Assessment of 2PL, 3PL and 4PL Model-Fit on the Wechsler Preschool and Primary Scale of Intelligence (WPPSI-IV) Using Item Response Theory

Adepoju, Olufemi Adetunji., Dauda, Isiaka Kayode., Lawal, Abduljaleel Kola

Department of Educational Foundations and Curriculum Studies, Faculty of Specialised and Professional Education, Emmanuel Alayande University of Education, Oyo

DOI: https://dx.doi.org/10.47772/IJRISS.2025.907000337

Received: 12 June 2025; Accepted: 19 June 2025; Published: 16 August 2025

ABSTRACT

The purpose of this paper is to investigate into which of the models of item Response Theory that fits into the Wechsler Preschool and Primary Scale of Intelligence (WPPSI-IV) with a view to adapting the WPPSI-IV to measure the intelligence of Nigerian children aged approximately 7. The design of the study was survey which allows for data to be collected through the administration of the WPPSI-IV to 387 pupils. Data collected was analysed with the use of factor analysis on 2PL, 3PL and 4PL models of item Response Theory. Findings from the study indicate that 2PL model was more fit for WPPSI-IV than 3PL and 4PL models. It was recommended that the WPPSI-IV should be adapted or adopted by schools to measure pupils intelligence approximately aged 7 because the instrument is moderately difficult and discriminated moderately among the testees and that schools should be well furnished with audio-visual materials to promote critical thinking of children or pupils, this will go a long way in making them to do well in intelligence tests.

Keywords: Item Response Theory, Intelligence, Tests, Model-fit, measurement

INTRODUCTION

Item Response Theory is a psychometric approach based on the assumption that a testee's response to a particular test item is a function of the qualities of the individual or is influenced by the testee's ability. This agrees with Camilli and Shepheard(1994) that IRT is a measurement approach that relates the probability of a particular response on an item to overall examinee's ability level. Findings have revealed that the IRT model that most fit the (WPPSI-IV) UK, is the 2PL model which was evidence in Zanon et al (2016) and Salomen et al (2022) in their separate studies. IRT has therefore become a vital framework to resolve a lot of measurement problems.

Psychologists have come up with several models of IRT; the 1PL, 2PL, 3PL, 4PL and mixture IRT model. Finding out the model fits of a data set involves assessing how well a statistical model fits the data, it therefore means that a model fits tests help determine if a statistical model accurately represents the relationships within the data comparing the observed data to the model's predictions and provide a measure of how well the model fits the data.

According to Maydeu-Olivares (2013), item response theory modeling involves fitting a latent variable model to discrete response obtained from questionnaire/test items intended to measure educational achievement, personally attitudes and so on. Similarly, Swamina Than et-al (2006) positioned that assessment of the fit of the model to data is multi-faceted and must be carried out at the test level as well as the item level. Maydeu Olivares (2015) further added that the Goodness of Fit (G O F) of a statistical model such as an IRT model defines how well the model matches a set of observations. Lu (2006), stated that "accessing fit in IRT models usually involves validating assumptions underlying the models and evaluating Goodness-of-Fit, which specifically refers to how effectively the model describes the outcome data." Pi Hambleton et-al (1991), therefore argued that violations of model data might lead to the rejections of null hypothesis in Goodness- of – Fit tests even though it is not in all cases.

Watkins et-al (2014), studied the Bi-factor structure of the WPPS I-IV, submitted that the WPPS I-IV fits a bi-factor model or 2PL model than the other models tested in the study. Delhamid et-al(2021) assessed the precognitive abilities using the WAIS-IV based on IRT approach submitted that the 2PL model provided a good fit for dichotomous subsets while the graded response model fitted the polynotomous.

Also, Sara- Ann (2015) while exploring the item difficulty and other psychometric properties of the core perceptual; verb; and working memory of subsets of the WAIS-IV using IRT approachs discovered that the Rasch modeling, a form of IRT is one parameter logistic model that is appropriate for items with only two response options.

Statement Problem

Assessing the IRT model fits for a data set obtained from the administration of an instrument is an essential area in the study of 1RT. Adapting or adopting an instrument requires that the fitness of the instrument is investigated. This study will therefore provide an avenue to select a suitable IRT model that will ensure that the WPPSI-IV is adapted or adopted for use to measure the intelligence of Nigerian Children approximately aged 7.

Questions

The following research questions were generated for the study:

1. What model fits the data on the WPPSI-IV plata between 2PL and 3PL?
2. What model fits the data in the WPPSI-IV between 2PL and 4PL?

METHODOLOGY

The study adopted a survey design, as data was collected in raw form through the administration of the WPPS I-IV without any modification. The population for the study comprises of pupils in public primary schools out of which 387 participants were sampled through a purposive sampling method. The pupils were approximately aged 7. The WPPSI-IV was administered on the selected pupils. This instrument has fifteen sub-scales comprises of 270 items. Reliability of the WPPSI-IV was estimated through internal consistency, test-retest stability, and inter-scorer agreement as reported by the Technical and Interpretive Manual (2012). Internal consistency co-efficient across nine age groups were 0.96 – 0.96 for the Full Scales Intelligence Quotient (FSIQ) and ranged from 0.85 to 0.96 for index scores and from 0.71 to 0.95 for the sub-test score (Wechsler 2012). Data collected were analysed with the use of factor analysis computed on 'R' supported IRT software at 0.05 level of significance.

RESULT AND DISCUSSION

Table 1: Comparison of 2 PL and 3 PL of WPPSI-IV

Model	AIC	AICc	SABIC	HQ	BIC	logLik	X2	Df	P
2	132112	110631	132684	133005	134392	-65518	NaN	NaN	NaN
3	131970	127564	132829	133311	135390	-65178	679.637	269	0

Given that the AIC estimate for 2 PL is lower than that of 3 PL then it can be said that 2 PL model is more fitting for the data than 3 PL

Table 2: Comparison of 2 PL and 4 PL of WPPSI-IV

Model	AIC	AICc	SABIC	HQ	BIC	logLik	X2	Df	P
2	132112	110631	132684	133005	134392	-65518	NaN	NaN	NaN
4	133525	129422	134670	135312	138085	-65686	-336.96	538	1

The 2 PL model is lower than the 4 PL in this comparison, hence, 2 PL can be said to be better than 4 PL.

DISCUSSION OF THE FINDINGS

From the study, it was clearly shown that AICE of 2PL < AICE of 3PL i.e 1.10631 < 1.27564 for table 1; while AICE of 2PL is also < AICE of 4PL i.e, 1.10631 < 1.29422. This is an indication that 2PL model is mom fit than 3PL and 4PL models, a further indication that the WPPSI-IV is moderately difficulty and moderately discriminating, making it very suitable to be adopted on adapted for use on Nigerian children approximately aged 7. This Zanon et al., (2016) and Saloman et al., (2022).

CONCLUSION

From this study, it was revealed that when 2PL model was tested against the 3PL model for the WPPSI –IV data, 2PL is more fitting than 4PL. It was therefore be concluded that 2PL model is more fit for WPPSI-IV instrument than other models indicating that the WPPSI-IV instrument possess the attribute of difficulty and discrimination which agreed with the finding of Watkins et-al (2014)

RECOMMENDATIONS

The following recommendations were made from the study:

1. WPPS I-IV(UK) is a reliable instrument to measure children's intelligence should therefore be adapted for use in primary schools in Nigeria
2. Since the findings is in favour of 2PL model of IRT, the WPPS I-IV is suitable for testing children's intelligence for their overall development
3. School managers should equip their schools with audio-visual aids to promote critical thinking of children or pupils, this will go a long way in making them to do well in intelligence tests.

REFERENCES

1. Abdelhamid GSM, Bassiouni MGA & Gomez- Benito J. (2021). Assessing cognitive abilities using the WAIS-IV. An Item Response Theory Approach. International Journey of Environmental Public Health. 18.13, 1-15
2. Hambleton R.K, Swammathan H & Roser H. (1991) Fundamentals of item response theor. Newbury Park, C A: SADE
3. LU. Y. (2006). Assessing fit of item response theory model. A dissertation presented to the graduate school of the University of Massachusetts, U.S. A
4. Maydeu Olivares A. (2013) Goodness-of-Fit Assessment of Item Response Theory Models
5. Maydeu Olivares A. (2015). Evaluating the fit of IRT model. Ins. P Deise & D.A Revicki(Eds) Handbook of item Response theory modeling: Applications to typical performance assessment, U.S.A, Reutledge/Taylor & Francis Group.
6. Salomen, J., Haavisto, A. & Rosenquist, J. (2022). Comparison of WPPSI-IV and WISC-IV cognitive profiles in children – findings from the finished study – *Child Neoropsychology*, *29*(5), 687-709
7. Sara Ann M.S. (2015). Exploring the Item Difficulty and other Psychomeric Properties of the core Perceptual, Verbal and working Memory Subtests of the WAIS –IV using Item Response Theory . A Ph.D dissertation in clinical psychology submitted to the Centre for Psychological Studies. NOVA Southeastern University UK.
8. Swammathan, H., Hambleton K. R. & Rogens J. H. (2006). Assessing the Fit of Item Response Theory Models. Handbook of statistics. 26, 683-718
9. Watkins W. M, Beaujean A. A. (2014). Bifactor structure of the Wechsler Preschool and Primary Scale of Intelligence- Fourth edition school Psychol. Q 29.1, 52-63
10. Zanon, C., Hutz, C. S., Yoo, H. & Hambleton, R. K. (2016). An application of item response theory psychological test development: Psychologia. Reflexao e-critica.29. https://doi.org./10.1186/541155-016-0042x