

Taxonomy of Aptitude Test Items: A Guide for Item Writers

Carlo Magno

De La Salle University, Manila, Philippines

Abstract

A taxonomy of aptitude test items is organized based on widely used aptitude tests. The taxonomy is based on three-stage Delphi technique from experts in measurement and test development practitioners. The issues on the need to arrive with a taxonomy of aptitude test items are discussed. The classifications of these items are described and examples of items are provided. The taxonomy created showed that the nature of variables measured for each type varies. Through the taxonomy, item writers can be guided on what items to write given the nature of the construct that needs to be measured. The utility value of the taxonomy is discussed.

Aptitude tests are commonly used in the educational, clinical, and industrial settings to determine the potential abilities of individuals for a wide variety of purposes. Most of the people in these settings commonly use available and standardized aptitude tests. However, for some experts in measurement and assessment, available aptitude tests in the market do not provide them with certain needs such as measuring domain specific-variables like potentials of nurses and engineers. These experts address this concern by constructing their own aptitude tests. Constructing aptitude tests involves not only the skills and expertise in psychometrics but also expertise in visual artists, mathematics, drafting, computer graphics, and other related fields in the arts. For a test development expert, his procedural knowledge in constructing the entire test and the theoretical background on test construction is given. The dilemma faced by the expert is identifying certain aptitude test types that can ascertain the measurement of a specified construct needed. Guidelines in test types are provided mostly by assessment references but what is needed is a systematized classification of aptitude test types to guide test developers especially when constructing aptitude test items. Even standardized test manuals are not too keen on reporting the basis of including and creating certain test types such as verbal and figural reasoning. Published articles in the *Journal of Psychological Assessment*, *Psychological Assessment*, *Assessment and Evaluation in Higher Education*, and *Educational and Psychological Measurement* show absence of resources in classifying certain test types. The closest article to include a classification of test item is the study of De Raad and Barelds (2008) where they constructed a list of 2,365 personality descriptive items from a computerized database of the Dutch language. In the same way, Alreck and Settle (2004) identified numerous scale types used for questionnaires, inventories, and surveys (ex. Lickert, Verbal frequency, Linear numeric etc.). But these classifications are only useful for non-cognitive assessment and not for aptitude and ability tests. One of the major aims of scientific inquiry is to describe and classify phenomenon and the field of psychological testing, measurement, and evaluation should start adhering to this basic scientific premise.

There is a need to construct a taxonomy of aptitude test items primarily to (1) aid test developers in constructing items, (2) frame appropriate constructs measured for particular test types, and (3) target particular skills and potentials

for each type of item. The available schemes of tests are classified according to their function such as cognitive vs. noncognitive, achievement vs. aptitude, verbal vs. nonverbal, and individual vs. group tests. This classification is useful in reference to the use of the tests and the kind of variable measured. For ability measures, test items are commonly classified using methods of factor and cluster analysis (ex. Principal Components Analysis, Confirmatory Factor Analysis, Joining Tree Clustering, K Means Clustering, Multidimensional Scaling etc.). The goal of factor analysis is to group variables (or items in the case of a test) according to their common variance (Anastasi & Urbina, 2002). Such classification are the Two-Factor theories by Charles Spearman (general and specific factors), and Cattell's fluid and crystallized intelligence. There are also theories centered on multiple factors such as Thurstone's Primary mental abilities (Verbal comprehension, word/verbal fluency, inductive reasoning, spatial visualization, number, memory, and perceptual speed), and Guilford's Structure intellect model (Kaplan & Saccuzzo, 1997). Some contemporary classification of abilities are Ennis' (1987) taxonomy of critical thinking and disposition of abilities, Sternberg, Castejon, Prieto, Hantamaki, and Grigorenko (2001) used confirmatory factor analysis to come up with the successful intelligence components (practical, creative, and analytic) across different cultures. Other schemes on abilities are discussed in the paper of Lohman and Hagen (2001). These theories account for classification of abilities like intelligence. It is reflected in the factors that distinct types of test items are more appropriate for each kind of ability. For example, the verbal comprehension in Thurstone's Primary Mental Abilities is best measured by vocabulary test items such as reading comprehension, verbal analogies, disarranged sentences, verbal reasoning, and proverb matching. Placing certain types of items on specific abilities means that items can vary in their functionality.

The primary purpose of this article is to present and describe a taxonomy of aptitude test items organized on two schemes (verbal and nonverbal) and according to the ability measured. Secondly, the usefulness of such a taxonomy is discussed. The taxonomy was constructed based on a series of Delphi technique among different test users and developers (see Vázquez-Ramos, Leahy, & Hernández, 2007). This proposed taxonomy was arrived at by compiling different test types from widely used standardized aptitude and ability tests that are published. These catalogued test types were first presented to selected experts in visual arts, multimedia arts, architects, industrial designers, philosophy (logic), and mathematics to rank the items that best measure the abilities exemplified in their fields. Items that had very low rankings were then removed (see Appendix A). In a second-stage Delphi technique, experts in test development at the Center for Learning and Performance Assessment¹ confirmed the ranks based on three criteria: (1) Commonly used by standardized tests, (2) generalizability for real life skills, and (3) critical skills measured. In a third-stage Delphi technique, the final selected items were presented to experts in test development from the Philippine Educational Measurement and Evaluation² and Asian Psychological Services and Assessment Corporation³. The purpose of the third-stage Delphi technique was to categorize the items based on similarities of the items selected and other criteria that the group will identify. The final configuration of taxonomy resulted in 12 sets of item types.

Description of the Taxonomy

The taxonomy of aptitude ability test items describes a systematic organization of items according to the task the examinee underwent, skills measured, and courses or fields that tapped these skills. All similar items that tapped the same skill were catalogued as one category. These catalogues resulted to 12 classifications. These 12 classifications are categorized under verbal and nonverbal. The schemes in the verbal are verbal analogy, syllogism, and number or letter series, while the nonverbal is composed of topology, visual discrimination, progressive series, visualization, orientation, figure ground perception, surface development, object assembly, and picture completion.

The taxonomy provided is limited on ability test types that measure general intelligence and potential. Performance aptitude test types for individual tests are not included (Ex. "In what ways are the horse and the cow alike?"). Most of the test types in the taxonomy are taken from group tests. Including test types from individual and performance tests is infinite because test developers can always come up with new ways of measuring it.

The classification in the taxonomy is described according to the task required, example of test with containing the item type, skills measured, field/courses that taps into these skills.

Verbal Analogy. The examinee is presented a pair of words and another word without its pair. The examinee is tasked to identify the word that has the same relationship to the first pair. For example, Pyramid is to Egypt, as Parthenon is to _____ (a. Babylon, b. Greece, c. Persia, d. Mesopotamia). Verbal analogy is a measure of vocabulary and verbal comprehension which is the most stable and least deteriorating aspect of intelligence (Rapaport, Gill, & Schafer, 1969). This ability is relevant in courses that requires communication ability. Verbal analogies are used in tests like Otis Lenon School Ability Test (OLSAT), Cognitive Abilities Test (cogAT), and the Wide Range Intelligence Test (WRIT).

Syllogism. A series of statements are presented to the examinee and the examinee should choose the most logical conclusion based on the given statements.

For example:

No heroes are cowards.

Some soldiers are cowards.

Therefore some soldiers are...

- a. not heroes*
- b. cowards
- c. heroes
- d. not cowards

*Correct answer

Syllogism is a measure of logic specifically deriving conclusions based on a set of premises. This skill is important for critical thinking and reasoning. Items on syllogistic reasoning are measures of inductive or general reasoning. These skills are exemplified by lawyers, scientists, researchers, and philosophers. Syllogism test types are found in the Watson Glaser Critical Thinking Appraisal (WGCTA) and Stanford-Binet Intelligence Scale: Fifth Edition (SB-5).

Number/Letter Series. The examinee is given a group of letters and/or numbers, in which he/she is to provide the next group in the series.

For example:	
Number series	Letter Series
7, 13, 21, 33, <u>?</u>	A, D, H, M, S, <u>?</u>
a. 49*	a. X
b. 64	b. Y
c. 56	c. Z*
d. 72	d. A
*Correct answer	*Correct answer

Number and letter series measure understanding of basic quantitative concepts and relationships that are essential for learning mathematics. Tasks measure both the understanding of relational concepts and the ability to discover relationships and to figure out a rule or principle that explains them. This type of test also measures speed and accuracy of simple arithmetic computations. These skills are reflected in mathematics and other computational courses. This test type is widely used in many ability tests such as the Wonderlick Personnel Test (WPT), the quantitative battery of the cogAT, Wechsler Adult Intelligence Scale-Revised (WAIS-R), OLSAT, SB-5, and Slosson Intelligence Test: Primary (SIT-P).

Topology. The examinee is tasked to select from the choices which duplicate the condition presented. This type tests how the examinee apprehends figures presented for his observation, sees the relations between them, conceives the nature of the figure completing each system of relations presented, and by doing so, develops a systematic method of reasoning.

For example:

Duplicating Conditions	Logical Pattern with Missing Design
<p>a. </p> <p>b. </p> <p>c. </p> <p>d. </p> <p>Answer: B</p>	<p>Source: Raven, J. (1986). <i>Manual for Raven's Progressive Matrices and Vocabulary Scales</i>. London: H. K. Lewis.</p>

Topology is a measure of spatial relationship which is the ability to see two or more objects in relation to each other. This also tests perceptual speed which involves quick and accurate grasping of visual details, similarities and differences. These skills are required in courses that need analytical skills. This test is very evident in the Raven's Progressive Matrices.

Visual discrimination. The examinee identifies the similarities and differences between or among objects. Answering this test type involves the ability to recognize and sense objects in original or altered environments.

For example:
Select the object that does not belong in the group.

(A)	(B)	(C)	(D)

Correct answer: C

Discrimination can also be in a verbal form where the examinee selects the concept that does not belong in a group.

For example:

Which of the four is least like the other three?

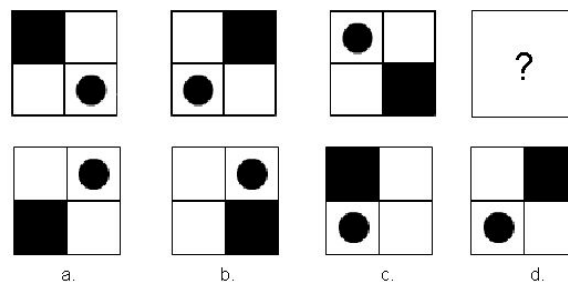
- a. Africa
- b. China
- c. Japan
- d. Korea

Correct answer: A, Africa is not an Asian country

Test 5 of the nonverbal battery of the cogAT is a measure of visual discrimination. The Form Perception subtest of the General Aptitude Test battery (GATB) measures the ability to make visual comparisons and discriminations, see slight differences in shapes and shadings of figures and width and length of lines, and mentally manipulate graphic figures in the mind. The Test of Visual-Reasoning and Processing Skills (TVRPS) also includes a visual discrimination test type that measures a person's ability to determine visually fine distinguishing characteristics among four objects and things, in which one of them would be exactly the same as the stimulus. The Test of Visual Perceptual Skills: Third Edition (TVPS-3) also contains a visual discrimination subtest. These skills are necessary in content and social science courses.

Progressive Series. The examinee is tasked to select one of the images from the options that should come next given the image sequence presented.

For Example:



Correct answer: a

The progressive series is a measure of fluid general intelligence which involves understanding of abstracts and novel relations. Considering it is a perceptual test that involves sequence of figures, it involves the ability to look into the relations of figures. Courses that require logic tap into these skills. The Culture Fair Intelligence Test purely used progressive series.

Classes of Visualization. The examinee imagines how pictorially presented objects will appear when they are rotated, twisted, or inverted. It measures the ability to mentally manipulate patterns in unsped conditions (Colom, Escorial, Shih, & Privado, 2007). This test type also involves the ability to recognize the identity of an object when it is seen from different angles.

When tasked with this type of item, the examinee imagines the movement or internal displacement among parts of a configuration (Thurstone, 1938).

The subclasses of visualization can vary from two-dimension to three-dimension objects and figures as stimuli. It can also vary from manipulation to rotation of objects and figures.

Figure 1
Two-dimensional Object Rotation

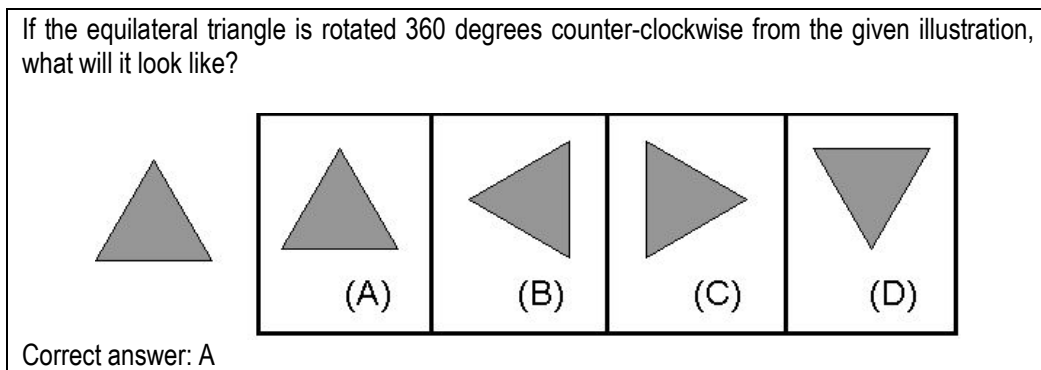


Figure 2
Object Manipulation

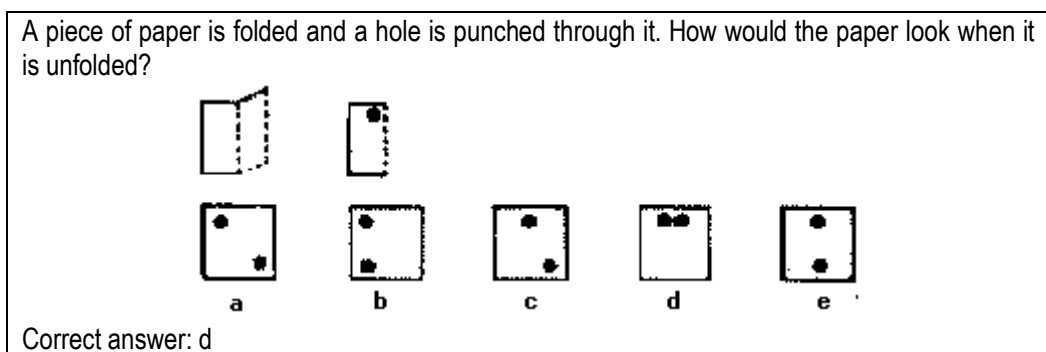
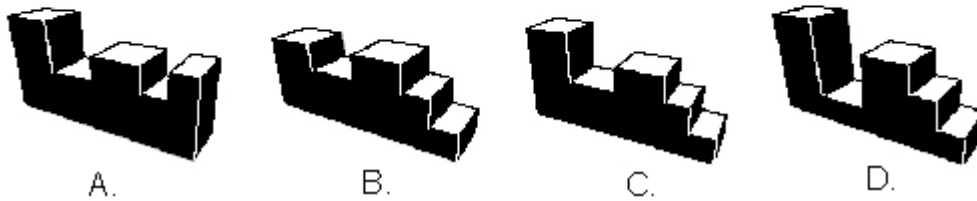
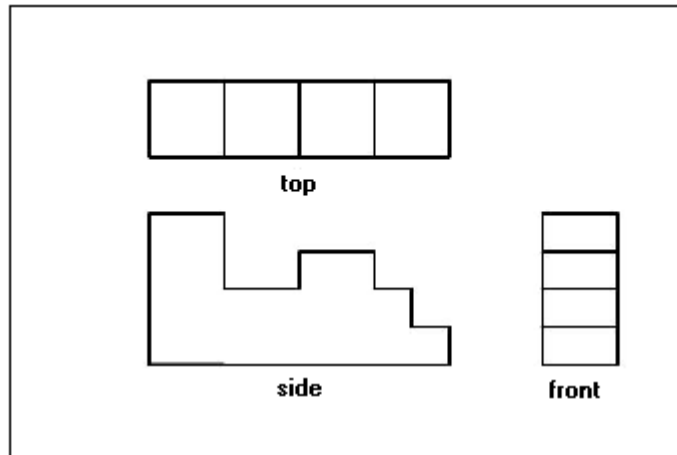


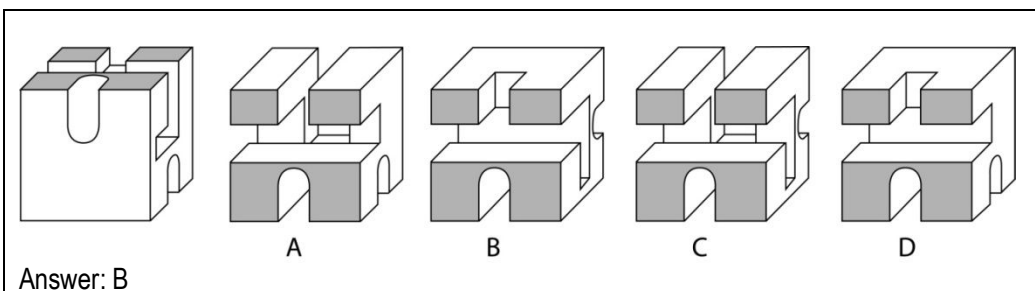
Figure 3
Three-dimensional Object Manipulation

Given the following views of an object, combine the three views and choose the corresponding 3D object that matches it.



Answer: D

Figure 4
Three-Dimensional Object Rotation



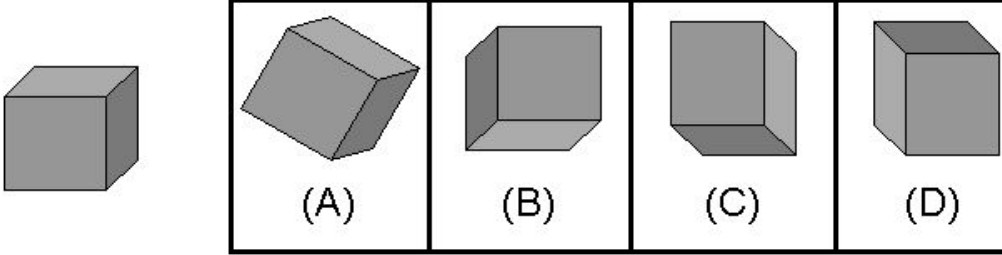
Answer: B

The tests that contain some classes of visualization are the Spatial Test Battery (STB), SB-5, and Slosson Intelligence Test: Primary (SIT-P). Courses that tap into these skills are computer graphics, multimedia arts, design courses, and drafting.

Orientation. The examinee is tasked to detect arrangement of elements within a pattern and the ability to maintain accurate perception in the face of

changing orientations (Owen, 1999; Ozer, 1987). Orientation is required in courses such as animation, computer graphics, engineering, design, and drawing.

For example:
If a cube is rotated 90 degrees clockwise from the given illustration, what will it look like?

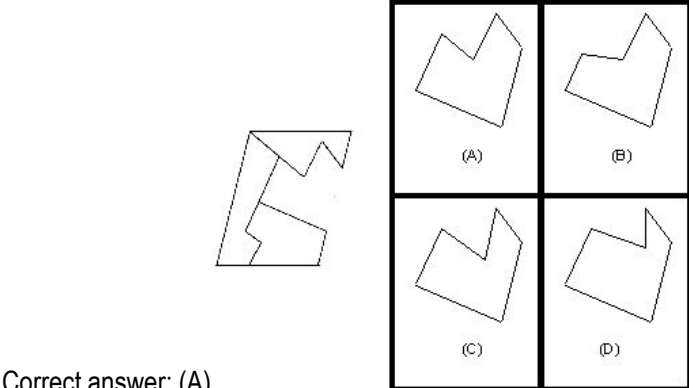


Correct answer: C

Examples of test that have orientation items are the Embedded Figures Test (EFT), Wechsler Adult Intelligence Scale-Revised (WAIS-R), Test of Visual Perceptual Skills: Third Edition (TVPS-3), Spatial Test Battery (STB), and some items of the OLSAT.

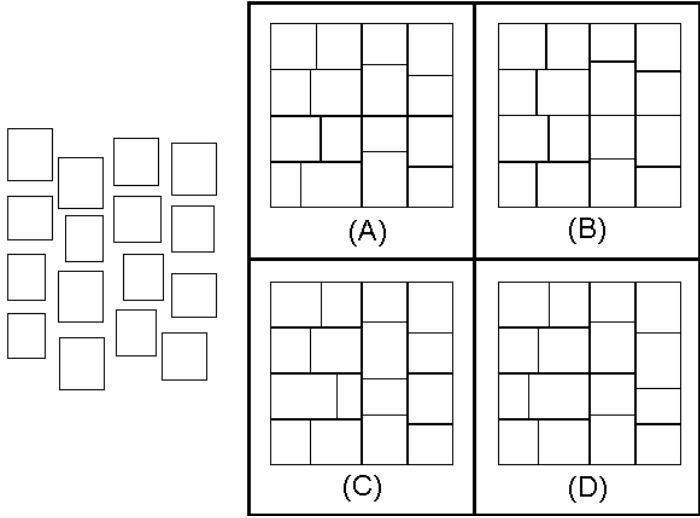
Figure and Ground Perception. The examinee identifies a specific component in a situation that involves shifts in perception of figures against complex backgrounds where interesting and hidden forms are used (Bengamin, Hopkins, & Nation, 1994).

For example:
Find the missing piece to complete the parallelogram.



Correct answer: (A)

What will the dismantled object look like when assembled?

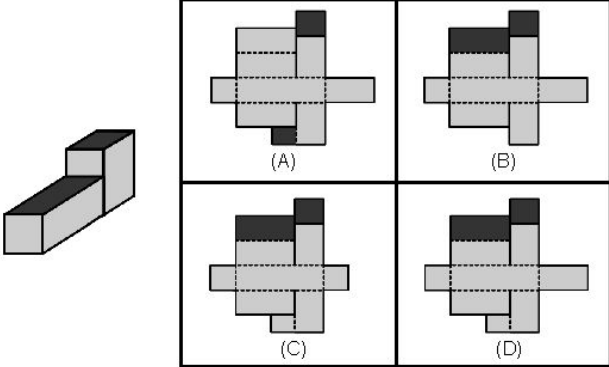


Correct answer: B

This test type requires the individual to find a common geometric shape within a larger, more complex design. Examples of tests with the figure ground perception are the cogAT, form 5, Group Embedded Figures Test (GEFT), Embedded Figures Test (EFT), Wechsler Adult Intelligence Scale-Revised (WAIS-R), and Test of Visual Perceptual Skills: Third Edition (TVPS-3).

Surface Development. The examinee is presented with a flat image that can be transformed into a three dimensional object when folded. The task is to find out the end product after folding. In this type of test, individuals are usually shown irregular, flat shapes, such as pieces of paper or cardboard, and boxes created by folding these shapes. It will be the person's task to find out what lines on the flat shapes correspond to what lines on the boxes. These skills are needed in drafting, mechanics, physics, and courses that have on analysis.

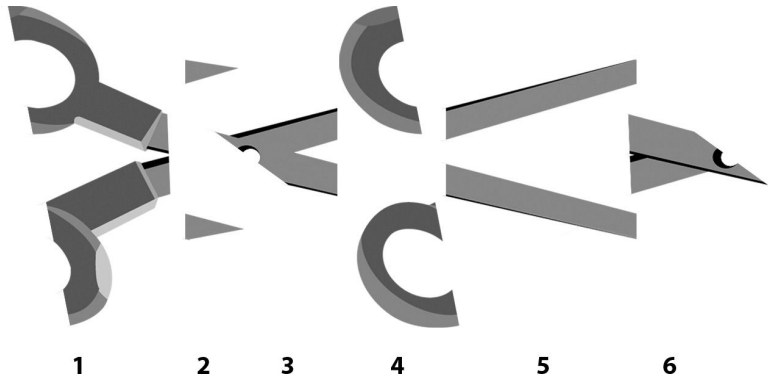
For example:
Which pattern can be constructed from the given figure?
Correct answer: D



Examples of test with surface development are the Cognitive abilities test (cogAT, form 5), Wechsler Adult Intelligence Scale-Revised (WAIS-R), Spatial Test Battery (STB), Test of Visual Perceptual Skills: Third Edition (TVPS-3), and Test of Visual Reasoning and Processing Skills: (TVRPS).

Object Assembly. The examinee is tasked to choose the order, from left to right, in which the parts should be placed to form the object. This skill is necessary in courses that require building and creating objects such as mechanical, electrical, drafting, and engineering.

For example:



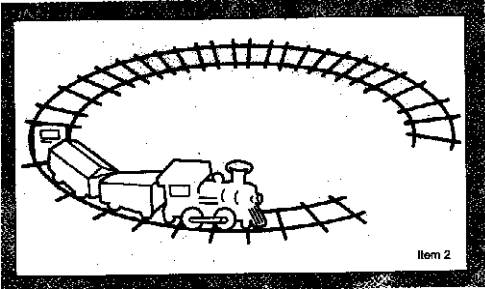
A) 145263
B) 143652
C) 413652
D) 416352

Correct answer: D

Test of object assembly measures visual analysis, synthesis, and construction. These skills are important in the fields of engineering, mechanics, automotive, and electronics. Tests with object assemblies are Wechsler Adult Intelligence Scale-Revised (WAIS-R) and Test of Visual-Reasoning and Processing Skills (TVRPS).

Picture completion. The examinee is tasked to identify the missing part from a whole object and select among the options any indicator of the missing part. This test type requires the examinee to identify what important part is missing from pictures or common objects and scenes. These skills are needed in content courses such as communication, social science, and humanities.

For example:
What letter does the missing part start with?



a. P b. T c. R d. S

Correct answer: c

This subtest measures visual scanning of printed materials. Examples of test that contain this test type are the WISC-III, WPPSI-R, and WAIS-III.

Evaluation of the Test Types

The test type presented is evaluated of its weaknesses in terms of its exclusivity, specificity, and generalizability of skills. Some tests use the verbal analogy, syllogism, and number and letter series that are not exclusively used in aptitude tests. These test types are also used in tests that measure achievement in Mathematics, English, and tests of general intelligence. Another is that a multitude of skills can be measured by each item type and it depends on the prerogative of stakeholders on which skill to select. The accurate skills related to these item test types should be framed. Majority of the items on the classification are nonverbal in nature and the items are artificial ways of pointing at certain skills. A test taker may have used a different cognitive and metacognitive skill to derive the answer which may not be intended with what the item is suppose to measure. Since the items are artificial representation of skills as they are not context specific to a defined field, work, or task, the problem of translation to real skills is a question. However, this is solved by determining the criterion-prediction validity for a future domain-specific performance.

Utility of the Taxonomy

The taxonomy provided is useful in three ways. First, it aids test developers in creating aptitude test items. The items can fit to majority of the available frameworks that can be used as basis on test development. The items also cover a wide range of test types that serve as a model in determining which type of test items to use. The taxonomy helps in operationalizing conceptual frameworks which becomes more well-defined (Powers & Dwyer, 2003). Second, the taxonomy provides the skills measured by a particular kind of test. Assessment frameworks usually begin with the competencies needed to be

developed whether in the educational or industrial setting. Once these competencies are framed, specific test types can be used and identified that can directly assess these competencies. Third, it guides in the selection of appropriate tests for various uses. Given the system of classification provided, stakeholders can decide a suitable ability test that would address their needs. The construction and selection of item types for an aptitude test is not characterized by its novelty and variety of test items. It should depend on the intended skill that needs to be measured.

Endnotes

¹The Center for Learning and Performance Assessment (CLPA) is based De La Salle-College of Saint Benilde, see www.dls-csb.edu.ph

²The Philippine Educational Measurement and Evaluation Association (PEMEA) is a professional organization of experts in the field of measurement that aims for promoting a culture of assessment.

³Details about the Asian Psychological Services and Assessment is found in <http://www.apsa-esba.com/mission.htm>

References

- Alreck, P. L., & Settle, S. B. (2004). *The survey research handbook*. Boston: McGraw-Hill/Irwin.
- Anastasi, A., & Urbina, S. (2002). *Psychological testing* (7th ed.). New Jersey: Prentice Hall.
- Colom, R., Escorial, S. Shih, P., & Privado, J. (2007). Fluid intelligence, memory span, and temperament difficulties predict academic performance of young adolescents. *Personality and Individual Differences, 42*, 1503-1514.
- Bengamin, L. T., Hopkings, J. R., & Nation, J. R. (1994). *Psychology*. New York: Macmillan College Publishing Company.
- De Raad, B., & Barelds, D. (2008). A new taxonomy of Dutch personality traits based on a comprehensive and unrestricted list of descriptors. *Journal of Personality and Social Psychology, 94*, 347-364.
- Ennis, R. H. (1987). A taxonomy of critical thinking dispositions and abilities. In J. B. Barron & Sternberg (Eds.), *Teaching thinking skills: Theory and practice* (pp. 9-26). New York: W. H. Freeman and Company.
- Kaplan, R., & Saccuzzo, D. (1997). *Psychological testing* (4th ed.). London: International Thomson Publishing Company.
- Lohman, D. F., & Hagen, E. (2001). *The cognitive abilities test Form 6: Interpretative guide for teachers and counselors*. Itasca, IL: Riverside Publishing.
- Owen, D. (1999). *None of the above: The truth behind the SATs*. USA: Rowman & Littlefield.
- Ozer, D.J. (1987). Personality, intelligence, and spatial visualization: Correlates of mental rotations test performance. *Journal of Personality and Social Psychology, 53*, 129 - 134.

- Powers, D. E., & Dwyer, C. A. (2003). *Towards specifying a construct of reasoning* (Tech. rep. no. RM-03-01). Educational Testing Service, Research and Development Division, Princeton, NJ.
- Rapaport, D., Gill, M., & Schafer, R. (1968). *Diagnostic psychological testing*. New York: International Universities Press.
- Raven, J. (1986). *Manual for Raven's Progressive Matrices and Vocabulary Scales*. London. H. K. Lewis.
- Sternberg, R., Castejon, J., Prieto, M., Hantamaki, J., & Grigorenko, E. (2001). Confirmatory factor analysis of the Sternberg triarchic abilities test in three international samples: An empirical test of the triadic theory of intelligence. *European Journal of Psychological Assessment, 17*, 1-16.
- Vázquez-Ramos, Leahy, & Hernández. (2007). The delphi method in rehabilitation counseling research. *Rehabilitation Counseling Bulletin, 50*, 111-118.
- Thurstone, L. (1938). *Primary mental abilities*. Chicago: University of Chicago Press.

Appendix A

Rankings of the Aptitude Test Items

	Sum of Ranks	Mean of Ranks
Syllogism	95	4.52
Visual Discrimination	102	4.86
Visualization	102	4.86
Orientation	110	5.24
Verbal Analogy	115	5.48
Number/Letter Series	116	5.52
Progressive Series	119	5.67
Figure Ground Perception	128	6.1
Surface Development	143	6.81
Topology	146	6.95
Object Assembly	147	7

Note. N=21

Author Notes

Further correspondence can be addressed to Dr. Carlo Magno, Counseling and Educational Psychology Department, De La Salle University-Manila, 2401 Taft Ave. Manila, Philippines. E-mail: carlo.magno@dlsu.edu.ph or crlmgn@yahoo.com.

Special thanks to Ms. Charisse Tan for compiling the sample items and Ms. Angeli Malaki for providing the examples of standardized tests for the taxonomy.

About the Author

Dr. Carlo Magno is presently an assistant professor at the Counseling and Educational Psychology Department at the De La Salle University, Manila, Philippines. His research interests are in the areas of educational psychology, measurement and assessment, and learning that employ quantitative techniques. Further correspondence can be addressed to him at carlo.magno@dlsu.edu.ph.