organized using the Stanine scale, a standard system in education that divides test scores into nine levels, making it easier to compare individual results to the larger group.

The stanine system classifies students into three categories: above average, average, and below average. As indicated in Table 3, those who scored 19 and above were placed in the above-average range, meaning they likely have strong psychological aptitude and critical thinking skills that could help them excel in psychology-related subjects. Most students, who scored between 13 and 18, fell within the average range, indicating they have the foundational skills necessary for success but may perform at varying levels. Meanwhile, students who scored 12 and below were categorized as below average, suggesting they may need extra support in psychology-related coursework.

Establishing a test norm like this is important because it provides a straightforward way to interpret results based on a larger sample rather than looking at scores in isolation. Psychologists and education experts like Anastasi and Urbina (1997) have emphasized how

norming makes test scores more meaningful by showing how an individual's performance compares to others. The stanine system, which is commonly used in standardized assessments, helps simplify score interpretation while maintaining accuracy (Thorndike & Thorndike-Christ, 2010). Since the test was reduced to 30 items through detailed analysis, only the most reliable and relevant questions were kept. This process follows established research on test development, highlighting how refining questions improves accuracy (Nunnally & Bernstein, 1994). Research also supports the idea that stanine scores can help predict academic performance, making them helpful in identifying students who might struggle or excel (De Ayala, 2009).

Table 3. Psychology Aptitude Test Norm for Incoming College Students

Raw Score	Stanine	Interpretation
23 and above	9	Above Average
21-22	8	Above Average
19-20	7	Above Average
17-18	6	Average
15-16	5	Average
13-14	4	Average
11-12	3	Below Average
9-10	2	Below Average
8 and below	1	Below Average

n=387

Digon and Alvarado

Journal of Humanities and Education Development (JHED)
J. Humanities Educ. Dev.- 7(2)-2025

IV.   CONCLUSIONS AND RECOMMENDATIONS

In summary, the development and validation process of the Psychology Aptitude Test has yielded positive outcomes, indicating its potential as a valuable tool for evaluating psychological aptitude in incoming college students. By meticulously validating content and selecting items, the test now consists of 30 items with strong content validity, ensuring they accurately measure the intended constructs. Analysis of item performance revealed a balanced range of difficulty levels, and careful prioritization of items with notable discrimination indices emphasizes the test's accuracy in distinguishing between individuals with varying levels of aptitude. Moreover, the high Cronbach's Alpha coefficient affirms the test's internal consistency, reinforcing its reliability in consistently assessing psychological constructs. In the subsequent timeframe, continuous evaluation, diverse validation studies, and refined discrimination analysis will enhance the test's efficacy and suitability for assessing psychological aptitude in college settings.

Recommendations

Following a comprehensive analysis of the development and validation of the Psychology Aptitude Test, practical recommendations are presented here to enhance its effectiveness in evaluating psychological aptitude among incoming college students. Based on the study's findings, these recommendations aimed to strengthen the test's utility and relevance in the college setting.

Continuous Item Review and Revision. Regularly review and update test items to ensure relevance and effectiveness in assessing psychological aptitude.

Expanded Validation Studies. Conduct additional studies with diverse samples to confirm the test's effectiveness across different populations and contexts.

Predictive Validity Assessment. Conduct a predictive validity assessment to examine the extent to which scores on the Psychology Aptitude Test predict future academic success and retention in the BS Psychology program.

Longitudinal Studies. Undertake studies tracking students' academic and career outcomes over time to assess the test's predictive validity.

REFERENCES

[1] Anastasi, A., & Urbina, S. (1997). *Psychological Testing* (7th ed.). Prentice Hall.

[2] Bridgeman, B., & Wendler, C. (2016). Fairness in college admissions testing: Validity and validity issues. In C. R. Reynolds, K. J. Vannest, & E. Fletcher-Janzen (Eds.), Encyclopedia of special education: A reference for the education of children, adolescents, and adults with disabilities and other exceptional individuals (pp. 841-844). Wiley.

[3] Candiasa, I. M., Natajaya, N., & Widiartini, K. (2018). Vocational aptitude test. *SHS Web of Conferences, 42*, 00044. http://dx.doi.org/10.1051/shsconf/20184200044

[4] Cohen, R. J., & Swerdlik, M. E. (2018). Psychological testing and assessment: An introduction to tests and measurement. McGraw-Hill Education.

[5] Crocker, L. (1986). Methods for determining the difficulty index of test items. *Journal of Educational Measurement, 23*(2), 121–132.

[6] De Ayala, R. J. (2009). *The Theory and Practice of Item Response Theory*. Guilford Press.

[7] Ebel, R. L. (1979). Techniques for determining the discrimination index of test items. *Educational and Psychological Measurement, 39*(4), 647–655.

[8] Graham, J. R. (2016). MMPI-2: Assessing personality and psychopathology (5th ed.). Oxford University Press.

[9] Gupta, H. (2010). Graduate pharmacy aptitude test. *Journal of Pharmacy and Bioallied Sciences, 2*(1), 55. http://dx.doi.org/10.4103/0975-7406.62715

[10] Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2013). Academic performance, career potential, creativity, and job performance: Can one construct predict them all? J*ournal of Personality and Social Psychology, 86*(1), 148-161.

[11] Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology, 28*(4), 563–575.

[12] Lim, H. A., et al. (2015). The validity of the Modern Language Aptitude Test (MLAT) in a bilingual context: A case study of Singapore. *Language Testing, 32*(2), 193-218. https://doi.org/10.1177/0265532214531319

[13] Lloyd, B. (1991). Test your own aptitude. *Long Range Planning, 24*(2), 122. http://dx.doi.org/10.1016/0024-6301(91)90130-g

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

[15] Propp, J. (2005). Self-referential aptitude test. *Math Horizons, 12*(3), 35. http://dx.doi.org/10.1080/10724117.2005.12021810

[16] Raza, M. A., Deeba, F., & Faqir, R. (2022). A comparative analysis of school teachers' teaching aptitude. *Global Educational Studies Review, 7*(3), 45–52. http://dx.doi.org/10.31703/gesr.2022(vii-iii).05

[17] Thorndike, R. M., & Thorndike-Christ, T. (2010). *Measurement and Evaluation in Psychology and Education* (8th ed.). Pearson.

Appendix

Sample Test Items

This section presents selected test items from the Psychology Aptitude Test for Incoming College Psychology Students. These items assess fundamental psychological concepts and critical thinking skills relevant to incoming psychology students.

Item 9

Your friend has a hard time understanding classical conditioning. As a psychology student, how can you explain this concept to your friend?

a. Whenever you bring a baseball bat home, you take your child to the park to play. As a result, every time your youngster sees you bring home a baseball bat, he becomes delighted because he associates your baseball bat with a trip to the park

b. Your parents gave you a reward every time you get a high score in an exam

c. Every time you clean your room, your mother rewards you with 1000 pesos

d. All of the above

Item 42

Which kind of concept do psychometricians employ when determining the status of the criterion in relation to a test's accuracy?

a. Predictive Validity

b. Content Validity

c. Concurrent Validity

d. Construct Validity

Item 60

What is the probability of rolling two dice and getting a sum of 7?

a. 1/6

b. 1/12

c. 1/36

d. 6/36