



**RESEARCH PAPER**

**Item Discrimination of Intelligence Test Items: A Quality Assurance Test for an Assessment Tool**

<sup>1</sup>Nighat Gul    <sup>2</sup>Dr. Sonia Shagufta    <sup>3</sup>Dr. Shagufta Parveen\*

1. PhD. Scholar, Department of Psychology, Shaheed Benazir Bhutto Women University Peshawar KPK Pakistan,
2. Head of the Department, Department of Psychology, Shaheed Benazir Bhutto Women University Peshawar, KPK, Pakistan,
3. Assistant Professor, Department Psychology, Hazara university of Mansehra, KPK, Pakistan

PAPER INFO	ABSTRACT
<p><b>Received:</b> February 28, 2022</p> <p><b>Accepted:</b> April 10, 2022</p> <p><b>Online:</b> April 15, 2022</p> <p><b>Keywords:</b> Assessment Tool, Discriminatory Power, Intelligence Test, Item Discrimination, Observational</p> <p><b>*Corresponding Author</b> Ali25_moh@yahoo.com</p>	<p>Item discrimination is an essential technique for making the item analysis for selecting the appropriate items for intelligence test. It helps to identify items which can be revised or discarded, thus building a quality items for IQ assessment tool. The study focussed on item analysis of 100 items of four intelligence types. These items of intelligence test conduct on 1500 people of different age groups adolescence to adulthood. Through item discrimination explored the items discriminatory power with the help of MS excel 2010 and SPSS. Through item analysis we rejected 23 items which item discriminatory power was zero, negative and less than 0.5. item discrimination range in present study was 0.5 -0.45. reliabilities of all types of the intelligence likewise linguistic, observational, mathematical and logical were .88, .80, .89 and .90 respectively. These reliabilities were higher and indicating the present tool of intelligence authentic and reliable. This is a quality assurance test for measuring the individuals intellectual capabilities.</p>

**Introduction**

Intelligence is one of the foremost quantifiable traits of human person which assumes a giant element in instructive, phrase associated, social and economic outcomes (Duckworth, et al., 2011). The task of nature and guide has been extensively taken into consideration and it's been presumed that it's miles a few combination of each. Assessment of a perfect diploma of information can reflect a novel practicality for intellectually soliciting for callings and on this way can count on ensuing fulfilment. Achievement in lifestyles bills in extensive scope of capacities and there may be without a doubt now no longer a solitary variable that can foresee lifestyles outcomes higher as compared to perception (Smith, Philips and Reichard, 2015). Insight is considered because the focal and maximum comprehended subject matter in mind technological know-how which has been a mark of dialogue in a actual experience for millennia, basically from the hour of Plato and Aristotle. In spite of the stretched records and concern of dialogue for big range of 12 months nonetheless the subject matter has severa questions. The factor is interesting and exciting because it offers information close to character capacities and offers information into why some humans are greater certified for particular positions (Kaufman et al., 2013).

The time period information has received a ways and extensive ubiquity on account that its start line as a philosophical concept to tremendously common-sense usage of distinguishing character with numerous scholarly capacities and putting them as in keeping with their studying possibilities. Knowledge is a complete detail of human person with heap

of recommendations and implications. Assessment of a perfect diploma of information can reflect a novel suitability for intellectually soliciting for callings and therefore may be beneficial in foreseeing fulfilment in schooling, paintings execution and different lifestyles outcomes (Duckworth, et al., 2011). The lengthy status troubles with perception because the unitary popular potential or a mixture of numerous insights is as but status out sufficient to be observed of scientist even after lengthy records and multiplied verbal exchange. Spearman (1904) proposed the unitary angle on information.

The degree to which an item correctly differentiates among test takers in the behaviour that the test is supposed to measure is referred to as item discrimination (Anasthasi and Urbina, 2004). It's a metric that determines how well an item can discriminate between educated and uninformed examinees, or between masters and non-masters. It is defined by Cohen and Swerdlik (2005) as a statistic that indicates how well a test item differentiates or discriminates between high and low scores.

Many indices are used in test construction theory to determine the property of item discrimination. Some of them presume that the underlining attribute has a normal distribution. Despite their differences in methodologies, the majority of item discrimination indices produce results that are very comparable (Anasthasi and Urbin, 2004).

He inferred that the wonderful contributory detail in execution on undertakings and assessments is the 'g' factor. In competition to unitary potential, Thurston (1938) added bunch speculation of information and proposed vital intellectual capacities. Thurston's speculation of information appreciably affected later speculations. Guilford (1988) subtle Construction of Knowledge (SI) version with a hundred and eighty capacities. He pointed that that information is not one strong first-rate but a mix of various capacities. Sternberg added triarchic angle on information even as Gardner (1983) gave speculation of severa insights which hypothesize that people are delivered into the sector with numerous diploma of information on 11 capacities.

With the development of first Binet Insight Test the shift of awareness from subjective to quantitative is considered because the giant development in discipline of information trying out. The maximum severe and valid evaluation got here from the minority gatherings and people from distinctive social orders who have been attempted making use of take a look at apart from their mother language. The manner of lifestyles decency, educability and monetary basis arose out as giant mark of query amongst take a look at engineers and take a look at takers, mainly with inside the international milieu. The abuse of perception assessments and distortion of outcomes likewise faced the brunt of professionals, analysts and different giant quarters of society (Gottfredson and Saklofske, 2009). It is relied on that hereditary enrichment and character weather each expect their element being evolved of perception. With the increasing notoriety, the information assessments have been likewise reprimanded for his or her biasness and misinterpretations in the direction of minorities. It became said that the differences in information are basically due to the maintain as opposed to the hereditary enrichments. What's greater, development in psychometrics and logical investigations undermined greater mounted views and information trying out performed the scenario with intelligence amongst informed laymen unbiased of shading, ideology and cast. However differential task of nature and guide in information is as but now no longer indisputable, however alternatively maximum professionals be given that there can be some predispositions but aren't kidding to the factor of imprinting the overall believability of perception assessments. The discipline of perception trying out is pushing in advance and being carried out international over with realistic fulfilment. The headway of psychometrics, prescient really well worth of perception assessments in training and enterprise blended with utility in human asset attractiveness

driven upon the want of greater reliable and full-size information assessments. During Universal Conflicts the want to check large portions of applicants made equipped for development of collecting manipulate successful information assessments for immediate comparing and enlistment for navy (Kaufman, 2009). With the development of time and coming approximately pundit a giant range of those predispositions had been diminished, essentially due to the presentation of normalized way of life truthful execution and nonverbal assessments. Albeit the excessive degree assessments have modern-day speculations of psychometrics but, analysts maintain on bantering on numerous hypothetical views and the verbal exchange is progressing. The vital ramifications of social reasonableness associated with the usage of intellectual assessments and translation desires specific contemplations mainly close to unknown dialect assessments (Goldstein et al., 2008). The following dialogue befell into the development of way of life truthful and unfastened trial of perception. In those conditions, the nonverbal perception assessments arose out because the sensible manner ahead as they restriction the effect of language functionality. It is proposed that the nonverbal take a look at are much less willing to lean closer to particular elegance or social basis (Cattell and Cattell, 1973; Freedle, 2010; Lewis et al., 2007). The usage of information in instructive settings is being applied for the reason that improvement of first Binet scale in 1904. Albeit, the usage of perception assessments started out with the very want to apprehend understudies who want notable tutoring but with the headway in trying out, the usage of information unfold to almost in all potential regions. The instructors, educationists and scientists are making use of collection of information assessments to apprehend understudies' proper potential in numerous potential regions. Various investigations brought about have a look at the relationship among perception rankings and scholastic fulfilment proved that information is extremely good indicator of scholarly fulfilment Intelligence testing is a lengthy process that assesses a person's capacity to comprehend concepts and words, reason abstractly, solve problems using various forms of reasoning, learn from feedback and experience, and process information in many modalities, such as visually or audibly. This examination technique usually yields the intelligence quotient (IQ). Intelligence testing is sometimes known as intellectual or cognitive evaluation, or IQ testing, all of which are interchangeable words. Before offering a brief history of intelligence testing in the United States, this section covers intelligence and theories of intelligence. It then examines why intelligence testing is employed, as well as the psychometric challenges surrounding intelligence assessment, intelligence testing across the lifespan, and intelligence's strengths and shortcomings, as well as common misconceptions regarding intelligence (Gorden, 2013).

### **Material and Methods**

This is the cross sectional study conduct on three stages of life people children, adolescence, adult and adulthood. Recruited total numbers of participants were 1500, while children, adolescences, adult and adulthood were hired 200, 500, 500 and 300 respectively by applying the simple random sampling technique Before conducting the research researcher get the approval to internal research committee. Before collecting the data to participants get the institutional consents and as well participants consents and informed them about purpose of study and your participation is purely voluntarily no monetary reward for your valuable participation. After collecting data to participants pay thanks of them, after completing the data collection phase we made the item analysis by using excel and SPSS by item discrimination and accept and reject the intelligent test items

**Results and Discussion**

**Table 1**  
**Items discrimination of intelligence test items**

	ID	Item no	ID	Item no	ID	Item no	ID
1	<b>0.20</b>	26	0.71	51	0.5	76	.83
2	<b>0.40</b>	27	0.72	52	0.73	77	.89
3	<b>0.29</b>	28	0.77	53	0.5	78	.86
4	<b>0.30</b>	29	0.87	54	0.73	79	.79
5	<b>0.34</b>	30	0.84	55	0.69	80	.82
6	0.50	31	0.82	56	0.7	81	.78
7	<b>0.44</b>	32	0.81	57	0.71	82	.71
8	0.56	33	0.83	58	0.52	83	.72
9	0.81	34	0.91	59	<b>0.4</b>	84	.77
10	0.84	35	0.67	60	<b>0.4</b>	85	.87
11	<b>0.34</b>	36	0.75	61	<b>0.2</b>	86	<b>0.4</b>
12	0.5	37	0.73	62	0.5	87	<b>0.2</b>
13	0.73	38	0.69	63	0.6	88	.81
14	0.69	39	0.78	64	0.6	89	.83
15	0.7	40	0.71	65	0.5	90	.91
16	.71	41	0.82	66	0.7	91	.67
17	0.52	42	0.84	67	0.52	92	.75
18	<b>0.4</b>	43	<b>0.04</b>	68	<b>0.4</b>	93	.73
19	<b>0.4</b>	44	<b>.034</b>	69	<b>0.4</b>	94	.69
20	<b>0.2</b>	45	<b>.05</b>	70	<b>0.2</b>	95	.78
21	0.5	46	0.6	71	0.5	96	.71
22	0.6	47	0.5	72	0.6	97	.82
23	0.6	48	0.5	73	0.6	98	.84
24	0.5	49	<b>.045</b>	74	0.5	99	.83
25	0.7	50	<b>.07</b>	75	0.7	100	.89

Table is depicting the item discrimination of the all items of IQ test items all those items were accepted for second draft those item discrimination was less then 0.5 as well was zero and negative. 23 items were rejected due to poor discriminatory power of the items and modified the items of IQ test for developing reliable construct. Rejected items numbers of the intelligence test were 1, 2, 3, 4, 5, 7, 11, 18, 19 20, 43, 44, 45, 49, 50, 59, 60, 61, 68, 69, 70, 86 & 87 these items ID range was 0.5-.45.

**Table 2**  
**Alpha reliability of different types of intelligence**

Factors	Alpha
Linguistic IQ	.88
Observational IQ	.80
Mathematical IQ	.89
Logical IQ	.90

Table II alpha reliability for linguistic IQ, Observational IQ, Mathematical IQ and logical IQ were .88, .80, .89 . and .90 respectively. These reliabilities are showing all types of intelligence are reliable and valid for measuring the intelligence.

## **Discussion**

Study tries to develop of appropriate difficulty level and item discrimination items for intelligence test. Table 1 depicting the items discriminatory power of the intelligence test on the base of items discriminatory power researcher reject the all those items those ID was less then 0.5 also discriminatory power was zero and negative . in present study rejected items were 23 these items discriminatory power was poor these were not able to make the discrimination in higher intelligent and low intelligent people's. Lange (1967) argued item analysis is the best method for taken the decision about selection and rejection of the test items. This study examined the quality of English test items using psychometric and linguistic characteristics among Grade Six pupils. Contrary to the conventional approach of relying only on statistics when investigating item quality, this study adopted a mixed-method approach by employing psychometric analysis and cognitive interviews. The former was conducted on 30 Grade Six pupils, with each item representing a different construct commonly found in English test papers. Qualitative input was obtained through cognitive interviews with five Grade Six pupils and expert judgements from three teachers. None of the items were found to be too easy or difficult, and all items had positive discrimination indices. The item on idioms was most ideal in terms of difficulty and discrimination.

Difficult items were found to be vocabulary-based. Surprisingly, the higher-order-thinking subjective items proved to be excellent in difficulty, although improvements could be made on their ability to discriminate. The qualitative expert judgements agreed with the quantitative psychometric analysis. Certain results from the item analysis, however, contradicted past findings that items with the ideal item difficulty value between 0.4 and 0.6 would have equally ideal item discrimination index ( Shanmugam, 2020)

The Item Discrimination Index, or D, examines how well an item may distinguish between better and less able students (Mehta & Mokhasi, 2014). An index value of +1 indicates that the item is extremely effective, whilst a value of 0 shows that the item is completely incapable of discrimination. If the discrimination index is negative, the item is incorrect, as more students in the lower ability group were able to select the correct responses more frequently than students in the higher ability group (Bichi & Embong, 2018)

## **Conclusion**

On the base of findings of the research we concluded selected items of the intelligent test posses the appropriate discriminatory power. These items can make the discrimination between the high intellectual capabilities and low intellectual capabilities. Alpha reliability of all types of intelligence is higher and acceptable for a reliable test. Thus, intelligent test is reliable construct for measuring the intelligence.

## References

- Allan Lange, Irvin J. Lehmann, William A. & Mehrens. (1967). Using Item Analysis to Improve Tests. *Journal of Educational Measurement*, 4(2).65-68
- Suppiah Shanmugam, S. K., Wong, V., & Rajoo, M. (2020). Examining the quality of English test items using psychometric and linguistic characteristics among grade six pupils. *Malaysian Journal of Learning and Instruction*, 17(2), 63-101.
- Kaufman, J. C., Kaufman, S. B., & Plucker, J. A. (2013). Contemporary theories of intelligence. In D. Reisberg (Ed.), *The Oxford handbook of cognitive psychology* (pp. 811–822). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780195376746.013.0051>
- Duckworth, A. L., Quinn, P. D., & Tsukayama, E. (2012). What No Child Left Behind leaves behind: The roles of IQ and self-control in predicting standardized achievement test scores and report card grades. *Journal of Educational Psychology*, 104(2), 439–451. <https://doi.org/10.1037/a0026280>
- Bruce B. Frey. (2018). *Intelligence test*. The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation, <https://dx.doi.org/10.4135/9781506326139.n338>
- Gardner, H. (2013). *Frequently asked questions—Multiple intelligences and related educational topics*. [https://howardgardner01.files.wordpress.com/2012/06/faq\\_march2013.pdf](https://howardgardner01.files.wordpress.com/2012/06/faq_march2013.pdf)
- Thurston, L. L. (1938). *Primary Mental Abilities*. Chicago: University of Chicago Press.
- Anastasi, A., & Urbina, S. (2004). *Psychological Testing, 7th Edition*, Pearson Education, New Delhi.
- Cohen, R. J., & Swerdlik, M. E. (2005). *Psychological Testing and Assessment: An introduction to tests and measurement*. McGraw Hill- New York
- Guilford, J. P. (1982). *Psychometric Methods*. Mc Graw Hill , New York
- Ashraf ZA, Jaseem K. (2020). Classical and modern methods in item analysis of test tools. *International Journal of Research and Review*. 2020; 7(5): 397-403.
- Kaplan, R. M., & Saccuzoo, D.P. (2001). *Psychological Testing - Principles, Applications and Issues -5 th Edition*, Wadsworth, Stanford
- Mehta, G., & Mokhasi, V. (2014). Item analysis of multiple-choice questions - an assessment of the assessment tool. *International Journal of Health Sciences and Research*, 4(7), 197-202
- Bichi, A. A. (2015). Item analysis using a derived science achievement test data. *International Journal of Science and Research (IJSR)*, 4(5), 1656- 1662.
- Cattell, R.B. (1973). *Personality and Mood by Questionnaire*. San Francisco: Jossey-Ba
- Freedle, Roy O. (2018). On Replicating Ethnic Test Bias Effects: The Santelices and Wilson Study, *Harvard Educational Review*, 80(3): 394-403,435.

- Lewis, S., Gambles, R. & Rapop, R. (2007). The constraints of a 'work-life balance' approach: An international perspective. *The International Journal of Human Resource Management*, 18(3):360-373 [http://: DOI: 10.1080/09585190601165577](http://doi.org/10.1080/09585190601165577)
- Gottfredson, L., & Saklofske, D. H. (2009). Intelligence: Foundations and issues in assessment. *Canadian Psychology/Psychologie canadienne*, 50(3), 183–5. <https://doi.org/10.1037/a0016641>
- Goldstein, R., Almenberg, J., Dreber, A., Emerson, J., and Herschkowitsch, A. (2008). Do More Expensive Wines Taste Better? Evidence From A Large Sample of Blind Tastings. *Journal of Wine Economics*, 3(2), 1–9.
- Strindberg, August (1995). *McLeish, Kenneth (ed.)*, Miss Julie, London: Nick Hern Books