REVIEW OF SOCIETY EDUCATION WORK

ISSN: 2507-749X EISSN: 2602-7755

https://www.asjp.cerist.dz/revues/414

Volume :09/N°:01(2024) PP 393/404

Grading test of logical reasoning using the Rasch Wright map

تدريج اختبار التفكير المنطقي باستخدام خريطة (راش) رايت

SNOUSSAOUI Okkacha Bachir *
University of Oran2, (Algeria)
snoussaoui.bachir@univ-oran2.dz

HEBALI Karima
University of Oran2, (Algeria)
hebali.karima@univ-oran2.dz

Abstract :

The study implemented the logical reasoning test and measured its level of difficulty, The Rasch Wright map was used to facilitate the scaling of scores and to establish the level of difficulty of each item in the test, using the WINSTEPS statistical software based on a sample of 318 employees. The results showed that the average levels of the item difficulty of the test were very close to the levels of sample abilities, as well as the classification of the test from the easiest item to the most difficult, which allows us to use the test with a certain warranty.

Article info

Received :29/09/2023

Accepted :28/05/2024

Keyword :

- ✓ *Grading test*
- ✓ *Logical reasoning*
- ✓ *Rasch model*

ملخص:

في هذه الدراسة ثم تطبيق اختبار التفكير المنطقي وقياس مستوى صعوبته، ثم استخدام خريطة راش رايت لتسهيل تحديد مستوى صعوبة كل فقرة من الاختبار باستخدام البرنامج الإحصائيWINSTEPS على عينة مكونة من 318 موظفا. أظهرت النتائج أن المتوسطات متقاربة جداً بين مستوى صعوبة الاختبار ومستوى قدرات العينة، كذلك سهولة تصنيف فقرات الاختبار من الفقرة الأسهل إلى الأصعب، وهذ اما يسمح لنا باستخدام الاختبار بثقة.

معلومات المقال

تاريخ الارسال:2023/09/29

تاريخ القبول:2024/05/28

الكلمات المفتاحية:

- ✓ *تدريج الاختبار*
- ✓ *التفكير المنطقي*
- ✓ *نموذج راش*

*Corresponding author

❖ **Introduction :** The assessment of individual's cognitive abilities is critical in various fields, including education and psychology. There are several approaches to measuring human cognitive skills, and one such method is the use of psychometric tests such as Raven's test. In recent years, the Rasch model has gained popularity as a powerful tool, for analyzing data obtained from psychometric testing. The Rasch model enables researchers to estimate individual ability levels, based on their responses to test items and provides insight into item difficulty and discrimination; studies have investigated the applicability of different variants of the Rasch model, for calibrating and scaling test, forms with dichotomous or non-dichotomous response options. For instance, Gallini applied the Reduced SPM version of Raven's tests, to examine testing time constraints on fitting Student Progress Monitor, items to a Rasch model. Similarly, Dossar & Mesbah demonstrated that while both Partial Credit Model and Graded Response Model variants could be useful in practice, there were differences between them when assessing latent traits.

The Rasch measurement model can be used to process the examination data and provide answers to the following questions:

- What are the reliability indices when calculating person/item Cronbach Alpha value, Cronbach Alpha (KR-20) value and separation value?
- What is the level of difficulty of the items of the test?

To respond to the question, we must first understand the Rasch measurement model and WINSTEPS program to analyze data.

❖ **Intelligence testing:**Intelligence testing generally refers for charmorroer furnham (2005) to the method used to gauge a person's cognitive abilities and intellectual potential. Traditional intelligence tests, such as IQ tests, typically measure aspects like problem-solving abilities, analytical thinking, and abstract reasoning. However, some critics have argued that these tests mainly measure academic or "book smart" abilities and do not account for other forms of intelligence. The idea of 'multiple intelligences', advocating for various forms of intellectual ability was proposed by Guilford, who described up to 150 different types of abilities based on operations, products, and content categories. Yet, there is continued support for the positive manifold of correlated ability test scores, i.e., the original 'G' hypothesis of general intelligence, due to its predictive validity.

❖ **Theories of intelligence:**Bernaud (2013) report that Psychologists have proposed several theories of intelligence, these theories aim to explain the nature and structure of intelligence and provide frameworks for understanding and measuring it:

- **Spearman's unifactorial conception:** This theory postulates that intelligence is primarily determined by a general factor, denoted "G", which concerns all intellectual tasks. Individual differences in this "G" factor would lead to differences in intellectual performance.

- **Multifactorial conceptions**: These theories reject the idea of a single general factor and suggest that intelligence is better understood as a set of independent specific skills. An example would be Thurstone's theory, which identifies seven primary factors of intelligence.
- **Sternberg's theory**: This theory conceptualizes intelligence as consisting of three parts: analytical, creative, and practical intelligence. Each of these parts is necessary for successful intellectual performance.
- **Gardner's theory of multiple intelligences**: This theory proposes that intelligence is not a single innate ability, but rather a set of multiple independent intelligences, including linguistic, logical-mathematical, spatial, musical, bodily-kinesthetic, interpersonal, intrapersonal, and naturalistic intelligence.

Each of these theories offers a different view of what intelligence is and how it can be measured.

✓**Personality Traits and Intelligence:**According to Chamorroet Furnham (2005) Research has shown that certain personality traits can have an impact on IQ test performance. For instance, traits of the Five Factor Model (FFM), or the "Big Five" openness; conscientiousness; extraversion; agreeableness; and neuroticism - can influence how a person performs on such tests. For instance, 'openness to experience', which includes characteristics like imagination and insight, has shown a positive relationship with IQ scores. This relationship occurs because this trait involves being open to new ideas and being able to think abstractly, which are abilities tested in IQ tests.

However, the relationship between personality traits and IQ test performance is complex and can be influenced by a variety of other factors, including motivation, test-taking abilities, and the specific design of the IQ test.

✓ **Logical reasoning:**For Dowden (2019) Reasoning logically involves using clear and sound logical principles to conclude. This involves structured thinking, where one concept logically flows to the next, leading to a conclusion that is reasonable based on the provided evidence or premises. Logical reasoning is deeply rooted in critical thinking and helps in decision-making processes by providing a structured approach. There are various types of logical reasoning, including deductive, inductive, and abductive reasoning, each with its own rules and processes.

Logical reasoning and intelligence, though interconnected, refer to two different concepts.

- Logical reasoning is a sub-component of intelligence. It refers to the ability to analyze information and solve problems systematically and logically. These processes include deducing new information from already-known facts, evaluating arguments, and establishing solutions to issues. This ability can be developed and improved over time through learning and practice.

- Intelligence, on the other hand, is a broader concept that encompasses a range of cognitive capabilities, not just logical reasoning. Definitions vary, but intelligence generally refers to the human ability to understand complex ideas, learn from experience, apply knowledge to new situations, adapt to changing environments, and engage in various forms of reasoning. It includes multiple facets like spatial intelligence, social intelligence, emotional intelligence, linguistic intelligence, and more. Intelligence is traditionally assessed using IQ tests, though these do not capture all aspects of human intelligence.

In essence, logical reasoning is a specific component of the broader concept of intelligence.

- **The main tests of intelligence:** Bernaud (2013) affirmed that the main tests of intelligence include the Stanford-Binet Intelligence Scale, Wechsler Intelligence Scale for children and adults, Cattell intelligence and culture and the Raven's Progressive Matrices. These tests, among others, are designed to measure various aspects of intelligence, including logical reasoning, problem-solving, verbal comprehension, perceptual organization, and processing speed:

  - Stanford-Binet Intelligence Scale: This test originated as a method to identify intellectually challenged children. Today, it assesses intelligence and cognitive abilities in individuals from two to 85+ years of age. It evaluates five factors: knowledge, quantitative reasoning, visual-spatial processing, working memory, and fluid reasoning.

  -Wechsler adult intelligence Scale (WAIS): The WAIS is designed to measure a person's intellectual ability in both verbal and performance capacities. It is mostly used with adults but has variations for different age groups.

  -Cattell clture fair III: This test attempts to measure a person's intelligence in a way that is least affected by the cultural and environmental background. The test includes four subtests each in two sections: verbal and non-verbal.

  -Raven progressive matrices: This non-verbal group test was originally developed to measure educational aptitude. It assesses observation skills, clear-thinking ability, and intellectual capacity without language and reading skills coming into play.

  According to Khairani et al, (2020) The Raven’s test have gained significant attention in the field of education and assessment in the world for many years because it is a standardized psychometric test typically used to measure abstract reasoning and fluid intelligence. This non-verbal multiple-choice test presents visual patterns with a piece missing, and the examinee is asked to identify the missing piece from several

options. It is designed to minimize the influence of cultural or language biases, making it widely applicable across different populations. The "progressive" aspect refers to the test's structure where items become increasingly complex as the test progresses.

Nurhudaya et al, (2019) added that Advanced Progressive Matrices (APM) are part of Raven's Progressive Matrices, which are multiple-choice IQ tests. British psychologist Dr John C. Raven developed these tests in 1936. The APM, specifically, is a set of 48 matrix problems, ordered by increasing difficulty.

This carrectéristic of the APM test is what makes it a valuable tool in assessing general intelligence, as it is less influenced by cultural or language biases compared to other intelligence tests and is used in the world but after thoroughly evaluating its reliability and validity through techniques such as Rasch analysis.

❖ Rasch analysis:

- ✓ **Rasch analysismodel and classical test theory:** Nurhudaya et al (2019)point out that In statistical methods, Rasch models hold à significant place due to their inherent special measurement properties. These models allow for the transformation of discrete response data into useful, interpretable measurements, making them essential tools in fields such as psychology, sociology, and education. The beauty of the Rasch models is that they provide an objective measurement, which is independent of the specific sample of items. Moreover, Rasch models impose strict conditions on data, which, if met, allow individual item calibrations and person measures to be independent of each other. This effectively means researchers can compare measurements from different tests and samples, making this model extremely advantageous for comparing individual or group performances across different tasks, time points, and measurement tools. Furthermore, if data do not conform to the Rasch model, it implies there is a need for further work on the substantive problem of scale construction, not on the identification of a more complex model to fit the data. This distinctive feature makes Rasch models particularly attractive for measurement in the social sciences. Lastly, Rasch models also provide à valuable diagnostic tool to identify problematic items within a scale. In summary, Rasch models provide strictly invariant measurement, which means the comparison between person measures is independent of the particular items used, and the comparison between items' difficulty is independent of the people used. However, CTT does not have this property. Instead, CTT scores depend on the specific items used and the group taking the test. Therefore, scores from one test are not comparable with those from another.

The Rasch model in Rasch analysis, as mentioned earlier, provides valuable insights into the measurement properties of a test or survey.

- ✓ **Grading Test Of Reasoning Skills With Rasch Wright Map:**To grade a test of reasoning skills using the Rasch model and Wright map, several steps need to be followed. First, it is necessary to understand the Rasch measurement model and the WINSTEPS program. Second, pilot data should be collected and à Rasch analysis should be conducted using the WINSTEPS program. Third, Presentation of results and discussion.

First:

❖ Understanding the Rasch measurement model:

- **The Rasch Model Definition:**Maclean et al, (2005) mention that The Rasch model, named after Danish mathematician Georg Rasch, is a model used in the field of psychometrics, which is the science of measuring mental capacities and processes. It is an ideal model that converts qualitative data into quantitative data. The Rasch model expects that the probability of a correct response to a test item is a logistic function of the difference between the person's ability and item difficulty. It's especially useful in the design and analysis of tests, particularly in education and psychology. The Rasch model represents a transformational approach to understanding test scores, fundamentally changing how we interpret a person’s score on a test. This model was developed to remedy problems of subjectivity in interpretation associated with classic test theory, moving towards an objective, interval-level measurement.

  There are several key tenets of the Rasch model. One key element is the idea of "specific objectivity". This means that a person's measure on a given trait (like ability or attitude) is independent of the particular set of items used to measure that trait. This is an essential property for making fair comparisons between individuals or groups.

  Another significant aspect is the measurement of both person ability and item difficulty on the same linear scale. The model presents a logistic function expressing the probability of item success, contingent on the difference between person ability and item difficulty.

- **Understanding Person and Item Reliability Separation:**In the context of the Rasch model, the term "item" refers to à question or statement in a test, quiz, or survey that the respondent needs to answer. The "difficulty" of the item refers to how challenging it is for respondents, which can vary according to the complexity of the item.

  The term "person" refers For Bonne (2016) to the individual respondent who is answering the items on the test or survey. The "ability" of the person refers to their innate capacity or knowledge in relation to the variable under consideration. This ability can vary from person to person and can influence how they respond to the items in the test.

Reliability in this context generally refers to the consistency of measures. Item reliability indicates how consistently the items rank the respondents, i.e., whether the items yield similar rankings of respondents if used in repeated measures. Person reliability, on the other hand, relates to the consistency of the responses given by each respondent across items or over time. A high person reliability suggests that people's responses to different items are consistent and there is less measurement error.

The Rasch model is used to examine both item difficulty and person ability, allowing for the generation of measures that are independent of the specific sample of items or respondents. This can potentially provide more accurate and reliable measurements than some other traditional methods.

The person separation reliability is an estimate according to Bond, Fox (2015) as how well one can differentiate persons on the measured variable. That is, it estimates the reliability of the person placement across other items measuring the same construct. The estimate is based on the same concept as Cronbach's alpha. That is,the fraction of observed response variance that is reproducible.

Item reliability and item separation refer to the ability of the test to define a distinct hierarchy of items along the measured variable. The higher the number, the more confidence we can place in the applicability of item placement across other samples.

- **Wright's Map:**The term "Wright's Map" generally refers for Stone, Wright, & Stenner (1999) to a visual representation used in Rasch analysis, named after Benjamin D. Wright, one of the key contributors to Rasch modelling. This map enables researchers to simultaneously visualize and compare the distribution of person abilities and item difficulties along à common logit (log-odds units) scale. This shared scale forms the basis for measurement in Rasch analysis. On a Wright Map, respondents and items are placed on the same vertical continuum, representing the logit scale, essentially showing the person's ability and the difficulty of the items. If a person and item have the same logit value, there is a 50% probability that the person will succeed on the item. It offers analysts a way to investigate the alignment between the person's abilities and item difficulties, identifying any gaps or patterns.

- **The winstep software program:**Linacre (2023) difined Winsteps as a statistical software program that is used for performing Rasch analysis. Rasch analysis or Rasch modelling is a form of probabilistic test theory used in the creation and analysis of assessments or questionnaires. The software takes in data and runs it through the Rasch model, resulting in parameters that can be used for statistical analysis and reporting. This is useful in fields such as psychometrics, health outcomes, market research and any field that requires the creation or interpretation of measurements.

Second:

Boone (2016) report that pilot data should be collected and a Rasch analysis should be conducted using the Winsteps program to evaluate the functioning of the measurement instrument. After recording the specific data in the study by SPSS statistical programs, we convert it to WINSTEPS and then build a file for statistical analysis through Rasch. The pilot data and Rasch analysis with Winsteps can provide insights into the reliability of the test in terms of person and item measures. The person reliability values of the test, which indicate the consistency and accuracy of the measurements of individual abilities, can be determined using the Rasch Wright map and the Rasch model. Person reliability values of the test can also be assessed using (KR-20) Cronbach's coefficient alpha. Furthermore, person and item reliability and separation indices can be calculated to assess the quality of the test.

Third:

Presentation of results and discussion: The results of the Rasch analysis using WINSTEPS should be presented in the form of a Wright map, which depicts the items arranged according to their difficulty levels and the persons positioned according to their level of competency.

❖ Methods:

✓ **Research Methodology:**In this study, we use the Rasch Wright model to grade the version PMA (progressive matrice advanced) to grading the test of reasoning skills and SPSS, winsteps program to analysis the data.

✓ Sample:

- **Respondents*:*** A sample of 318 employees was selected randomly from Algerian company «Sonatrach», the distribution of study is shown in table 1- as follow:

Table 1.Frequency distribution of the sample according to sex

Sex	Frequency	The percentage
Male	190	59.7%
Female	128	40.3%
Total	318	100%

Source: SPSS V 25

The sample consisted of 318 respondents, including 190 males (59.7%) and 128 females (40.3%).

- ✓ **Measures:**Descriptive statistics were examined using SPSS Table - 1, and WINSTEPS to Analysis the data (Rasch Measurement using Winsteps).

  Rasch analysis was then conducted on the 48 items of the test Raven using Winsteps. In general, the Rasch model is the simplest and easiest-to-use single-parameter model.

- ✓ **Expected Results and Discussion:**

To answer this question 1:What are the reliability indices when calculating person/item Cronbach Alpha value, Cronbach Alpha (KR-20) value and separation value?

We need to Analysis and explain the data reliability and separation analysis, the result are in table 2:

Table 2. Person/ item separation and reliability

PERSON	318 INPUT	318 MEASURED			INFIT		OUTFIT	
	TOTAL	COUNT	MEASURE	REALSE	IMNSQ	ZSTD	OMNSQ	ZSTD
MEAN	73.7	48.0	.17	.44	.98	-.1	1.21	.1
S.D.	5.8	.1	1.02	.04	.29	1.3	1.22	1.2
REAL RMSE	.44	TRUE SD	.92	SEPARATION	2.08	PERSON RELIABILITY		.81
ITEM	48 INPUT	48 MEASURED			INFIT		OUTFIT	
	TOTAL	COUNT	MEASURE	REALSE	IMNSQ	ZSTD	OMNSQ	ZSTD
MEAN	488.1	318.0	.00	.19	.99	-.2	1.32	.6
S.D.	105.1	.1	2.31	.07	.07	.9	1.11	1.7
REAL RMSE	.21	TRUE SD	2.30	SEPARATION	11.20	ITEM RELIABILITY		.99

CRONBACH ALPHA (KR-20) PERSON RAW SCORE "TEST" RELIABILITY = .82

Person separation is used to classify people. Low person separation (< 2, person reliability < 0.8) with a relevant person sample implies that the instrument may not be not sensitive enough to distinguish between high and low performers. More items may be needed. Item separation is used to verify the item hierarchy.

Low item separation (< 3 = high, medium, low item difficulties, item reliability < 0.9) implies that the person sample are not large enough to confirm the item difficulty hierarchy (= construct validity) of the instrument. In general, test Reliability reported by Classical Test Theory Cronbach's Alpha, KR-20 is higher than Rasch Reliability.

The value of Cronbach's Alpha, KR-20 in this study is 0.82 and it is very good. The Person Reliability value is 0.81 so it is good; the Item Reliability value is 0.99 so it

is excellent or special. The value of person reliability is high and the item reliability value is excellent, indicating that the consistency of student answers are good and the quality of the items in the instrument is very good. The person separation value 2.08 indicated that the people who were included were sufficiently separated from one another. The device was able to distinguish between and separate the objects, as evidenced by the item separation figure of 11.20 the participant hierarchy (item map) shows how the participants and objects fit together on a continuum.

The first conclusion on reliability and separation states that the test has a high internal consistency reliability and a good separation of item difficulties.

To answer to the second question that it says: What is the level of difficulty of the items of the test?

After analyzing the data with the WINSTEPS program, we extracted the Wright map in the next figure:

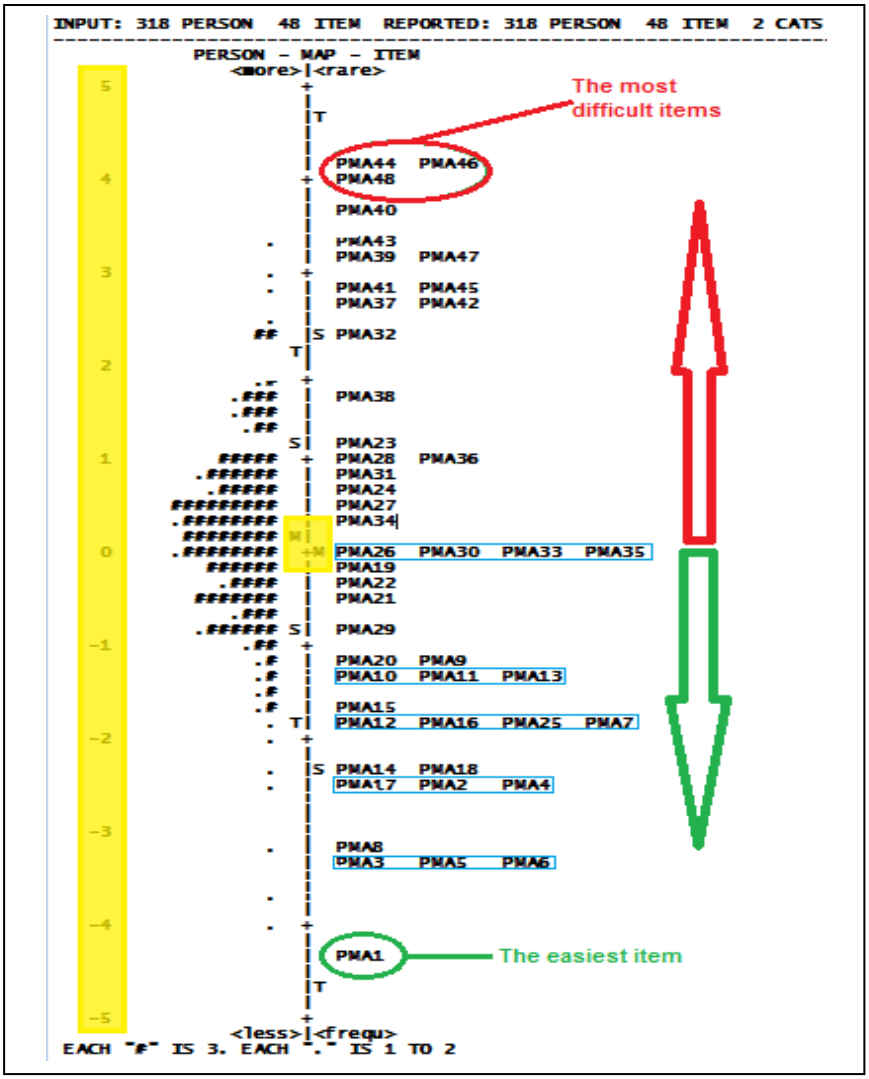

Figure 1. Wright -map of persons ability and items difficulty

Fig.1 represents Wright's map of 318 Individual, the first observation is the approaching averages symbolized by **M** on the right and left sides (in yellow) The average difficulty of the items appears in **M** on the left side, it corresponding to (0) Logit, which means the possibility of answering the items is 50%, average capacity should not be away from average difficulty than (1) logit, so that the test is not too difficult or easy to compare with the capabilities of individuals, this is illustrated in the figure, as we note that the average abilities approach the average items by less than half logit, and this convergence suggests that the items are difficult as close to the capabilities of the sample in this study. That is, testing is not very difficult and not very easy, it is in favor of individuals' abilities. The upper part of the shape (red) in the roofing effect (Ceiling effect), is resulting of difficult items, to which no member of the sample has succeeded or answered correctly due to its difficulty, the items: (44, 48, 46). Gap (in orange) its meaning that some abilities of individuals are not covered. The (green color) at the average of the shape observes the most easy item, that has been properly answered by all members of the sample, such as the item (1). This part is called the base of floor (floor effect). Extension any item must cover all capabilities difficulty level (low, medium, high).

Also in the line of groups (in blue) for example: (26, 30, 33, 35) and another group (25, 7, 16, 12). These groups in blue indicate the possibility of combining information with the group to a certain level of difficulty. Some items could be deleted to reduce the recurrence phenomenon, particularly when the number of items was large on the same line.Through the previous map, we can read and deduce the difficulty and ease of the item, the level of its coverage of an individual’s abilities and the way it spreads, which enables us to modify or delete the item that helps us to sensitize the test, which is an important element of structural reliability.

Over the average and about distribution, it is close to normal for individuals, and for the distribution of items, it should be a vertical line. With regard to targeting, where most items are preferred to correspond to individuals capabilities, at least one item should exist, some capabilities of individuals are not covered, and many items versus the level of capabilities of individuals.

In the second conclusion, we can gain valuable insight into the difficulty and ease of each item of the test and the level of coverage of the individual's abilities. By analyzing this point, we can identify items that may need to be changed or removed to improve the sensitivity of the test, by improving the structural reliability of the test; we can increase its validity and usefulness in accurately assessing an individual's cognitive abilities.

❖ **Conclusion:**Grading the logical ability test using the Rasch Wright map shows that most items cover almost all abilities at low, medium and high levels. Nevertheless, there are also many items with a very high level of difficulty. We can reserve these items for people with a rare high level of logical ability, such as highly cognitive employees or students. For the low difficulty items, we can delete some items

that are very easy to minimize the number of items and the response time, the analysis and results show that the test of reasoning ability is characterized by reliability and sensitivity, which qualifies it as useful and adaptable within the scope of the permissible and qualified testing law.

❖ Bibliography :

1. *Bernaud, J.L. (2013). Tests Et Théories De L'intelligence. Paris: Dunod.*

2. *Bond, TG & Fox, CM. (2015). Applying the Rasch Model: Fundamental Measurement in the Human Sciences. Routledge: New York.*

3. *Boone, W. J. (2016). Rasch Analysis for Instrument Development: Why, When, and How? Life Sciences Educaion, 15(4).*

4. *Chamorro-Premuzic, T., & Furnham, A. (2005). Personality and intellectual competence. New York: Lawrence Erlbaum Assocaites Publishers*

5. *Dowden, B. H.. (2019). Logical Reasoning.California State University Sacramento. Accessedfrom:https://www.csus.edu/indiv/d/dowdenb/4/logical-reasoning-archives/Logical-Reasoning-2020-05-15.pdf(consulted on 13/06/2022)*

6. *Khairani, A. Z., Matore, M., Effendi, M., & Khairani, M. Z. (2020). Psychometric Properties of Raven's Advanced Progressive Matrices with a Sample of Malaysian Youth. Pertanika Journal of Social Sciences & Humanities, 28(1), 267–277.Accessed from:http://journals jd.upm.edu.my/resources/files/Pertanika%20PAPERS/JSSH%20Vol.%2028%20(1)%20Mar.%202020/17%20JSSH-1908-2016.pdf (consulted on 12/01/2023)*

7. *Linacre, J. M.(2023). A User's Guide to Winsteps MiniStep Rasch-Model Computer Programs. Accessed from:https://www.winsteps.com/winman/copyright.htm(consulted on 24/01/2020).*

8. *Maclean, R., Watanabe, R., Baker, R., Boediono., Cheong Cheng, Y., Duncan, W., Keeves, J., Mansheng, Z., Power, C ., Rajput, J. S., , Thaman, K- H., Alagumalai, S ., . Curtis, D- D., Hungi, N. (2005). Applied Rasch Measurement: A Book of Exemplars. Asia-Pacific Educational Research Association : Springer.*

9. *Stone, M. H., Wright, B. D., & Stenner, A. J. (1999). Mapping Variables. Journal of Outcome Measurement, 3(4); 308-322.*

10. *Nurhudaya, T.A; Yudha, E.S, Suryana, D.(2019).The Raven's Advanced Progressive Matrices in Education Assessment with a Rasch Analysis. Universal Journal of Educational Research, Vol. 7, No. 9, pp. 1996 - 2002, Accessed from: DOI: 10.13189/ujer.2019.070921.*