Feature Data Generation for Computer Adaptive Testing: A Novel method for Transdisciplinary Psychometrics Improvements using Post-hoc Simulation Approach

Jumoke I. Oladele

Department of Science and Technology Education, Faculty of Education, University of Johannesburg, South Africa, Email: jumokeo@uj.ac.za

Received 2 June, 2022; Revised 5 July, 2022; Accepted 5 July, 2022

Available online 5 July, 2022 at www.atlas-tjes.org, doi: 10.22545/2022/00188

There is no gain-saying that machines can learn from data to derive patterns and insights to aid various applications, also known as artificial intelligence, and fast-gaining relevance today. This study implemented feature data generation (FDG) as a novel technique for psychometrics improvements using the Post-hoc Simulation approach. The descriptive design of the correlation type was adopted for this study and deployed quantitatively. The instrument for the study was a test aligned to the behavioral objectives of the Postgraduate Certificate Curriculum with the program enrollees as the study participants. The test underwent a thorough validation procedure which yielded a reliability coefficient of 0.98. The item parameters of the test were analyzed using XCalibre 4.2 to analyze the real data from 38 respondents, while the WINGEN application through the post-hoc approach was used to generate the simulated data with 500 respondents. The findings of the study revealed that the 3 Parameter Logistic Model fit the generated data determined using chi-square goodness of fit statistics. The FDG is a viable approach with a strong and positive correlation between real data and simulated data, which enables the generalization of findings on the basis on which conclusions were made. The developed FDG method for psychometric improvement has wide applicability, a plus for the novel technique while strengthening transdisciplinary research.

Keywords: Artificial Intelligence, machines learning, feature data generation, psychometrics, post-hoc simulation, transdisciplinary research.

1 Introduction

Psychometrics is the development and use of observational procedures yielding quantitative data to measure psychological attributes such as aptitudes, attitudes, and abilities applicable to multiple fields. The attributes can be measured using various instruments, including tests, questionnaires, checklists, and rating scales, to mention a few (Institute of Medicine, 2015). These instruments can either be deployed manually (paper-pencil) or via computer terminals popularised by the advent of technology. One of such platforms is Computer Adaptive Testing (CAT-a second-generation technology with computer-based testing), which requires many items. Developing such an item bank for CAT is time-consuming on the path of item developers and examinees' sitting time required for item validation purposes. This challenge calls for Artificial Intelligent (AI) based interventions as a transdisciplinary approach to real-world problem-solving as envisaged by the conceptual framework for this study.

2 Literature Review

Data science collects data premised on standardized processes for preparing the data (texts, numeric, graphic, imagery, transitionary, visuals, among others) into meaningful form through classification, scoring or estimation for making predictions and decisions referred to as machine learning for various applications. Features in machine learning refer to the inputs to machine learning models or numerical representations of raw data. Feature generation is a method for data pre-processing where new features are extracted from raw data, which can be transformed into suitable formats for the machine learning models or further statistical analysis (van den Bosch, 2017; Xiao, 2021). Feature Generation takes one or more attributes from a dataset and creates a new "feature" from them (RapidMiner, 2016). Feature data generation (FDG) creates new features from one or multiple existing features, potentially for use in statistical analysis and machine learning. This process adds new information to be accessible during the model construction and hopefully results in a more accurate model (Cohen, 2019). Blum and Langley (1997) stressed selecting only the relevant features as a central problem in machine learning in predicting novel test cases. FDG is apt for psychometric analysis, which requires analyzing big data, an important research field contributing to the development of 4IR technologies (Rustam & Kasmawati, 2021).

Psychometric analytics refers to the statistical analysis of test data to evaluate the quality of measurement by closely examining the performance of items, sub-scores, and the test as a whole (Assessment Systems, 2022). A major feature of psychometric analytics is item parameters. Regarding ability estimation, the relevant parameters are difficulty, discrimination, guessing, and the theta (θ) (Schonpflug, 2000). An ample number of these parameters usually referred to as psychometric data are required to generalize psychometric research (McEwan, 2020). Artificial Intelligence (AI) influences people and businesses on a massive scale. It has become an inseparable component of our lives today, making living simpler, as evident with smartphones used to navigate around the city, with live insights on traffic, suitable and fastest routes, and other recommendations and virtual digital assistants such as Cortana or Alexa are making our lives simpler than ever (GreatLearning, 2022).

AI, sometimes called machine intelligence, has been applied in speech recognition, learning, planning, and problem-solving. The field of AI has evolved from humble beginnings to a field with global impact fueled by the fourth industrial revolution (4IR), which ushers the world into a technological revolution that is fast changing the way people function with unprecedented effects already experienced (Bartneck et al., 2021). Considering this reality, its response must be integrated and comprehensive, involving all stakeholders of the global polity, from the public and private sectors to academia and civil society (Schwab, 2016). AI addresses the crucial questions of what knowledge is required in any aspect of thinking, how should that knowledge be represented; and how should that knowledge be used (Mohammed, 2009). AI has been applied to automated item generation for CAT in psychometrics, but this application is at the testing stage. This study is geared towards applying AI at the validation stage as an identified gap in psychometrics research.

Data Science and Machine Learning leverage AI platforms for intelligent applications as the most dominant technology in the 4IR (Marwala, 2020). Waykole and Sharma (2022) leveraged on data science to develop a novel data hiding technique using an image as cover media. Other platforms offer pre-built algorithms and simplistic workflows with such features as drag-and-drop modeling and Graphical user interfaces (GUI) that easily connect necessary data to the end solution (Han & Hambleton, 2014). In contrast, others require a greater knowledge of development and coding. These platforms combine intelligent decision-making algorithms with data, enabling developers to create solutions in various areas (G2, 2022). To qualify for inclusion in the AI Platforms category, a product must be capable of intelligence building, allow users to create machine learning algorithms and offer pre-built machine learning algorithms for more novice users to build applications, and present a way for developers to connect data to the algorithms for them to learn and adapt (University of Johannesburg Library, 2022).

Considering the relationship between psychometrics and data analytics, a prime motivation of computational psychometrics is dealing with complex performance situations for more psychologically sophisticated views of students' capabilities (Mislevy & Bolsinova, 2021; Rustam & Kasmawati, 2021). In simulation-based tasks, for example, quite voluminous data can be captured about every detail of a student's actions; with each mouse click, time-stamped, keystroke and in some cases, heart rates, eye fixations, and physical locations. Therefore, reasoning from micro-features of performance to higher-order inferences about students' capabilities which can be quite challenging, is made possible through simulation. There are three simulation approaches Monte-Carlo, Post-hoc, and Hybrid (Oladele, 2021). While these three methods were valid for data generation, the results derived from MC simulations may not always be generalizable to operational studies. This important factor makes the post-hoc or hybrid simulations better options to overcome this problem. With the Post-hoc simulations approach, real item-response vectors, gathered from either a paper-pencil test or operational adaptive test, are used to run simulations (Thompson & Weiss, 2011). The Post-Hoc simulation was employed by Seo and Choi (2018), while Halil (2020) studied post-hoc vs Hybrid Simulations revealing that both methods generally resulted in comparable outcomes.

Transdisciplinarity, which is about four decades old, has emerged due to a critique of the standard configuration of knowledge outlined in the curriculum of various disciplines while taking cognisance of moral and ethical concerns (Bernstein, 2015; Oladele, 2022a). Bernstein (2015) further stressed the need for a paradigm shift from individual to systemic thinking as the solution to new, highly complex, global concerns, beginning with climate change and sustainability and extending into many fields such as science, technology, social problems, and policy, education, and the arts. Transdisciplinarity involves work that creatively re-imagines the disciplines and the possibilities for combining them while challenging the entire framework of disciplinary thinking and seeks to assemble new approaches from scratch, using materials from existing scholarly disciplines for novel purposes. Transdisciplinarity is seen as a developer's holistic view of reality as filtered by that developer's sensory input and perception of that reality and this opens up limitless possibilities for the mind with respect to problem-solving applicable to different contexts of reasoning (Denard, 2021).

Psychometric models use mathematical structures to model observable data patterns that reveal peoples' psychological capabilities or tendencies in relevant situations using theoretical methods (Mislevy & Bolsinova, 2021). The authors described Item Response Theory (IRT) as a generative model, which advances unobservable or latent variables associated with persons, upon which observable variables such as item responses are posited to depend scholastically through a so-called link function on the nature of the data, the targeted inferences. Similarly, Progar et al. (2008) described IRT as a model-based paradigm that starts with modeling the relationship between the latent variable being measured and the item response. According to Fredriksson et al. (2021), IRT corresponds to a family of statistical models developed to evaluate how latent traits of students (such as intelligence) when evaluated by a set of items (an exam). Therefore, the basic idea of IRT lies in the performance of an examinee, which can be predicted by the latent trait and the relationship between the performance of the test item and the set of characteristics underlying the item's performance explained by the item characteristic function or item characteristic curve (Rustam & Kasmawati, 2021). CAT is a simple example of the modular assembly for probability-based inference in observational situations that psychometric latent-variable models afford. Since then, IRT has been standard practice in developing psychometric scales with standardized examinations, while recent research employs techniques for assessing datasets for machine learning research (Fredriksson et al., 2021).

Mislevy and Bolsinova (2021) reiterated that the first is the notion of a "psychometric backbone"; that is, a persistent variable (*theta*) for characterising higher-level psychological constructs, which are posited to hold meaning for understanding performance across multiple real or hypothetical specific situations. The second is serving in stages in evidence-identification, in a multilayered chain of reasoning from low-level data to indicators of psychologically relevant performance patterns to serve as high-level variables that enter psychometric models. This study aimed to leverage AI possibilities for generating ample and higher-level data for psychometric analysis using the WINGEN application for simulation. WINGEN qualify as an AI platform capable of intelligence building with in-built algorithms. It also has a GUI for connecting data to the algorithms for learning and adaptation.

3 Conceptual Framework

The need for transdisciplinary thinking has been raised owing to the rapid convergence of technologies such as artificial intelligence, brain-like computer chips, bioengineering, and intelligent pharmacology while living in the reality of complex societal problems (Yeh, 2019). In this study, transdisciplinarity is conceptualized using the four philosophical research Nicolescuian axiom, which consists of multiple levels of reality mediated by the Hidden Third (ontology); knowledge as complex, emergent, cross-fertilized and embodied (epistemology); inclusive logic (the logic of complexity) to facilitate contradiction reconciliation; and transdisciplinary value formation (axiology) (McGregor, 2018) for addressing the complexities facing educational psychometric research as shown in Figure 1.

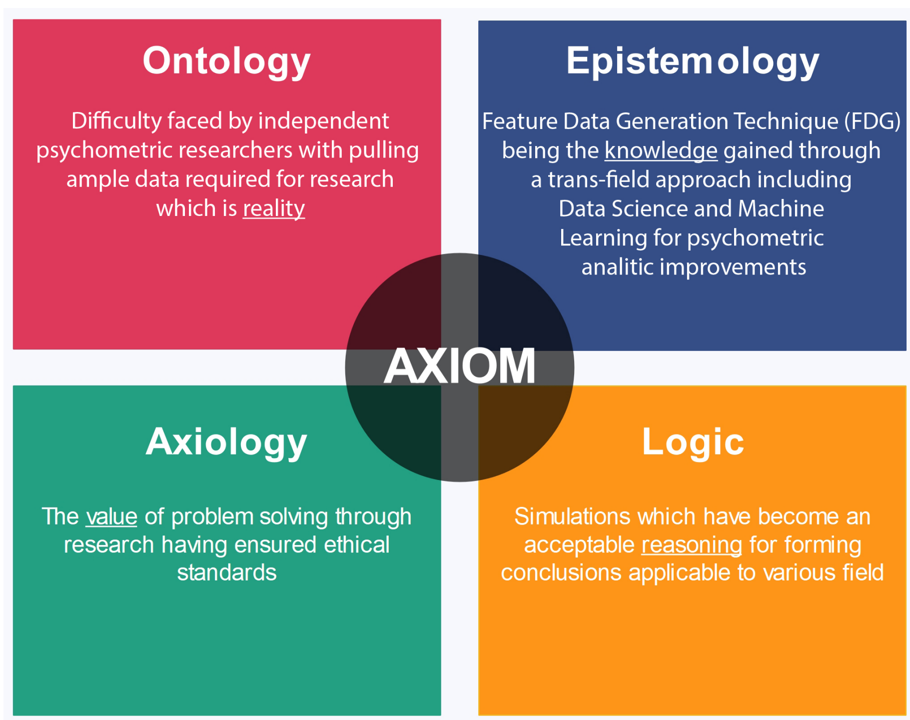


Figure 1: *Four philosophical research axioms (Adapted from McGregor, 2018).*

As shown in Figure 1, the first square, named ontology, is related to the realities of difficulty researchers face with gathering sufficient data required for psychometric research. Unfortunately, this reality applies to researchers in Africa as those in developed countries have access to large-scale assessments such as the program for international student assessment (PISA), progress in international reading literacy study (PIRLS), and trends in international mathematics and science study (TIMSS) among others. The need for Africans to own solutions to Africa's challenges has become necessary (Fiquremariam, 2008; Momoh, 2016). Research is a sure direction for inventing Africa-appropriate solutions in the face of the continents' realities in line with the axiom of ontology. The second square, named epistemology, as adapted for this study, speaks to the novel FDG approach, which leverages on data science and machine learning aspect of AI and provides a solution for generating larger data sets from smaller ones for improving psychometric research outcomes in Africa. Technology is regarded as a solvent for Africa's Biggest Problems (Bernstein, 2015; Tachev, 2020). Leveraging technology gain ensures a trans-field approach to knowledge generation with empirical evidence. This solution is apt considering that new developments in AI-related educational assessment are attracting increasing interest to improve assessment efficacy and validity, contributing to analyzing large volumes of process data being captured from digital assessment contexts (Gardner et al., 2021). The third square, named Axiology, as adapted for this study, reiterates the need for values through ethical considerations while solving problems through research in the aspects of non-coercion of participants and assessment sitting time and ensuring data model fit required for ascertaining the appropriateness of the measurement model employed for the generated data. The issue of measurement misfit would raise serious concerns which could invalidate any "invention", no matter how lofty and promising it is. Lastly, the fourth square on logic as adapted to this study leverage the power of the post-hoc simulations approach, which draws on information from real data samples in the data generation process as the main goal of the study, which allows for generalization through which conclusions are drawn, Thompson and Weiss (2011) which occupies the novelty seat of this research. The forgoing clarifies how the concepts of ontology, epistemology, axiology, and logic, as propounded by McGregor (2018), translate to transdisciplinary research as applied to computational psychometrics for educational assessments. Mokiy & Lukyanova (2022) stressed that meaningful change does not necessarily involve a one-size-fit-all approach. Rather, general patterns within this process are discernable through a broad perspective offered by science and in the context of psychometric research.

4 Problem Statement and the Gap

Psychological testing, which quantifies human characteristics, behavior or performance, requires large data (Statistics.com, 2022). Data obtained from large-scale testing in advanced countries are relatively accessible for research. However, the reverse is the situation in developing countries and worst still with independent researchers who are left struggling to garner sufficient data within a reasonable length of time to ensure sampling validity related to psychometric research with implications on ethicality where participants cannot be coerced, sitting time and time spent on research data collection (Guo et al., 2013; Anthoine et al., 2014; Martin & Martin, 2017; He at al., 2021). This ordeal can be remedied using AI. Feature data generation (FDG) for machine learning has been applied to various fields such as email spam filtering in communication, credit card fraud detection and digit recognition on checks in finance, facial recognition and scene classification in security, MRI image analysis in medicine, recommendation system, search engines in research and handwriting recognition applicable to various fields (G2, 2022). Concerning the applications of FDG to educational assessment, Straehle et al. (2018) study centered on high-level strategic approaches for conducting big data studies in assessment, He et al. (2021) illustrated how processed data can be used to identify behavioral patterns in a computer-based problem-solving assessment. This study is poised to apply FDG to psychometrics for improving educational assessment using the Post-hoc Simulation approach while answering the following research questions:

1. What are the item parameters of the simulated data from the study?
2. How viable was the novel FDG method?

5 Methodology

Design: This study is non-experimental with a descriptive research design of correlation type approached quantitatively. This research design enables the researcher to measure and describe the degree of association among variables or sets of scores (Creswell & Creswell, 2017; Asenahabi, 2019). This research design is appropriate as it attempts to find relationships between the features of real and simulated item parameters for optimising feature generation (Cohen, 2019).

Measurement Model: This study adopts the 3 Parametric Logistic Model, which is premised on IRT with the item difficulty (b), discriminating (a), and pseudo-guessing (c) parameters informed by subjecting the generated data to chi-square goodness of fit statistics employed in determining how well the observed data correspond to the fitted (assumed) model as shown in Table 1.

Table 1: *Model Fit Statistics of the generated IRT parameters.*

Chi-square value	*df*	*p-value*
13.55318	12	0.42127

As shown in Table 1, the overall chi-square value for the generated IRT parameters was 13.55 (put to 2 decimal places) with a degree of freedom of 12 and a p-value of 0.42. Considering that the p-value of 0.42 is greater than the significant value of 0.05, the generated item fits the projected 3-PLM. Also presented below is a systematic selection of some of the item response function (IRF) plots (the complete output is available on request). This IRF is produced by plotting the observed proportion correct using a black line, and the fit line is plotted using a red line for evaluating the fit of an item as it asymptote toward zero. IRFs for items 40, 80, 120, 160, and 200 through a systematic sampling of nth interval=40 is shown in Figure 2a to 2e.

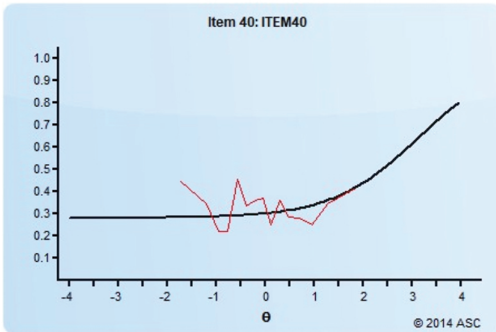
Figure 2a to e shows that the chosen three-parameter logistic model fit line asymptotes toward zero. As such, discrimination, difficulty, and pseudo-guessing parameters were relevant for estimating students' ability levels using the multiple-choice test format.

Participants: The study population were Postgraduate Certificate in Education (PGCE) and Postgraduate Diploma in Education (PGDE) students enrolled for the 2021 academic session in a South African and Nigerian university. This set of students was deemed appropriate as they have been fully exposed to the PGCE and PGDE Mathematics curriculum, which informed the developed item bank. The participants responded to multiple-choice items (a total of 68 items) from which the base data was obtained. A total of 38 respondents volunteered in the study while adhering to the ethical standards of human participants.

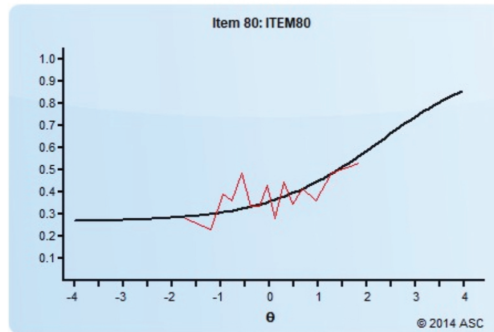
Instrumentation: The instrument for the study was a multiple-choice test with sixty-eight items and a 4-optioned response format premised on the Postgraduate Diploma in Education (PGDE) curriculum. The test went through the face and content validation and yielded alpha reliability of 0.98, a high-reliability index (McCowan & McCowan, 1999). Also, the items constructed on each of the content areas across the cognitive learning domains were based on Bloom's taxonomy of educational objectives (McCowan & McCowan, 1999; Anderson & Krathwohl, 2001).

Feature Generation Technique: The data generation was carried out in two stages. In the first stage, the administered test was scored dichotomously (1 for correct answers and 0 for incorrect answers). Psychometric analysis was carried out to calibrate the a, b, and c parameters for the sixty-eight multiple-choice items from thirty-five volunteered participants and their θ estimates. In the second stage, Feature data generation (FDG) was carried out using the Post-hoc Simulation approach. With FDG, new features were generated from existing ones. In deploying FDG for this study, the relevant features were the a, b, and c- parameters and their θ estimates. Therefore, the parameters (from 68 items) and θ estimates (from 35 respondents) were simulated to produce more robust data of 200 items from 500 responses.

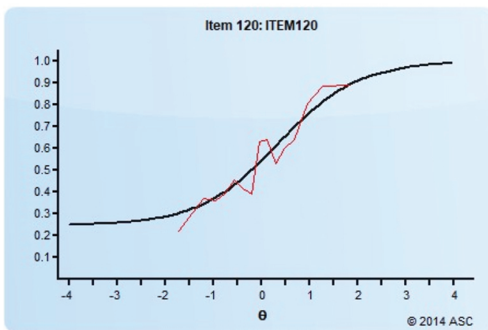
Data Analysis Techniques: The psychometric analysis was performed using Xcalibre 4.2 (Licensed version) to answer the first research question. The simulated data was generated using WINGEN (a software for generating examinee/item parameters and responses) to answer the second research question. The summary statistics of the item parameters were subjected to Pearson correlation statistics to test the viability of the FDG method using SPPS 27.0 to answer the third research question.



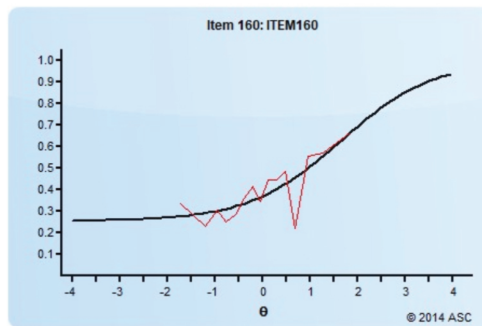
(a)



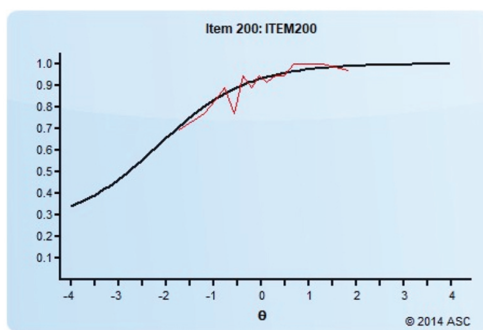
(b)



(c)



(d)



(e)

Figure 2: *(a) Item Response Functions (Item 40), (b) Item Response Functions (Item 80), (c) Item Response Functions (Item 120), (d) Item Response Functions (Item 160), (e) Item Response Functions (Item 200).*

6 Feature Data Generation Procedure and Results

Preliminary Information: The 68 multiple-choice items developed are shown in Oladele (2022b).

In deploying the FDG technique, the 68 items were subjected to psychometric analysis, and item 66 was removed due to no variance as all participants answered the item correctly. As such, 67 items were utilized for the procedure. The item parameters a, b and c and their θ estimates obtained from the psychometric analysis serve as the base data for the study shown in Table 2 and Table 3.

Table 2 presents the summary statistics for the item a-discrimination, b-difficulty, and c-guessing parameters for all calibrated items, while Table 3 summarizes the theta estimates for the full test.

Table 2 shows that the a-parameter had a mean of 0.73 and standard deviation of 0.086, the b-parameter

Table 2: *Summary statistics for all calibrated items.*

Parameter	Items	Mean	SD	Min	Max
a	67	0.732	0.086	0.558	0.915
b	67	0.789	1.503	-2.507	3.476
c	67	0.246	0.015	0.199	0.336

Table 3: *Summary statistics for the theta estimates.*

Test	Examinees	Mean	SD
Full Test	35	-0.030	0.8497

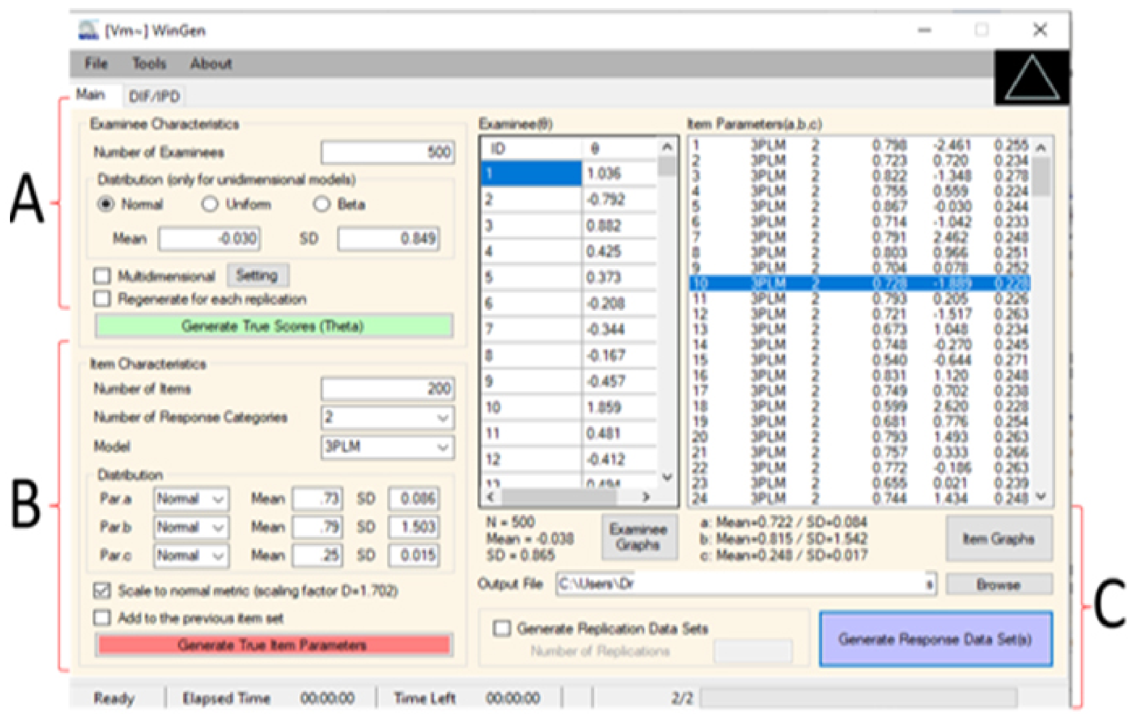


Figure 3: *: WINGEN Screen inputs for generating examinee/item parameters.*

had a mean of 0.79 and standard deviation of 1.503, while the c-parameter had a mean of 0.25 and standard deviation of 0.0156. Also, a normal distribution with a mean of -0.030 and a standard deviation of 0.849 (Table 3). These parameters were used to generate 500 examinees and theta scores for 200 items using WINGEN (a software for generating item responses) specified in the fields A and B, as shown in Figure 3.

As shown in Figure 3, the parameters obtained from the psychometric analysis (see Table 2) were entered into the required fields (A- examinee parameters and B- item parameters). Once the inputs have been made, the researcher would specify the output file and then click on the "Generate Response Data Set" button as shown in field C. The summary statistics from the output is shown in tables 4 and 5 for answering the research questions generated for this study

Table 4: *Summary statistics for all calibrated items.*

Parameter	Items	Mean	SD	Min	Max
a	197	1.261	0.752	0.231	4.348
b	197	0.990	1.460	-2.666	4.00076
c	197	0.252	0.013	0.117	0.2956

Table 5: *Summary statistics for the theta estimates.*

Test	Examinees	Mean	SD
Full Test	500	0.004	0.962

6.1 Answering Research Questions

Research Question 1: What are the item parameters of the simulated data from the study?

Of the 200 simulated items, items 1, 2 and 3 were removed leaving 197 items (as shown in Table 4). The items were removed due to no variance being answered correctly by all participants. Also, Table 4 shows that the a-parameter had a mean of 1.261 and a standard deviation of 0.752; the b-parameter had a mean of 0.99 and a standard deviation of 1.460, while the c-parameter had a mean of 0.25 and standard deviation of 0.013. As shown in Table 5, the simulated data for the full test had a normal distribution with a mean of 0.004 and a standard deviation of 0.962.

Research Question 2: How viable was the novel FDG method?

The viability of the FDG method was tested by the strength and pattern of the relationship between the summary statistics for the item parameters of the real and simulated data using Pearson's correlational statistics. Results are shown in Table 6.

Table 6: Pearson's correlational statistics of the Summary Statistics for the Real & Simulated Data

Correlations	**a-Discrimination**		**b-Difficulty**		**c-Pseudo-Guessing**		**Theta Estimate**	
	Real	Simulated	Real	Simulated	Real	Simulated	Real	Simulated
Real	1	.669	1	.998	1	0.954	1	1.000
Simulated	.669	1	.998	1		0.046	1.000	1

As shown in Table 6, the correlation coefficients for the real and simulated data were a-Discrimination: 0.7, b-Difficulty: 1, c-Pseudo-guessing: 1, and the Theta Estimate: 1. These results show a strong correlation between the two data sets. Also, positive correlation coefficients were obtained, which connotes that the values from the simulated data were higher than those obtained from the real data. This result connotes that the novel FDG method for psychometrics was viable.

7 Discussion

The study's findings revealed that the values produced from the real and simulated data were both within the specified ranges for the b, a, and c parameters. This result shows that the items of desirable difficulty would discriminate correctly between the high and low-ability candidates while giving no room for guessing. McCowan and McCowan (1999) stressed the importance of items' quality to improve test reliability. Furthermore, the c-parameter from the simulated data provided a better pseudo index. Considering that guessing is inevitable with multiple-choice items, this parameter becomes important for psychological testing. Lastly, the persistent variable (Theta estimate-θ), regarded as the psychometric backbone for characterizing higher-level psychological constructs, aids the understanding of performance across multiple real or hypothetical specific situations were positively correlated, which served as a multilayered chain of reasoning from low-level data to high-level variables that for the psychometric models (Mislevy & Bolsinova, 2021).

This study was premised on the IRT model. Considering that IRT is an item model-based paradigm with emphasis on the relationship between the latent variable being measured and the item response, the proposed method is item based, which is alienable to the IRT; more so that IRT has been established as empirically superior to CTT parameters once fitted rightly and gaining popularity in many other fields, such as educational research, health sciences research, quality-of-life research, and even marketing research (Progar et al., 2008; An & Yung, 2014). The model fit is another important aspect of the IRT modeling, which is usually chosen arbitrarily and, at best, subjectively at the initiation stage required for modeling the relationship between the performance of the test item and the set of characteristics underlying the item's performance (Rustam & Kasmawati, 2021). Considering that the test is a multiple-choice test, all the three parameters of discrimination, difficulty, and pseudo-guessing were regarded as germane, which informed the choice of 3PLM. This study's findings show that the projected 3PLM fitted the generated data. This finding is in line with that of Zhang et al. (2020) and Yu et al. (2020), who revealed that the item parameters in the multiple-choice test aligned with the 3PL model. While the two-parameter logistic model is often used because of its nice mathematical properties and plausible stochastic response mechanisms, Yu et al. (2020) stressed that a model is a good approximation in a real-world application, and researchers must be aware of the risk of model misspecification. This finding is not without implication for the appropriateness of multiple-choice tests, generally regarded as less rigorous than the essay forms of a test as accommodating the pseudo-guessing index would cater to the tendencies to choose from the available options arbitrarily. Students' ability levels can be accurately estimated using the multiple-choice test format once carefully designed and following standardized test development procedures.

This study also revealed that the novel FDG method is viable considering the high and positive correlation coefficient between the real and simulated data. Also, there is a perfect alignment between the real and simulated data features. This finding aligns with ensuring that only the relevant features were included as a central problem in machine learning (Blum & Langley, 1997). Furthermore, the novel FDG procedure has proved efficient for generating new information, resulting in a more accurate model for an optimal feature-generating technique (Cohen, 2019). FDG can be referred to an "out of the box" approach for psychometric analysis Yeh (2019), which requires analyzing big data, an important research field contributing to internet and web technologies (Rustam & Kasmawati, 2021). This finding implies that the FDG method guarantees psychometric improvements while easing the associated stress of gathering data for psychometric analysis (Martin & Martin, 2017; He et al., 2021).

8 Conclusion

This study concluded that the novel FDG method is viable for providing ample psychometric data with a higher chain of reasoning which is a requirement with artificial intelligence while strengthening the generalizability of psychometric research. Most importantly, the novel FDG method deployed transdisciplinarily can be applied to various disciplines considering that psychometrics as a discipline is cross-cutting and applied in various institutions, research, and continents.

Acknowledgment

The author is grateful to Dr Nathan Thompson of Assessment Systems, USA, for supporting this research through a generous provision of access to FastTest workspace (an online computer-based online platform) used for item writing, banking, and pilot-testing as well as X-Calibre 4.2 used for analyzing the collected data. The author also acknowledges Dr. Han, K. T., who designed the WINGEN software available for research at no financial cost.

Funding: This research received no specific grant from any funding agency.

Conflicts of Interest: The author declared no potential conflicts of interest concerning this article's research, authorship, and publication.

References

[1] An, X., & Yung, Y. F. (2014). Item response theory: What it is and how you can use the IRT procedure to apply it. *SAS Institute Inc,* 10(4), 364-2014. https://bit.ly/33kKgCn

[2] Anderson, L. & Krathwohl, D.A. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives.* Longman.

[3] Anthoine, E., Moret, L., Regnault, A., Sébille, V., & Hardouin, J. B. (2014). Sample size used to validate a scale: a review of publications on newly-developed patient-reported outcomes measures. *Health and quality of life outcomes,* 12(1), 1-10. https://hqlo.biomedcentral.com/track/pdf/10.1186/s12955-014-0176-2.pdf

[4] Asenahabi, B. M. (2019). Basics of research design: A guide to selecting appropriate research design. *International Journal of Contemporary Applied Researches,* 6(5), 76-89. http://www.ijcar.net/assets/pdf/Vol6-No5-May2019/07.-Basics-of-Research-Design-A-Guide-to-selecting-appropriate-research-design.pdf

[5] Assessment Systems (2022). What is Psychometric Analytics? https://assess.com/psychometric-analytics/

[6] Bartneck, C., Lütge, C., Wagner, A., & Welsh, S. (2021). *An introduction to ethics in robotics and AI,*(p. 117). Springer Nature. https://doi.org/10.1007/978-3-030-51110-4_2

[7] Bernstein, J. H. (2015). Transdisciplinarity: A review of its origins, development, and current issues. *Journal of Research Practice,* Volume 11(1), 1-20. http://jrp.icaap.org/index.php/jrp/article/view/510/412

[8] Blum, A. L., & Langley, P. (1997). Selection of relevant features and examples in machine learning. *Artificial intelligence,* 97(1-2), 245-271. https://doi.org/10.1016/S0004-3702(97)00063-5

[9] Cohen, I. (2019). Optimising Feature generation. Medium, October 31. https://bit.ly/3guGBow

[10] Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches.* Sage publications. http://www.ceil-conicet.gov.ar/wp-content/uploads/2015/10/Creswell-Cap-10.pdf

[11] Denard, S. (2021). Development Cycle Modeling: Transdisciplinary Implications. *Transdisciplinary Journal of Engineering & Science,*12, 100-118. https://doi.org/10.22545/2021/00159

[12] DeMars, C. (2010). *Item response theory. Understanding statistics measurement.* Oxford University Press.

[13] Fiquremariam, S. (2008). African Solutions to African Problems. Institute for Security Studies, September, 18. https://issafrica.org/iss-today/african-solutions-to-african-problems#: :text=The%20catchall%20phrase%20%E2%80%9CAfrican%20solutions,acquired%20a%20degree%20of%20autonomy.

[14] Fredriksson, T., Mattos, D. I., Bosch, J. & Olsson, H. H. (2021). Assessing the Suitability of Semi-Supervised Learning Datasets using Item Response Theory. 2021 47th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)|978-1-6654-2705-0/21/. http://doi.org/1109/SEAA53835.2021.00049

[15] G2 (2022). Best Data Science and Machine Learning Platforms. https://www.g2.com/categories/data-science-and-machine-learning-platforms

[16] Gardner, J., O'Leary, M. & Yuan, Li (2021). Artificial intelligence in educational assessment: 'Breakthrough? Or buncombe and ballyhoo?' *Journal of Computer Assisted Learning,* 37, 1207–1216. http://doi.org/10.1111/jcal.12577

[17] Guo, H., Pohl, E., & Gerokostopoulos, A. (2013, January). Determining the right sample size for your test: theory and application.*In Reliability and Maintainability Symposium.* https://www.weibull.com/pubs/2015_RAMS_right_sample_size.pdf

[18] GreatLearning (2022). Artificial Intelligence & Machine Learning. www.mygreatlearning.com/pg-program-online artificial-intelligence-machine-learning

[19] Han, K. T. & Hambleton, R. K. (2014). User's Manual for WINGEN: Windows Software that Generates IRT Model Parameters and Item Responses. University of Massachusetts Amherst. Retrieved from https://www.umass.edu/remp/software/simcata/wingen/WinGenManual2014.pdf

[20] Halil, S. A. R. İ. (2020). Testing Multistage Testing Configurations: Post-Hoc vs Hybrid Simulations. *International Journal of Psychology and Educational Studies,* 7(1), 27-37. https://doi.org/10.17220/ijpes.2020.01.003"

[21] He, Q., Borgonovi, F., & Paccagnella, M. (2021). Leveraging process data to assess adults' problem-solving skills: Using sequence mining to identify behavioural patterns across digital tasks. Computers & Education, 166, 104170. https://doi.org/10.1016/j.compedu.2021.104170

[22] Institute of Medicine. (2015). Psychological testing in the service of disability determination. https://www.ncbi.nlm.nih.gov/books/NBK305233/

[23] Martin, C. R., & Martin, C. J. H. (2017). Minimum sample size requirements for a validation study of the birth satisfaction scale-revised (BSS-R). *Journal of Nursing and Practice*, 1(2), 25-30. https://www.napier.ac.uk/ /media/worktribe/output-1022470/minimum-sample-size-requirements-for-a-validation-study-of-the-birth-satisfaction.pdf

[24] Marwala, T. (2020). Closing the Gap: *The Fourth Industrial Revolution in Africa.* Macmillan.

[25] McCowan, R. J., & McCowan, S. C. (1999). *Item Analysis for Criterion-Referenced Tests.* Online Submission. https://files.eric.ed.gov/fulltext/ED501716.pdf

[26] McEwan, B. (2020) Sampling and validity. *Annals of the International Communication Association,* 44(3), 235-247. https://doi.org/10.1080/23808985.2020.1792793

[27] McGregor, S. L. (2018). Philosophical underpinnings of the transdisciplinary research methodology. *Transdisciplinary Journal of Engineering & Science,* 9, 182-198. https://doi.org/10.22545/2018/00109

[28] Mislevy, R. J., & Bolsinova, M. (2021). Concepts and Models from Psychometrics. In *Computational Psychometrics: New Methodologies for a New Generation of Digital Learning and Assessment* (pp. 81-107). Springer, Cham. https://link.springer.com/chapter/10.1007/978-3-030-74394-9_6& Science, 13, 143-158 . https://doi.org/10.22545/2022/00184

[29] Mokiy, V., & Lukyanova, T. (2022). Prospects of Integrating Transdisciplinarity and Systems Thinking in the Historical Framework of Various Socio-Cultural Contexts. *Transdisciplinary Journal of Engineering & Science,* 13, 143-158 . https://doi.org/10.22545/2022/00184

[30] Mohammed, Z. (2019). Artificial intelligence definition, ethics and standards. *Electronics and communications: Law, standards and practice.* https://www.researchgate.net/publication/332548325_Artificial_Intelligence_Definition_Ethics_and_Standards #: :text=Artificial%20intelligence%20(AI)%20is%20a,understanding%20and%2For%20logical%20reasoning.

[31] Momoh, Z. (2016). African Solutions to African Problems: A Critical Appraisal. *Journal of African Union Studies (JoAUS),* 5(1), 39-62. https://www.researchgate.net/publication/319290174_African_Solutions_to_African_Problems_A_Critical_Appraisal

[32] Oladele, J. I. (2021). Computer-Adaptive-Testing Performance for Postgraduate Certification in Education as Innovative Assessment. *International Journal of Innovation, Creativity and Change*, 15(9), 87-101. https://www.ijicc.net/images/Vol_15/Iss_9/15929_Oladele_2021_E1__R.pdf

[33] Oladele J. I. (2022a). Estimating Measurement Error in University Students' Work Experience Programme (SWEP) Assessment using Generalizability Theory: Implications for Transdisciplinarity in Engineering Training. *Transdisciplinary Journal of Engineering & Science*, 13, 159-171. https://doi.org/10.22545/2022/00185.

[34] Oladele, J. I. (2022b). Feature Data Generation Item Analysis Data. Retrieved from https://osf.io/8xkr7/files/osfstorage

[35] Progar, Š., Sočan, G., & Peč, M. (2008). An empirical comparison of item response theory and classical test theory. *Horizons of Psychology,* 17(3), 5-24. http://psiholoska-obzorja.si/arhiv_clanki/2008_3/progar_socan.pdf

[36] Rustam, A., & Kasmawati. (2021). Development of Numbers Material Test Using the Item Response Theory (IRT) Approach for Elementary School Students. *Journal of Mathematics Education,* 6(2), 78-86. http://doi.org/10.31327/jme.v6i2.1585

[37] Schonpflug, W. (2000) Geschichte und Systematik der Psychologie (History and Structure of Psychology). Beltz/Psychologie Verlags Union, Weinheim, Germany

[38] Schwab, Klaus. (2016). *The Fourth Industrial Revolution.* World Economic Forum.

[39] Seo, D. G., & Choi, J. (2018). Post-hoc simulation study of computerised adaptive testing for the Korean Medical Licensing Examination. *Journal of Educational Evaluation for Health Professions,* 15. https://doi.org/10.3352/jeehp.2018.15.14

[40] Statistics.com (2022). Psychometrics. https://www.statistics.com/glossary/psychometrics/

[41] Straehle M. M., Munson, L. J., Fossey, A. & Kim, E. (2018) studied centered on high-level strategic approaches for conducting big data studies in assessment, In H. Jiao, R. W. Lissitz, and Wie, A.Van (Eds.). *Data analytics and psychometrics: Informing assessment practices.* Information Age Publishing, Inc.

[42] Tachev, V. (2020). How Technology Is Solving Africa's Biggest Problems. International Investor, February 18. https://www.intinvestor.com/news/how-technology-is-solving-africas-biggest-problems/

[43] Thompson, N.A., & Weiss, D.A. (2011). A framework for the development of computerised adaptive tests. *Practical Assessment, Research, and Evaluation*, 16(1): 1. https://doi.org/10.7275/wqzt-9427

[44] Xiao, Z. (2021). Feature Engineering for Categorical Data. June 28, Medium. https://medium.com/geekculture/feature-engineering-for-categorical-data-a77a04b3308

[45] University of Johannesburg Library (2022). 4IR Technologies: Artificial Intelligence. https://uj.ac.za.libguides.com/4ir/artificial#gsc.tab=0

[46] van den Bosch, S. (2017). Automatic feature generation and selection in predictive analytics solutions. Master's thesis, Faculty of Science, Radboud University. https://bit.ly/3rQyIjM

[47] Waykole, S., & Sharma, A. (2022). Reversible Secured Data Hiding using Binary Encryption and Digital Bit Modification Scheme. *Transdisciplinary Journal of Engineering & Science,* 13. https://doi.org/10.22545/2022/00181

[48] Yeh, Ph.D, R. T. (2019). Towards a Framework for Transdisciplinary Problem Solving. *Transdisciplinary Journal of Engineering & Science,* 10, 9-17. https://doi.org/10.22545/2019/0111

[49] Yu, X., Li, S. & Cheng. J. (2020). A three-parameter logistic regression model. *Statistical Theory and Related Fields,* 5(3), 1-10. https://doi.org/10.1080/24754269.2020.1796098

[50] Zhang, Z., Zhang, J., Tao, J., & Shi, N. (2020). A General Three-Parameter Logistic Model With Time Effect. *Frontiers in Psychology,* 11, 1791. https://doi.org/10.3389/fpsyg.2020.01791

About the Author

Jumoke I. Oladele holds a permanent teaching position with the University of Ilorin, Nigeria, and is currently on her postdoctoral research fellowship with the University of Johannesburg, South Africa. Her specialization is in educational evaluation and psychometrics. Her research interests are in item response theory-based psychometrics in educational assessments with computer adaptive testing, applicable to a wide range of subjects focusing on improving standardized assessment in Africa, educational evaluation, and innovation studies. So far, her research has generated a good number of publications in national/international journals. She is the initiator of the Association of Computer Adaptive Testing in Africa, of which she is a member and serves as the research/membership director. She is also a member of other learned societies such as the Teacher's Registration Council of Nigeria, the Association of Educational Researchers and Evaluators of Nigeria, and the International Association for Computerized Adaptive Testing. For her, transdisciplinarity is an evolving area of research interest driven by the need for more holistic views of problem-solving as applicable to pedagogical and assessment practices in higher education. Her contact email is jumokeo@uj.ac.za