

**THEORETICAL AND IMPLICATION ASPECTS OF ITEM RESPONSE THEORY IN
PSYCHOLOGICAL AND EDUCATIONAL PHENOMENA: REVIEW ARTICLE**

Izzeldeen Abdullah Awwad Alnaimi

Educational Measurement and Evaluation,

e-mail: ealnuaimi@imamu.edu.sa

Mobile: 00966507015462

Abstract:

Introduction: IRT is also known as latent trait theory, strong true score theory, or modern mental test theory, is considered as a pattern for various aspects related to design, analysis, and scoring of tests and questionnaires to measure attitudes and abilities.

Objectives: The main objective of this study is to review the literature about theoretical and implications of item response theory (IRT) in better understanding the phenomena of psychology and education.

Methods: We reviewed the literature to find out appropriate citing literature. We searched the most important sites including Science direct, sage journals, Scopus, and other citing sites accessible by Google. The articles were read, classified and summarized and organized in the current article.

Results: We gave a description of IRT and its constructs. We put focus on its unidimensional and multidimensional aspects. We also talked about some models used in IRT. IRT has the ability to give estimations as related to domains and items within the scales of instrument irrespective to respondents, an issue that gives its popularity. IRT has been reported as of having large implications in studies about education and psychology to reach patterns of responses of participants about test items. The importance of IRT was overemphasized.

Conclusion: IRT can be used to participate in better understanding of clinical phenomena. It has wide scopes of dimensionality either unidimensional or multidimensional. IRT has the ability to give estimations as related to domains and items within the scales of instrument irrespective to respondents, an issue that gives its popularity.

Keywords: item response theory, psychology, education, unidimensional, multidimensional, items, scales.

Corresponding author: Izzeldeen Abdullah Awwad Alnaimi

e-mail: eaalnuaimi@imamu.edu.sa

Mobile: 00966507015462

Introduction

An overview of IRT

IRT is also known as latent trait theory, strong true score theory, or modern mental test theory, is considered as a pattern for various aspects related to design, analysis, and scoring of tests and questionnaires to measure attitudes and abilities (Zanon et al., 2016).

It is a theory of testing based on the relationship between individuals' performances on a test item and the test takers' levels of performance on an overall measure of the ability that item was designed to measure. Several different statistical models are used to represent both item and test taker characteristics (Hambleton, 2003).

IRT is distinguished from others such as Likert scaling by giving each item its characteristic or weight of difficulty. As an example, in Likert scaling, all items are thought to have parallel

instruments (Zanon et al., 2016). In IRT, the difficulty of each item (the item characteristic curves, or ICCs) as information is treated to be incorporated in scaling items.

From another point of view, IRT is concerned with mathematical models for testing of data. Due to considerations that IRT is considered as superior to classical test theory (Watson et al., 198), it is the most preferred method for establishing scales in the United States. IRT has taken its name since it puts the focus on item compared with classical test theory that puts the focus on test-level. According to this context, IRT establishes the response for each examinee related to certain ability of each item in the test. One of the main features of IRT is that it depends on the principle of probability of a correct response regarding an item as a function of mathematics of an individual and variables of an item. The variable of an individual is considered as a single dimension. Illustrating examples include intelligence in general or the extent of attitude. On the other hand, the items characterizing parameters involve the associated difficulty, as location within the range of difficulty; the slope of correlation indicates the success rate of individuals to vary with their ability; and a pseudo-guessing variable is determining the lower rate in which the least able persons are likely to score as a result of guessing 25% by chance on a multiple choice item in which there are four possible responses. In a similar way, IRT measures human behaviors in social networks. The various views of participants can be collected and studied using IRT to categorize information as true or misinformation (Watson, 2005).

Study objectives:

The main objective of this study is to review the literature about theoretical and implications of item response theory (IRT) in better understanding the phenomena of psychology and education.

The implications of IRT in education and psychology

IRT has been reported as of having large implications in studies about education and psychology to reach patterns of responses of participants about test items and to highlight how participants respond to test items either participants were singles or in groups (Thissen and Wainer 2001). The framework of IRT postulates the existence of a functional relationship between a latent level of respondent such as the ability in educational testing, or what is called the factor score in the factor analysis tradition and the stimulant of an item level.

One of the major goals of IRT implies the isolation of the item variables and characteristics of sampled population from manifest data. This process of isolation requires complicated statistical analysis understanding.

IRT has two main models unidimensional and multidimensional parameter estimations. Unidimensional latent trait is the most popular model in social sciences and education studies due to historical considerations from one side, and from the other side, multidimensional parameter estimation approaches were not fully understood or developed (Baker and Kim 2004; Reckase 2009).

It has been shown that unidimensional models are characterized by their simplicity and have different interesting and important measurement properties (e.g., Rasch models), other psychological constructs are crucially multidimensional by their nature such as understanding of hidden constructs as a group of sub-scale parts within or along-side a more general construct, or it can be understood as factors that compensate and influence the probabilities of item response. The decision of using such models of IRT was challenged by the process of evaluation of parameters of the item within the space that is characterized by high dimensions because of

having increased number of factors, which makes difficulty in computing based on numerical techniques. The recent advances in theory of estimation and improved abilities of personal computers, multidimensional models of TRT have started to become as a feasible statistical approach (Wirth and Edwards, 2007, Reckase, 2009; Edwards, 2010).

Bi-factor IRT model

Toland et al (2017) conducted a study about the use of bi-factor IRT model in the explanation of the dimensions of a multi-layered questionnaire in which it is proposed that there is a tendency of continuous latent variables to respond to items. This model is used to give an idea about the sites of people involved in a general continuous latent variable and in continuous specific traits which specify responses to groups of items. According to Campell (2008), the bifactor graded response (bifac-GR) model has aspects such as specification, assumptions, estimation, and interpretation that can be illustrated through the reanalysis of data on the Shared Activities Questionnaire.

A variety of psychometric approaches exist to help in taking appropriate decision to decide if a set of categorical items is a unidimensional or multidimensional latent construct (Toland et al., 2017). If the construction is proposed to be unidimensional (e.g., Autism Spectrum Disorder, social skills, self-esteem, depression, school belongingness), it is expected that all items may show a single continuous latent variable. On the other hand, a multidimensional construct shows various latent variables that clarify relationships between the items and specified variables including engagement of students such as academic, behavior, cognitive, psychological aspects (Appleton et al., 2006; Beran et al., 2012).

IRT is considered a well-known psychometric method included for modeling unidimensional and multidimensional concepts (Lord, 1959; Lord and Novick, 1968; Embretson and Reise, 2000; De Ayala, 2009).

According to IRT, there is an assumption in which IRT models have items in scales and each of these items is under the effect of a single unidimensional factor. In psychology, scales have multidimensional nature that makes them to be rarely unidimensional (Reise et al., 2007).

Some scales based on IRT

As previously mentioned, IRT gives ideas regarding each item within an instrument as it represents a latent proficiency (Jessen et al., 2018). IRT has become popular through the works of Lord (1980), and it has been used further to develop various known educational tests including the SAT and National Assessment of Educational Progress (Lord, 1967). In these tests, IRT gives characteristics of a single dimension, including math or reading ability (Lord, 1967).

IRT enhances ‘classical test theory,’ in which its item statistics depend on the tested population (Lord and Novick, 1968). IRT works to give a description of question characteristics that are likely to be constant irrespective to the population (Jessen et al., 2018). It is worth mentioning that such item variables involve both information and location. Information defines the extent in which an item is likely to be discriminated among patients, whereas location determines where the scale including the proficiency an item provides this information. It has been shown that such location and information parameters are likely to be stable as far as the models proposed in IRT fit the data (Yen and Fitzpatrick, 2006).

IRT is used to develop ‘‘item maps,’’ which represent responses of an item for overall scores. Item maps help clinicians who experience difficulties in overall scores; these maps show specific

questions/ answers that are associated with latent ability levels. The resulting visions are likely to be used in individualized clinical care. As an example, hearing loss is a good condition that impacts 19% of adolescents and 60% of septuagenarians (Shargorodsky et al., 2010; Walling and Dickson, 2012). Its influences gave it a central role in Healthy People 2020 as well as in multiple professional organizations (Stachler et al., 2012; Alford et al., 2014).

One of the known scales is the Inventory of Interpersonal Problems 64 (IIP-64) (Alden, Wiggins, and Pincus, 1990; Horowitz et al., 2000). It is usually used as a psychological inventory to evaluate interpersonal problems either in centers of clinical treatment and/or in research applications. There are eight subscales in the IIP-64 that represent eight domains of interpersonal behavior, and each has eight items. Several researchers have investigated the psychometric properties of the IIP-64. The factor structure has been investigated through various studies including Acton and Revelle (2002), Grosse-Holtforth, Lutz, and Grawe (2006), and other studies targeted the sensitivity of this inventory to change detection (Huber, Henrich, and Klug (2007), and the extent to which its importance can be compared with other measures focusing on interpersonal behavior (Alden et al., 1990; Horowitz et al., 2000; Leising, Rehbein, and Sporberg (2007), and Vittengl, Clark, and Jarrett (2003). Other researchers investigated its relation to psychotherapeutic impacts (Horowitz, Rosenberg, and Bartholomew, 1993; Puschner, Kraft, and Bauer, 2004; Ruiz et al., 2004).

The IIP-64 scale has been used by many researchers to examine interpersonal distress, or its subscales have been used to distinguish the specific domains of interpersonal functioning. In clinical settings, IIP-64 subscale scores can be used for comparing purposes of either individual or a group of individuals with normal sample, or sometimes to compare person's distress within an interpersonal domain compared with the person's overall level of interpersonal distress,

which, in turn, permits the researcher to determine the domains that are categorized as problematic (Tracey et al., 1996; Pincus et al., 1998; Horowitz et al., 2000; Acton and Revelle, 2002; Vanheule et al., 2006).

Conclusions

IRT can be used to participate in better understanding of clinical phenomena. It has wide scopes of dimensionality either unidimensional or multidimensional. IRT has the ability to give estimations as related to domains and items within the scales of instrument irrespective to respondents, an issue that gives its popularity.

References:

- Acton, G.S., Revelle, W. (2002). Interpersonal personality measures show circumplex structure based on new psychometric criteria. *Journal of Personality Assessment*, 79, 446-471.
- Alden, L.E., Wiggins, J.S., Pincus, A.L. (1990). Construction of circumplex scales for the assessment of Interpersonal Problems. *Journal of Personality Assessment*, 55, 521–536.
- Alford RL, Arnos KS, Fox M, et al (2014). American College of Medical Genetics and Genomics guideline for the clinical evaluation and etiologic diagnosis of hearing loss. *Genet Med.*, 16:347-355.
- Annika Jessen, Andrew D. Ho, C. Eduardo Corrales, Bevan Yueh, Jennifer J. Shin (2018). Improving Measurement Efficiency of the Inner EAR Scale with Item Response Theory. *Otolaryngology–Head and Neck Surgery* 158(6): 1093-1100.
- Appleton, J. J., Christenson, S. L., Kim, D., Reschly, A. L. (2006). Measuring cognitive and psychological engagement: Validation of the Student Engagement Instrument. *Journal of School Psychology*, 44, 427–445. <http://dx.doi.org/10.1016/j.jsp.2006.04.002>.
- Baker FB, Kim SH (2004). *Item Response Theory: Parameter Estimation Techniques*. 2nd edition. Dekker, New York.

- Beran, T., Stanton, L., Hetherington, R., Mishna, F., Shariff, S. (2012). Development of the Bullying and Health Experiences Scale. *Interactive Journal of Medical Research*, 1 (2), e13. <http://dx.doi.org/10.2196/ijmr.1835>.
- Campbell, J. M. (2008). Brief report: Reliability and validity of the Shared Activities Questionnaire as a measure of middle school students' attitudes toward autism. *Journal of Autism and Developmental Disorders*, 38, 1598–1604. <http://dx.doi.org/10.1007/s10803-007-0534-6>.
- Cristian Zanon, Claudio S. Hutz, Hanwook (Henry) Yoo, Ronald K. Hambleton (2016). An application of item response theory to psychological test development. Zanon et al. *Psicologia: Reflexão e Crítica*, 29:18 DOI 10.1186/s41155-016-0040-x.
- De Ayala, R. J. (2009). *The theory and practice of item response theory*. New York, NY: Guilford Press.
- Edwards MC (2010). A Markov Chain Monte Carlo Approach to Confirmatory Item Factor Analysis. *Psychometrika*, 75(3), 474-497.
- Embretson, S. E., Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Grosse-Holtforth, M., Lutz, W., Grawe, K. (2006). Structure and change of the IIP-D preand postpsychotherapy. *European Journal of Psychological Assessment*, 22, 98–103.
- Hambleton RK (2003), Jodoin M. Item response theory: models and features. In: Fernández-Ballesteros R, editor. *Encyclopedia of Psychological Assessment*. London: Sage. p. 509–14.
- Horowitz, L. M., Alden, L. E., Wiggins, J. S., Pincus, A. L. (2000). *Inventory of Interpersonal Problems: Manual*. New York: The Psychological Corporation Harcourt.
- Horowitz, L. M., Rosenberg, S. R. Bartholomew, K. (1993). Interpersonal problems, attachment styles, and outcome in brief dynamic psychotherapy. *Journal of Consulting and Clinical Psychology*, 61, 549-560.
- Huber, D., Henrich, G., Klug, G. (2007). The Inventory of Interpersonal Problems (IIP): Sensitivity to change. *Psychotherapy Research*, 17, 474-481.
- Leising, D., Rehbein, D., Sporberg, D. (2007). Validity of the Inventory of Interpersonal Problems (IIP-64) for Predicting Assertiveness in Role-Play Situations. *Journal of Personality Assessment*, 89, 116-125.

- Lord FM (1967). An analysis of the verbal scholastic aptitude test using Birnbaum's three-parameter logistic model. *Educ Psychol Meas*, 2:989-1020.
- Lord FM (1980). *Applications of Item Response Theory to Practical Testing Problems*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Lord FM, Novick MR (1968). *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley Publishing.
- Lord, F. M. (1959). An approach to mental test theory. *Psychometrika*, 24, 293–302.
- Lord, F. M., Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Michael D. Toland, Isabella Sulis, Francesca Giambona, Mariano Porcu, Jonathan M. Campbell (2017). Introduction to bifactor polytomous item response theory analysis. *Journal of School Psychology*, 60, 41–63.
- Pincus, A.L., Gurtman, M. B., Ruiz, M. A. (1998). Structural Analysis of Social Behavior (SASB): Circumplex Analyses and Structural Relations With the Interpersonal Circle and the Five-Factor Model of Personality. *Journal of Personality and Social Psychology*, 74, 1629-1645.
- Puschner, B., Kraft, S., Bauer, S. (2004). Interpersonal Problems and Outcome in Outpatient Psychotherapy: Findings From a Long-Term Longitudinal Study in Germany. *Journal of Personality Assessment*, 83(3), 223-234.
- Reckase MD (2009). *Multidimensional Item Response Theory*. Springer-Verlag, New York.
- Reckase MD (2009). *Multidimensional Item Response Theory*. Springer-Verlag, New York.
- Reise, S. P., Morizot, J., Hays, R. D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcome measures. *Quality of Life Research*, 16, 19 –31.
- Ruiz, M. A., Pincus, A.L., Borkovec, T. D., Echemendia, R.J., Castonguay, L.G., Ragusea, S.A. (2004). Validity of the Inventory of Interpersonal Problems for Predicting Treatment Outcome: An Investigation With The Pennsylvania Research Network. *Journal of Personality Assessment*, 83, 213-222.
- Shargorodsky J, Curhan SG, Curhan GC, et al (2010). Change in prevalence of hearing loss in US adolescents. *JAMA.*, 304:772-778.

- Stachler RJ, Chandrasekhar SS, Archer SM, et al (2012). Clinical practice guideline: sudden hearing loss. *Otolaryngol Head Neck Surg*, 146(3):S1-S35.
- Thissen D, Wainer H (eds.) (2001). *Test Scoring*. Lawrence Erlbaum Associates, Mahwah.
- Tracey, T., Rounds, J., Gurtman, M. (1996). Examination of the General Factor with the Interpersonal Circumplex Structure: Application to the Inventory of Interpersonal Problems. *Multivariate Behavioral Research*, 31, 441.
- Vanheule, S., Desmet, M., Rosseel, Y. (2006). The factorial structure of the Dutch translation of the Inventory of Interpersonal Problems: A test of the long and short versions. *Psychological Assessment*, 18, 112-117.
- Vittengl, J. R., Clark, L. A., Jarrett, R. B. (2003). Interpersonal problems, personality pathology, and social adjustment after cognitive therapy for depression. *Psychological Assessment*, 15, 29-40.
- Walling AD, Dickson GM (2012). Hearing loss in older adults. *Am Fam Physician.*, 85:1150-1156.
- Watson D (2005). Positive affectivity: The disposition to experience pleasurable emotional states. In: Snyder CR, Lopez SJ, editors. *Handbook of Positive Psychology*. New York: Oxford University Press, 106–19.
- Watson D, Clark LA, Tellegen A (1988). Development and validation of brief measures of Positive and Negative Affect: The PANAS Scales. *J Pers Soc Psychol.*, 54(1):1063–70. doi:10.1037/0022-3514.54.6.1063.
- Wirth RJ, Edwards MC (2007). Item Factor Analysis: Current Approaches and Future Directions. *Psychological Methods*, 12(1), 58–79.
- Yen WM, Fitzpatrick AR (2006). Item response theory. In: Brennan R, ed. *Educational Measurement*. 4th ed. Westport, CT: American Council on Education/Praeger Publishers; 111-118.