Pak. j. life soc. Sci. (2025), 23(1): 3829-3835 E-ISSN: 2221-7630;P-ISSN: 1727-4915

Pakistan Journal of Life and Social Sciences
www.pjlss.edu.pk

https://doi.org/10.57239/PJLSS-2025-23.1.00301

RESEARCH ARTICLE

Influences of Item Responses Theory on Psychological Assessments

Izzeldeen Abdullah Alnaimi*

*Professor of Measurement and Evaluation, College of social Sciences, Psychology Department, Imam Mohammad Ibn Saud Islamic University (IMSIU), Saudi Arabia

ARTICLE INFO

Received: Nov 10, 2024

Accepted: Jan 27, 2025

Keywords

Item Response Theory

Psychological

Social

Cognitive

Psychometric

Validity and Reliability

***Corresponding Author:**

eaalnuaimi@imamu.edu.sa

ABSTRACT

Item Response Theory (IRT) helps discriminate authentic psychometrics of the scale. This articulated analysis provides reviews of recent years supporting the approach of Item Response Theory and its uses in the succession of multiple psychological assessment tools. Screening tools have been categorized into four dimensions, i.e., Cognitive, Psychosocial, Psychophysiological, and Language Assessments. The literature review method was used in the mapping and evaluation of the result. Research pieces of evidence suggested Item Response Theory has proven helpful in many subjects and assessments techniques. IRT approach on pictorial assessments has also been notified effective in the authentic formulation and calculation just as questionnaire and other assessment apparatuses. IRT approach has proven effective in constructing and paraphrasing an assessment if discovered in another language. This brief review concluded with the result of positive response and IRT to be a practical approach in the psychometric calculation of psychological assessments.

INTRODUCTION

Item response theory (IRT) has modernized the testing psychometrics in scales' validation and development. Screening apparatuses validating appropriate responses need to be calculated psychometrically. Multiple methods have been studied to administer in history. This study provides evidence to successful psychometric with the application of Item Response Theory. Many administrative tools these days are developed using IRT. Armed Forces, aviation, and academic tests SAT, GRE, & ASVAB are developed and validated by using IRT. Because the reliability and accuracy can be administered of every item better in a tool by using IRT, the significant reduction of poor validated things can help minimalize the inaccuracy of a tool.

IRT is a widely used method in the operational field for generating and certifying educational examinations (Meijer & Tendeiro, 2018). In language assessment, the IRT approach is being to investigate a variety of subjects, counting (i) dimensionality and local independence of linguistic assessment (Eckes, 2014; Lee, 2004); (ii) differential item functioning (Geranpayeh & Kunnan, 2007; Pae, 2012) (iii) computerized adaptive testing, equating, and scoring (Chan et al.,2019; Kim et al.,2020; Mizumoto et al., 2019).

Kazemi&Kajonius (2021) indicated that Item response theory (IRT) looks at how effectively diverse items provide well-distributed measurement accuracy at diverse stages of the latent feature they're

testing. (e.g. De Ayala, 2009). IRT is concerned with item performance. As a result, unlike traditional test theory, this is grounded on the hypothesis that all test items are, IRT usually does not suggest those test items are similarly problematic. furthermore, In IRT, the probability that a person with a definite level of a particular attribute will have a practical level of that attribute is determined by the person's true trait extent and item factors can be measured in two ways with respect to discrimination, and difficulty, as opposed to classical test theory, which accepts that a person's achieved score on an assessment is the sum of a true score and an error score. At trait ranges that match the difficulty parameter of the item, an item's aptitude to differentiate among responders is most specific. The reciprocal of this variance is the extent of data at a given skill level, hence IRT is better understood as the reciprocal of the accuracy with which an assured medium of aptitude or approach can be assessed. It is based on likelihood curvatures, which show the probability of picking any of the item choices for a certain item for reference on a Likert scale, given a definite level of the primary trait/attitude and the item difficulty for a specific response category.

The majority of computations in overall mental wellbeing assessments are constructed on pre-existing standard measurements established in more resource and cultural-rich environments, with restricted applicability to middle and low-income realms (Betancourt et al.2011a; Betancourt et al.2011b; Yang et al.2021; Patel et al., 2007a; Patel et al., 2007b; Van Ommeren, 2003; Yang & Kao, 2014).

Cognitive functioning tests are the gold standard for diagnosing dementia, as well as determining functional decline and tracking and measuring disease progression. Measures of cognition in dementia should be able to consistently diagnose and assess the severity of the disease in its early stages (McGrory et al. 2014). IRT is predicated on the likelihood of a person attaining a specific test score as a result of their aptitude on the latent construct. Difficulty and discrimination are two valuable measures that IRT can provide. The sensitivity of the items within a measure can also be examined using IRT.

For testing the legitimacy of assessment tools in nursing, the item response theory (IRT) approach is becoming progressively popular (Tavakol et al., 2014; Yang & Kao, 2014). IRT is based on mathematical equivalences that use a nonlinear monotonic function to explain the link amid test-taker ability and the probability of a correct/incorrect item response (Hays et al., 2000). IRT has the advantage of being able to disentangle the relationship among item difficulty and participant aptitude (Hambleton, Swaminathan, & Rogers, 1991). This attribute allows for the maintenance of test score comparability among test forms, the development of computer-adaptive testing methods, and the detection of test security failures (van der Linden, 2016).

As in classic psychometric approaches, these arrangements for diagnosing mental diseases characterize symptoms as causally originating from a basic illness (Schlechter et al,2021). This suggests that the symptoms of prevalent mental diseases are the same across nations and communities. Symptom presentations and manifestations, on the other hand, are heavily impacted by sociocultural factors (Lewis-Fernández & Kirmayer, 2019). It might be challenging to translate copies of original surveys when items are semantically sensitive or not indicative of the construct in a foreign culture. For example, depending on the translation, Arabic terms used to indicate a loss of interest, a basic symptom of depression, have distinct nuances, which might lead to varied interpretations of this symptom (Sulaiman, Bhugra, & De Silva, 2001).

To get a better understanding of how displaced people experience symptoms and to enhance the cross-cultural validity of assessment techniques, additional psychometric methodologies must be used in addition to CTT analysis. In (IRT) models, test information curves are useful for determining where the scales offer the most detail over the latent spectrum. These information curves identify the most explanatory range of a questionnaire's spectrum on the latent dimension; or the region whereas the questionnaire best discriminates between individuals. At the high point of the core disease, measures may be more discriminative. Much of the published work claiming to show the superiority of IRT over related methods from CTT focuses on technical issues that primarily capture the attention of statisticians and psychometricians. Given that CTT is more familiar, and that test

administration and scoring based on simple paper-and-pencil forms and sums of item scores is decidedly less complex, the onus is on proponents of IRT to demonstrate that the methodology directly impacts issues that are important to clinicians such as diagnostic accuracy, and power and effect size in clinical studies(Thomas,2019)

The benefits of early identification and treatment of dementia have been demonstrated in previous research. It is critical that professionals, including experts and non-specialists, have access to accurate, time-saving, and simple-to-use screening instruments that enable for early detection of dementia, The Clinical Dementia Rating (CDR) is widely used in Alzheimer's disease research studies, Item response theory (IRT) was used for item evaluation and model development, most CDR items discriminate well at mild and very mild levels of cognitive impairment(Yang et al.,2021). The shortened version of the CDR and the automatic scoring algorithm has established a good foundation for developing an eCDR and will ultimately improve the efficiency of cognitive assessment.

Choosing photos using subjective methods might be difficult for psych physiologists because one's subjective/self-reported reaction to a picture is not always indicative of the psychophysiological response that picture would evoke (e.g., Weinberg and Hajcak, 2010; Wilson et al.,2021). Item response theory (IRT) is a substitute for using normed ratings or expert consensus to choose photographs. IRT was originally designed for aptitude evaluation (Lord et al., 1968), however, because it is well-suited to a variety of areas within psychological research, interest in it has lately grown (Balsis et al., 2017; Egberink, Meijer, 2011).

The likelihood of a given reaction to each item is described in IRT models as a function of the item's total score on the latent characteristic of interest, known as theta (θ), and a set of item parameters, e.g., discrimination and location parameters (Kamata and Bauer, 2008). In contrast to traditional psychometric approaches. IRT psychometric assessments are sample-invariant, meaning that stimuli were chosen for their high content as well as discriminatory scores should execute consistently across populations. IRT offers a number of advantages that make it potentially well-suited to choosing inducements for use with psych physiological recordings in a rigorous, objective manner.

METHODOLOGY:

Psychometric, validity, Item response theory, and behavioral assessments were used as keywords to search suitable and relatable reviews on Internet Search Engines. Most of the data were extracted from Wiley, Elsevier BMC, etc. Studies from multiple years were studied and involved for review calculating psychometrics of several behavioral assessing tools. A fair quantity of researches from all over the globe was reviewed in the course of this study. In this contextual setting, a literature review method was adopted. A few behavioral assessments have been named and discussed in the literature. This paper provides psychometric responses of IRT on different behavioral assessments. Results are explained in four subheadings as seen in figure 1.

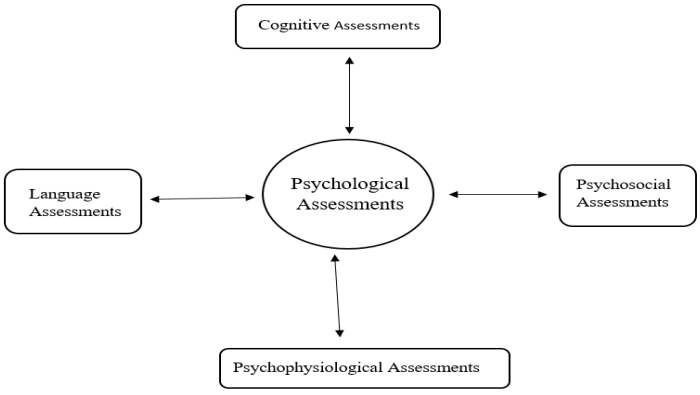

Fig. 1 ***Result Design of Psychological Assessments***

RESULTS & DISCUSSION:

Cognitive Assessments:

In the selection process, an electronical test was created to assess thinking as a generic cognitive capacity that reflects the necessities of a prospective job. To our understanding, no previous admission test based on a reasoning process (in which the test-taker gathers and processes information before coming to a conclusion) has been developed. In order to provide an in-depth research of the item characteristics for increasing the quality and equality of assessment, the psychometrical testing was based on IRT modeling rather than the CTT technique. (Tavakol et al., 2014).

Wiley published a paper in 2020 on the IRT approach on reasoning skills test for nursing students. The ReSki scale was developed electronically in three sections and to assess the parameters to its validity and development. A sample of more than a thousand participants was examined. The results determined ReSki testing to be a successful and impactful assessment tool. (Vierula et al., 2021).

BioMed Central contributed to a study in cognitive assessment on people with dementia. The IRT technique improved the cognitive scales' exegetical power and might be used to give clinicians essential items from a broader test battery with high predictive value. More research using IRT in a broader variety of studies with patients with dementia of various etiologies and intensity is needed. Four relevant published studies found IRT useful in the cognitive assessments of dementia. (McGrory et al., 2014)

In the Department of Psychology, Germans scrutinized a western developed screening tool of anxiety and depressive symptoms using IRT methodology. Later published in Wiley, the study researchers suggested that research into the psychometric features of tools for measuring mental strength concerns among refugees with an account of mass barbarity must go beyond traditional entirety score formulations of mental diseases. Either way, data did not deviate from its purpose. (Schlechter et al., 2021; Khairuldin et al., 2024).

A commonly used cognitive assessment, The Addenbrooke's Cognitive Examination III widely used to measure Chronic Cognitive Impairment was evaluated in a prior sample of diagnosed Alzheimer patients. The cognitive processing test reveals that the scale items are sensitive to various levels of impairment in diverse ways. These findings may aid in the identification of patterns of cognitive impairment and the accurate detection of various stages of dementia. (Calderón et al., 2021; Jam et al., 2019).

Psychosocial Assessments:

Item response theory (IRT) is a modern psychometric methodology that has changed the rules for developing, evaluating, and scoring psychological tests (Embretson, 1996;Jam et al., 2018). The methodology has broad applications in clinical assessment. In clinical assessment, item discrimination parameters (slopes of IRCs) are often surprisingly high (i.e., greater than 2.5; logistic metric), suggesting that the measured construct is conceptually narrow(Reise& Waller,2009), As an example, in a recent evaluation of the criteria for major depression in the Diagnostic and Statistical Manual of Mental Disorders, Third Edition, Revised (DSM-III-R), Aggen et al. (2005) report item discrimination estimates of 3.21 and 2.59 (logistic metric) for the items "depressed" and "disinterest," respectively. Hays et al. (2007) report item discriminations ranging from 1.88 to 4.24 on a physical functioning measure, and Chan et al. (2004) report item discriminations of 4.43 ("could not shake the blues") and 4.14 ("felt sad") on a popular depression measure.

In 2014 Wiley articulated a study on Psychosocial Assessment in the development of African youth under the IRT approach. Total item of 60 dimensionally categorized and shortlisted in the precision of 41 valid Items. The research showed how IRT analyzes may also be used to improve the psychometric performance of a qualitatively generated psychological assessment. Internalizing and externalizing emotional and behavioral disorders, physical complaints without a medical reason, and

prosocial attitudes and actions among youth can all be assessed using the African Youth Psychosocial Assessment (AYPA) test. (Betancourt et al.,2014).

Psychophysiological Assessments:

Data is not only collected in black and white. Pictorial methods are used in multiple psychological assessments to collect data. Elsevier publications provided the IRT approach in the collection of data through Emotional pictures. IRT is being used to help choose photos based on their accuracy in quantifying latent emotional responsiveness and their capacity to distinguish between distinctive sorts of emotional attention. (Wilson et al., 2021)

A Swedish sample of the nursing staff was examined to assess in response to person-centered care. The items have varying degrees of information. Nonetheless, for those items that did display any data, the distribution of evidence across the characteristic range was comparable for the majority of them, indicating that the items categorized well at reduced ranks of person-centeredness. (Kazemi& Kajonius, 2021).

Language Assessment

Elsevier in current years reported a review of multiple dimensions on language assessments. IRT formulated analysis was comprehended in the researches collected from two decades. Within the two streams of study, there was convergent evidence supporting the function of tests, linguistic skills, subskills, and verbal aspects as sources of multidimensionality, but miscellaneous verdicts were testified for the role of item layouts across research torrents. Both a unidimensional and multidimensional framework was used to measure listening, interpretation, communication, and writing skills. (Min& Aryadoust,2021)

CONCLUSION:

Item Response Theory resulted to be an effective technique in the evaluation and construction of screening tools. Multiple types of research from the current decade were reviewed in the study to determine the authentication of the IRT approach in discriminating and effectiveness of every individual item of a scale. Multidimensional factors and assessments were studied in the past and published in mainstream journals.

REFERENCES:

Aggen, S., Neale ,M.& Kendler K. (2005). DSM criteria for major depression: evaluating symptom patterns using latent-trait item response models. *Psychol. Med. 35*:475–87

Balsis, S., Ruchensky, J.R., Busch, A.J. (2017). Item response theory applications in personality disorder research. *Personality Disorders, 8* (4), 298–308.

Betancourt T.S., Meyers-Ohki S., Stevenson A., Ingabire C., Kanyanganzi F., Munyana M.C., Mushashi C., Teta S., Fayida I., Cyamatare F.R., Stulac S.N., Beardslee W. (2011a) Using mixed-methods research to adapt and evaluate a family strengthening intervention in Rwanda. *African Journal of Traumatic Stress, 2*(1), 32–45.

Betancourt T.S., Rubin-Smith J., Beardslee W.R., Stulac S.N., Fayida I., Safren S.A. (2011b) Understanding locally, culturally, and contextually relevant mental health problems among Rwandan children and adolescents affected by HIV/AIDS. AIDS Care, *iFirst*, 1–12. DOI: 10.1080/09540121.2010.516333

Betancourt,T., Yang, F., Bolton, P., & Normand, S.(2014). Developing an African youth psychosocial assessment: An application of item response theory. *International Journal of Methods in Psychiatric Research, 23*(2), 142–160. https://doi.org/10.1002/mpr.1420

Calderón, C., Beyle, C., Véliz-García, O., & Bekios-Calfa, J. (2021). Psychometric properties of Addenbrooke's Cognitive Examination III (ACE-III): An item response theory approach. *PloS one, 16*(5), e0251137.

Chan K., Orlando M., Ghosh-Dastidar B., Duan N.& Sherbourne C. (2004). The interview mode effect on the center for epidemiological studies depression (CES-D) scale. *Med. Care 42*:281–89

Chan, S. W. Y., Cheung, W. M., Huang, Y., Lam, W., & Lin, C. (2019). Development and validation of a Chinese character acquisition assessment for second-language kindergarteners. *Language Testing, 37*(2), 1–20.

De Ayala, R. J. (2009). The theory and practice of item response theory. Guilford Press.

Eckes, T. (2014). Examining testlet effects in the TestDaF listening section: A testlet response theory modeling approach. *Language Testing, 31*(1), 39–61.

Egberink, I.J.L., Meijer, R.R. (2011). An item response theory analysis of Harter's Self-perception Profile for children or why strong clinical scales should be distrusted. *Assessment 18* (2), 201–212.

Embretson SE (1996). The new rules of measurement. *Psychological Assessment, 8*(4), 341–349.

Geranpayeh, A., & Kunnan, A. J. (2007). Differential item functioning in terms of age in the certificate in advanced English examination. *Language Assessment Quarterly, 4*(2), 190–222.

Hambleton, R. K., Swaminathan, H., & Rogers, H. (1991). *Fundamentals of item response theory*. Newbury Park: Sage Publications, Inc.

Hays RD, Liu H, Spritzer K, Cella D. 2007. Item response theory analyses of physical functioning items in the medical outcomes study. *Med. Care 45*(5 Suppl. 1): 32–38.

Hays, R. D., Morales, L. S., & Reise, S. P. (2000). Item response theory and health outcomes measurement in the 21st century. *Medical Care, 38*(Suppl9), II28–II42.

Jam, F. A. (2019). CRYPTO CURRENCY–A NEW PHENOMENON IN MONETARY CIRCULATION. *Central Asian Journal of Social Sciences and Humanities*, *4*(1), 39-46.

Jam, F. A., Singh, S. K. G., Ng, B., & Aziz, N. (2018). The interactive effect of uncertainty avoidance cultural values and leadership styles on open service innovation: A look at malaysian healthcare sector. *International Journal of Business and Administrative Studies*, *4*(5), 208-223

Kamata, A., Bauer, D.J. (2008). A note on the relation between factor analytic and item response theory models. *Struct. Equ. Model. 15* (1), 136–153

Kazemi, A., & Kajonius, P. (2021). Assessing person-centered care: An item response theory approach. *International Journal of Older People Nursing, 16*(1), e12352.

Khairuldin, W. M. K. F. W., Anas, W. N. I. W. N., Suhaimi, M. A., Idham, M. A., Embong, A. H., & Hassan, S. A. (2024). Developing a holistic Islam-based academic integrity model for Malaysian higher education institutions. *Pakistan Journal of Life and Social Sciences*, *22*(1), 677-686.

Kim, S. Y., Lee, W.-C., & Kolen, M. J. (2020). Simple-Structure Multidimensional Item Response Theory Equating for Multidimensional Tests. *Educational and Psychological Measurement, 80(*1), 91-125. https://doi.org/10.1177/0013164419854208

Lee, Y. (2004). Examining passage-related local item dependence (LID) and measurement construct using Q3 statistics in an EFL reading comprehension test. *Language Testing, 21*(1), 74–100.

Lewis-Fernández, R., & Kirmayer, L. J. (2019). Cultural concepts of distress and psychiatric disorders: Understanding symptom experience and expression in context. Transcultural *Psychiatry, 56*(4), 786–803.

Lord, F.M., Novick, M.R., Birnbaum, A. (1968). *Statistical Theories of Mental Test Scores*. Addison-Wesley.

McGrory, S., Doherty, J. M., Austin, E. J., Starr, J. M., & Shenkin, S. D. (2014). Item response theory analysis of cognitive tests in people with dementia: a systematic review. BMC *psychiatry, 14*(1), 1-15.

Meijer, R. R., & Tendeiro, J. N. (2018). *Unidimensional item response theory*. In P. Irwing, T. Booth, & D. J. Hughes (Eds.), The Wiley handbook of psychometric testing: A multidisciplinary reference on survey, scale and test development (vol. 1, pp. 413–443). NJ, USA: Wiley Blackwell.

Min, S.&Aryadoust, V. (2021). A systematic review of item response theory in language assessment: Implications for the dimensionality of language ability. *Studies in Educational Evaluation, 68*, 100963, https://doi.org/10.1016/j.stueduc.2020.100963

Mizumoto, A., Sasao, Y., & Webb, S. A. (2019). Developing and evaluating a computerized adaptive testing version of the word part levels test. *Language Testing, 36*(1), 101–123.

Pae, T. (2012). Causes of gender DIF on an EFL language test: A multiple-data analysis over nine years. *Language Testing, 29*(4), 533–554

Patel V., Araya R., Chatterjee S., Chisholm D., Cohen A., De Silva M., Hosman C., McGuire H., Rojas G., van Ommeren M. (2007a). Treatment and prevention of mental disorders in low-income and middle-income countries. *The Lancet, 370*(9591), 991–1005. DOI: 10.1016/S0140-6736(07)61240-9

Patel V., Flisher A.J., Hetrick S., McGorry P. (2007b) Mental health of young people: a global publichealth challenge. *Lancet, 369*, 1302–1313. DOI: 10.1016/S0140-6736(07)60368-7

Reise SP, & Waller NG (2009). Item response theory and clinical measurement. *Annual Review of Clinical Psychology, 5,* 27–48.

Schlechter, P., Wilkinson, P. O., Knausenberger, J., Wanninger, K., Kamp, S., Morina, N., & Hellmann, J. H. (2021). Depressive and anxiety symptoms in refugees: Insights from classical test theory, item response theory and network analysis. *Clinical Psychology & Psychotherapy, 28*(1), 169-181.

Sulaiman, S. O., Bhugra, D., & De Silva, P. (2001). The development of a culturally sensitive symptom checklist for depression in Dubai. *Transcultural Psychiatry, 38*(2), 219–229. https://doi.org/10.1177/ 136346150103800205

Tavakol, M., Rahimi-Madiseh, M., & Dennick, R. (2014). Postexamination analysis of objective tests using the three-parameter item response theory. *Journal of Nursing Measurement, 22*(1), 94–105

Van der Linden, W. J. (2016). Introduction. In W. J. van der Linden (Ed.), *Handbook of item response theory, volume one*: Models (pp. 1–10). London: Taylor & Francis Group.

Van Ommeren, M. (2003) Validity issues in transcultural epidemiology. *The British Journal of Psychiatry, 182*(5), 376–378. DOI: 10.1192

Vierula, J., Talman, K., Hupli, M., Laakkonen, E., Engblom, J., & Haavisto, E. (2021). Development and psychometric testing of Reasoning Skills test for nursing student selection: An item response theory approach. *Journal of Advanced Nursing, 77*(5), 2549-2560.

Weinberg, A., Hajcak, G.(2010). Beyond good and evil: the time-course of neural activity elicited by specific picture content. *Emotion 10* (6), 767–782.

Wilson, K. A., Clark, D. A., & MacNamara, A. (2021). Using item response theory to select emotional pictures for psychophysiological experiments. *International Journal of Psychophysiology, 162,* 166-179.

Yang, F. M., & Kao, S. T. (2014). Item response theory for measurement validity. *Shanghai Archives of Psychiatry, 26*(3), 171–177. https://doi. org/10.3969/j.issn.1002-0829.2014.03.010

Yang,L et al.(2021). Item response theory analysis of the Clinical Dementia Rating. *Alzheimer's & Dementia, 17*(3), 534–542. DOI: 10.1002/alz.12210