International Journal of Innovation, Creativity and Change. www.ijicc.net
Volume 15, Issue 9, 2021

Computer-Adaptive-Testing Performance for Postgraduate Certification in Education as Innovative Assessment

Jumoke I. Oladele, University of Johannesburg, South Africa, Email: jumokeo@uj.ac.za

Learning in higher institutions of learning has moved online in a bid to keep students engaged amidst the lockdown. It becomes necessary to engage innovation-driven assessment solutions in ascertain the extent to which learning outcomes are being achieved. Computer Adaptive Testing (CAT), as an innovative assessment format has been empirically proven to hold a number of attractive advantages such as reduction in test length, the promise of an increase in precision and security extensively utilised by developed countries for certification and licensing. While a global adoption is expected, feasibility studies are needed to determine its workability for particular testing programmes. This study is a simulation of CAT for Postgraduate Certification in Education (PGCE), a consecutive route to teacher professionalisation. Monte-Carlo simulation was employed as a powerful research method that enables researchers to project future outcomes in terms of assessment performance for use at the higher education level. Utilising CAT can be adopted for educational assessment to compliment online teaching and learning for ensuring quality crop of teachers produced for senior schools. This study gives a clear direction on moving the African continent to the second of four generations of computer-adaptive testing in aligning to current trends in educational assessment.

Key words: *Computer Adaptive Testing; Postgraduate Certification in Education Assessments, Monte-Carlo simulation, 3-Parameter Logistic Model*

Introduction

National development has been linked to educational development. However, academic development may not be realisable without the production of effective and efficient teachers saddled to impart knowledge to learners and guide them as they engage in learning activities. Teachers are products of specific teacher education policies and practices in all parts of the world. Teacher education is the groundwork of a quality educational system and, as such, the key to aspects of development. Teacher education equips teachers with the necessary skills and competencies. These skills enable them to develop and instil educational and societal norms in learners (Jubril, 2007). The training received is termed as Initial Teacher Education (ITE) programmes at all public Higher Education Institutions (HEIs) where enrolment is open in two existing ITE qualifications, the Bachelor of Education (BEd) and the Postgraduate Certificate in Education (PGCE). The BEd qualification is an integrated teacher training programme where academic studies and professional preparation coincide, known as the concurrent model. On the other hand, the PGCE is a parallel teacher training programmes academic studies, and professional practice follow one another, also known as the consecutive model (Kachelhoffer, 1995; Sederevičiūtė-Pačiauskienė & Vainorytė, 2015; Zuzovsky & Donitsa-Schmidt, 2017).

In whichever of the above teacher training models, the curriculum for teacher training usually consists of three parts: academic studies (usually a degree or diploma specialisation in at least two school-related subjects leading to the BEd degree), school practice which may vary from short practical periods in school to more extended periods of internship under the supervision of teacher training lecturers (also termed as induction) and professional preparation (usually comprised by the study of educational theory from the founding disciplines and training linked with teaching skills necessary to be efficient and effective as a teacher in school, under which the PGCE falls) (Kachelhoffer, 1995; Nakpodia & Urien, 2011; Ejima, 2012).

Multiple training programmes utilising either of the models is germane to solving the shortage of teachers reported by the 2011 Centre for Development and Enterprise (CDE) study which indicated that South Africa was producing only a third of the country's requirement of some 25,000 new teachers a year, and few in key subjects such as mathematics, science, commerce and technology (Simkins, 2015). Therefore, the PGCE programme is part of the efforts geared towards producing enough qualified, competent teachers for all school phases and subjects over the next ten years in South Africa. Studies reveal that teachers lack essential knowledge and skills to inadequate preservice teacher training, which is provided through ITE programmes at HEIs in South Africa. These findings strengthen the consensus shared by researchers and government that most South African teachers' subject content knowledge and pedagogical knowledge are insufficient and that this is a major cause of inadequate learner achievement (Hofmeyer, 2015). In response to increasing concern about the inadequate supply of new teachers and ITE programmes' quality, the Integrated Strategic Planning Framework for Teacher Education and

Development in South Africa 2011-2025, also known as ISPFTED also known as the Plan was birthed.

ISPFTED is a high-level strategy that frames the quantitative and qualitative challenges for teacher education and development. It outlines a 15-year plan for improved and expanded teacher education and development opportunities with the primary outcome as to improve the quality of teacher education and development in order to improve the quality of teachers and to teach through expanded enrolments in two existing ITE qualifications, the Bachelor of Education (BEd) and the Postgraduate Certificate in Education (PGCE). With only five more years left of the ISPFTED period, it is also necessary to ensure improved teachers' quality through their knowledge base for professional practice. The needed improvement can be achieved by using effective educational assessment systems for gauging the extent to which learning has been achieved, leading to certification.

Computer Adaptive Test (CAT) is a technology-driven process of delivering educational assessments with about five decades of practice. Petersen (2016) stressed that importance of technology in improving assessment using adaptive tests which adjust the difficulty of assessment items as students' progress through a testing session for more precise measurement. Empirical studies have proved a CAT; an adaptive form of assessment advantageous over the linear forms, while giving credence to its capacity to improve absorption and retention of information leading to increased student engagement, motivation, and ultimately learning (Linacre, 2000; Thompson, 2011; Petersen, 2016). While several assessment programmes are grounded in the paper-pencil forms, feasibility studies are required for transiting to CAT carried out through computer simulations. These simulations provide answers to questions such as the availability of psychometric experts, or using an external consultant, the capacity to develop extensive item banks, availability and affordability of a CAT delivery engine, having the required technology to develop a new one, the prospects of CAT with an expected reduction in test length as one of the key advantages of CAT and does the reduced in test length translate to enough saved examinee seat time and actual monetary savings. In a situation where CAT costs more and does not substantially decrease seat time, the promise of an increase in precision and security makes it worthwhile (Thompson & Weiss, 2011). They further stressed that empirical evidence is gathered through simulation research, which informs CAT specifications, and this is the very foundation for validity.

Computer simulation is a powerful research method that enables researchers to look at artificial world and project future outcomes to improve performance. As such, simulation can be seen as a laboratory, safe from the real environment's risks, for testing out hypotheses and making predictions as one of the purposes of simulations (Dooley, 2002). The author further explained that simulation could be approached as a discrete event where variables and events describe systems changes and establish cause and effects relationships and lastly as agent-based where the reactionary system and the environment are described. Based on these approaches, computer simulations are excellent for studying variables that provide the necessary mechanism for finalising several decisions concerning CAT design before implementing an operational testing program (Eignor, 1993). Some of these variables are starting criteria; item selection procedures, test length as impacting stopping rules, item exposure control, score estimation procedures, test administration mode, and the interactions among the variables in the design of a CAT. There are various methods for controlling CAT item exposure rates, including fade-away, randomesque, progressive restricted and probabilistic methods. Olea *et al.* (2012) stressed a positive correlation between item exposure control used and the item selection criteria and of course, the choice of a method would also depend on the goal of a testing programme. The progressively restricted method was considered efficient in preserving ability estimation precision and increasing the item bank's security.

There are three simulations: Monte Carlo, post-hoc and hybrid (Seo & Choi, 2018). Monte Carlo simulations are studies carried out under varying conditions for a large number of imaginary examinees. Through this process, large examinee and item data can be generated by computer software for dichotomous IRT Models such as CATSim (David J. Weiss), FireStar (Seung W. Choi), SimuMCAT (Lihua Yao) and WINGEN and SimulCAT (Kyung (Chris) T. Han) (International Association for Computer Adaptive Testing, n.d.), PARDSIM (M. Yoes) (Thompson & Weiss, 2011) and for Polytomous IRT Models such as SIMPOLYCAT, which is a SAS program (Chen & Cook, 2009). While some of the software barely randomly generate data for use based on provided specifications (WINGEN and PARDSIM), others are mainly for simulating CAT (FireStar), CATSim and SimulCAT which can both generate data and simulate CAT.

As an illustration, Thompson and Weiss (2011) explained that Monte Carlo method could be used simulate a bank of 300 and 500 items respectively and results compared to determine which presents a better goal for the organisation and can be done before a single item is written. This submission is essential considering that CAT involves laborious procedures which require verification before commencement. Monte Carlo simulations can be conducted using either the one (Rache), two or three-parameter logistic IRT models that estimate the exact probability of a correct response to an item for a given value of θ. Researchers can quickly generate a response to an item through this process, given its item parameters and examinee ability level denoted as θ. In an example given in Thompson & Weiss (2011),

Hypothetically, an average examinee (θ = 0.0) could be calculated to have a 0.75 probability of a correct response to an item. A random number generated from a uniform distribution (usually with a mean-0, Standard Deviation-1) having a value is 0.75 or less shows that the generated response is 'correct.' On the other hand, if the value is greater than 0.75, then the generated response is 'incorrect.' Given item and examinee parameters, an entire dataset of correct/incorrect responses could be generated with just a click of the mouse depending on the researcher's interest at different stages of the CAT development process. With this, *basing the generation of data on specified parameters makes the simulation more defensible. (P.3)*

Therefore, Monte Carlo simulations provide answers to salient questions at the planning stage. Post-hoc simulations, on the other hand, require real data for determining the specifications for the final CAT. At this stage of the CAT development process, a calibrated item bank is needed using an appropriate IRT model which yields an actual item bank developed and data collected from real examinees responding to developed items. While real data is preferred to randomly generated data for determining the prospects of CAT for future use, a significant drawback with the post-hoc method of simulation is item redundancy which leaves many items unused. This drawback can be circumvented by the third method of a simulation called hybrid. The hybrid simulation uses real data where available while missing responses are generated using Monte Carlo methods based on each examinee's θ as estimated from the items he/she has answered. This method allows CATs to be simulated more effectively with a real item bank and real examinees (Thompson & Weiss, 2011).

With ability estimation, CAT begins with an initial trait estimate for the examinee which can be fixed, based on either the mean population distribution or prior information. As such, item selection aims to maximise information at the initial θ estimate leading directly to the next administered item through which an examinee's ability estimate is updated. Items are selected going by this iterative process until a stopping rule is satisfied, leading to the final trait estimation based on all the examinees' responses. Designing a CAT requires a decision on initial, interim and final ability estimation methods (van der Linden & Pashley, 2000; Davey, 2011); approached using maximum likelihood (ML with or without fences) or Bayesian (the Maximum and Expected a Posteriori) methods (Butterfield, 2016). According to Weiss (2011), one significant advantage of maximum likelihood estimation with Bayesian methods is that it considers all the information in an examinee's responses in conjunction with all the information available on each test item.

Theoretical Framework

This study was premised on the 3-Parameter Logistic Model (3-PLM) Item Response Theory (IRT) for dichotomously scored responses. IRT explains an examinee's response to test items via a mathematical function based on their ability (Al-A'ali, 2006). The theory establishes the examinees' interaction level with the items in the test, based on the probability of correct response to an item (Magno, 2009). The 3-PLM IRT model adopted for this study considered the estimates of difficulty (b-parameter), discrimination (a-parameter), and pseudo guessing (c-parameter).

Statement of the Problem

One of the four strategic outputs of Integrated Strategic Planning Framework for Teacher Education and Development in South Africa 2011-2025 is addressing the individual and systemic needs of teacher development with an overriding goal of expanding and improving the quality of initial teacher education (ITE), as well as their quality and ability to teach new ITE curricula effectively. This point to the extent to which learning is being achieved: quality of teacher education output. Hofmeyer (2015) reported a low quality of ITE programmes. In response to this, the Department of Higher Education and Training (DHET) gave notice that by July 2014, that all teacher education programmes had to be re-designed with particular emphasis on what is taught (content knowledge), how it is taught (pedagogical content knowledge), and a strong practise teaching component. However, the assessment component for gauging the extent to which learning has been achieved is mostly absent. CAT as an innovation in assessment practice is deemed appropriate for filling this gap.

In designing CAT, it is necessary to study related variables such as ability estimation methods, test length, item exposure and item selection criteria to provide the answers to salient questions with CAT requiring adequate research and documentation through computer simulations (Eignor *et al.*, 1993; Thompson & Weiss, 2009). Simulation studies are highly recommended for evaluating CAT administration performance. Given the item pool and simulee distribution, using CAT simulation for determining measurement precision is an important aspect of CAT performance evaluation and gains its importance as a practical way to study and evaluate CAT programs and their implementation (Han, 2018). A study by Ogunjimi *et al.*, (2021) revealed that the fixed-length test guarantees a higher testing precision but with a higher item exposure rate which can be handled by falling back on the item selection methods that rely less on the a-parameter and lesser item redundancy compared to the variable-length test. Wang and Kolen (2001) and Gibbons, *et al.* (2008) examined test administration differences concerning CAT and Paper Pencil tests modes to ensure comparability. A study by Oladele *et al.* (2020) established the a-stratification with b-blocking item selection method as optimal for CAT. Using the a-stratification with b-blocking this item selection method deployed as a fixed-length test, the general objective of the study is to provide simulated evidence on the precision of ability estimation with CAT by investigating the initial, interim and final score estimation while varying options concerning estimation methods and item exposure control aimed at developing and implementing CAT for postgraduate certificate examinations in a South African university. To achieve this objective, the following research questions were raised:

1. What is the performance of ability estimation precision starting CAT using fixed values, randomly chosen values at +-0.5 and mean performance values using Bayes Expected a Posteriori (EAP) with or without progressively restricted item exposure controls?

2. What is the performance of interim ability estimation precision on CAT while limiting the range of estimation and estimates by jumps using Bayes Expected a Posteriori (EAP) with or without progressively restricted item exposure controls?
3. What is the performance of the final ability estimation precision on CAT using Maximum Likelihood or not using Maximum Likelihood using Bayes Expected a Posteriori (EAP) with or without progressively restricted item exposure controls?

Methodology

Design: This study adopts a one-shot case study design deployed as a simulated study to improve high stakes assessment practice.

Simulee Protocol: This study was premised on the simulation protocol for the study on method choices for ability estimation; however, focusing on testing performance. The study revealed non-significance among the methods for Maximum Likelihood and Bayesian ability estimation methods. A preference for Bayes Expected a Posteriori (EAP) was considered flexible with consistent response patterns irrespective of the availability an optimal item pool typical with early CAT programmes as determinant factors (Le, 2013; Oladele & Ndlovu, 2021).

Data Analysis: The simulation yielded conditional statistics (Bias, Mean Absolute Error and Root Mean Square Error) which were used to compare the performance of ability estimation for CAT with the aid of line graphs plotted using Excel (Han, 2018).

Ethical statement: This study was exempted from the requirement to obtain informed consent by the Faculty of Education Research Ethics Committee of the University of Johannesburg because the data for the study were computer-simulated.

Results

Research Questions One: The performance of ability estimation of CAT using the fixed, random and data values were analysed using conditional statistics data saved in the SimulCAT output file (*.sca), and presented with line graphs as shown in Figures 1a, 2a and 3a-CBAIS, 4a, 5a and 6a-MAE and 7a, 8a and 9a-RMSE respectively.

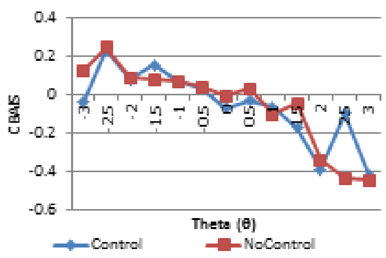

Figure 1a: CBAIS (Fixed-Initial Averaged θ range).

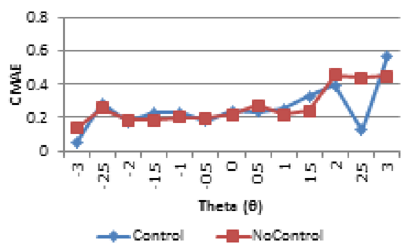

Figure 4a: CMAE (Fixed-Initial Averaged θ range).

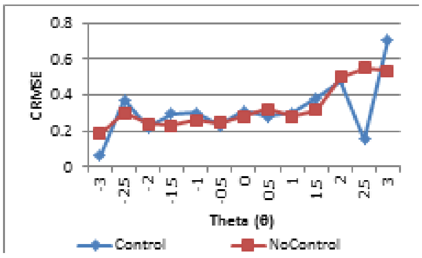

Figure 7b: CRMS (Fixed- Initial Averaged θ range).

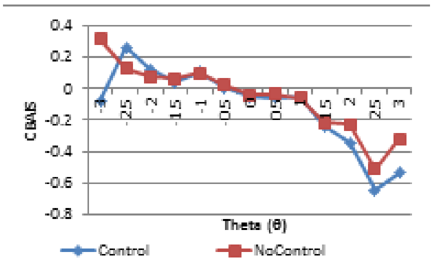

Figure 2b: CBAIS (Random- Initial Averaged θ range).

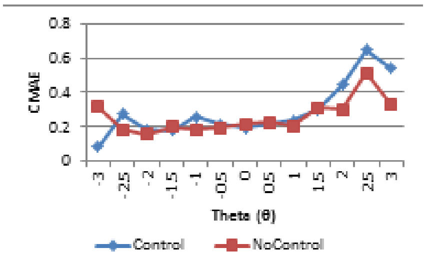

Figure 5b: CMAE (Random- Initial Averaged θ range).

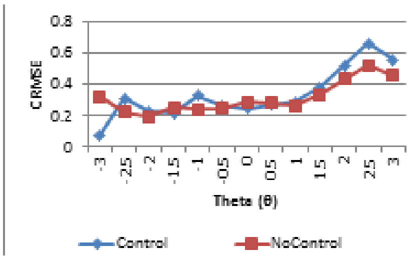

Figure 8b: CRMSE (Random- Initial Averaged θ range).

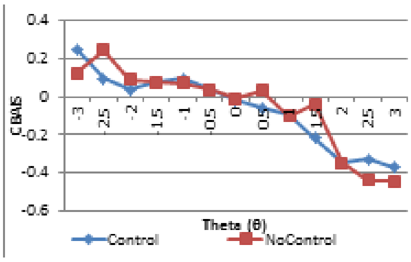

Figure 3b: CBAIS (Data- Initial Averaged θ range).

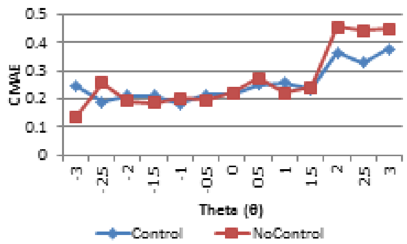

Figure 6b: CMAE (Data- Initial Averaged θ range).

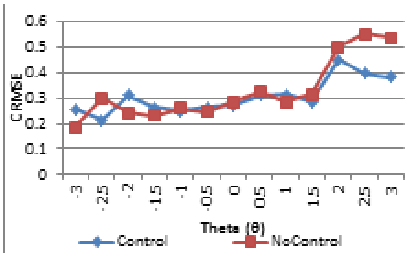

Figure 9b: CRMSE (Data- Initial Averaged θ range).

Research Questions Two: The performance of ability estimation at the interim of CAT with jumps and ranges was analysed using conditional statistics data saved in the SimulCAT output file (*.sca), and presented with line graphs as shown in Figures 1b, 2b and 3b-CBAIS, 4b, 5b and 6b-MAE and 7b, 8b and 9b-RMSE respectively.

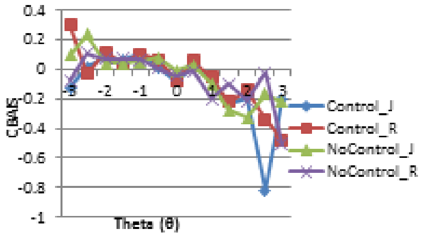

Figure 1b: CBAIS (Fixed-Interim Averaged θ range)

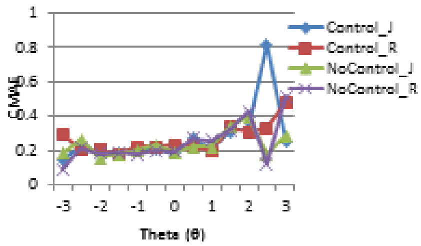

Figure 4b: CMAE (Fixed-Interim Averaged θ range).

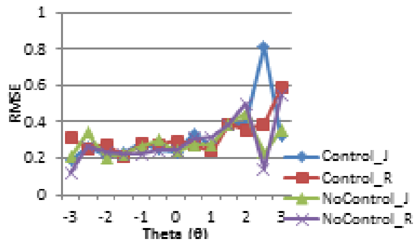

Figure 7b: CRMSE (Data-Interim Averaged θ range).

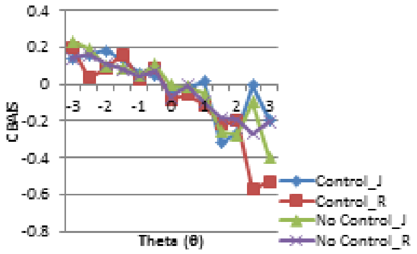

Figure 2b: CBAIS (Random-Interim Averaged θ range)

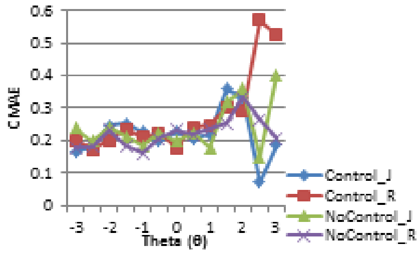

Figure 5b: CMAE (Random Averaged θ range).

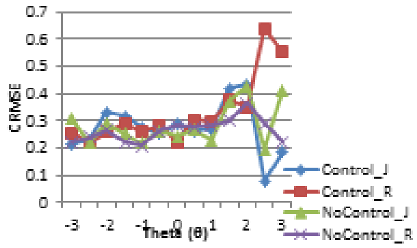

Figure 8b: CRMSE(Random-Interim Averaged θ range

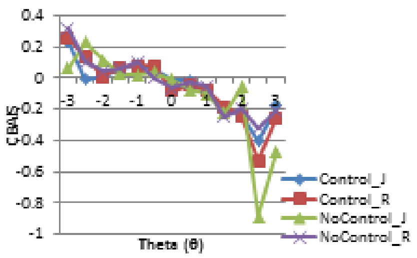

Figure 2b: CBAIS (Data-Interim Averaged θ range)

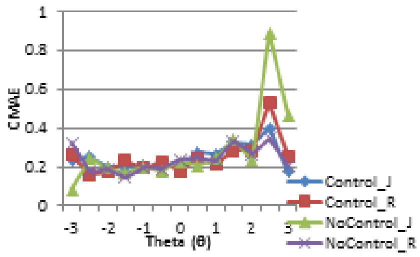

Figure 6b: CMAE (Interim Averaged θ range).

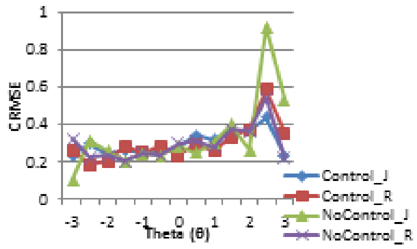

Figure 9b: CRMSE (Data-Interim Averaged θ range)

*J: Jump; R: Range

Research Questions Three: The performance of ability estimation at the final stage of CAT with/no Maximum Likelihood Estimation was analysed using conditional statistics data saved in the SimulCAT output file (*.sca), and presented with line graphs as shown in Figures 1c, 2c and 3c-CBAIS, 4c, 5c and 6c-MAE and 7c, 8c and 9c-RMSE respectively.

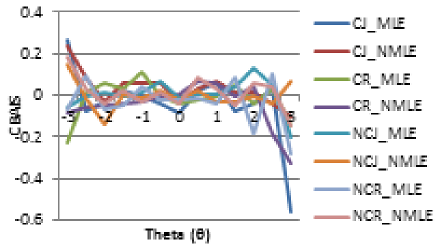

Figure 1c: CBAIS (Fixed-Final Averaged θ range)

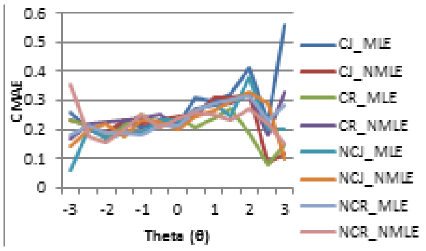

Figure 4c: CMAE (Fixed-Final Averaged θ range)

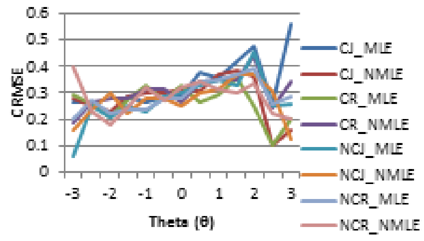

Figure 7c: CRMSE (Fixed-Final Averaged θ range)

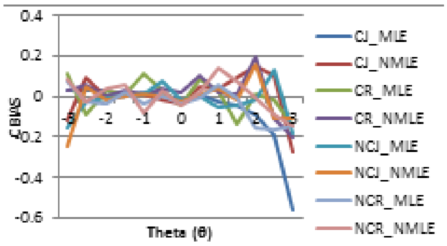

Figure 2c: CBAIS (Random-Final Averaged θ range)

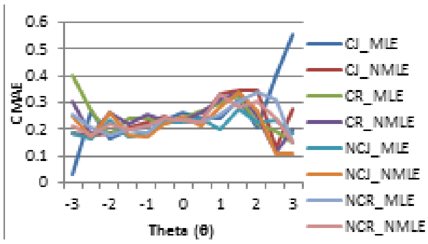

Figure 5c: CMAE (Random-Final Averaged θ range)

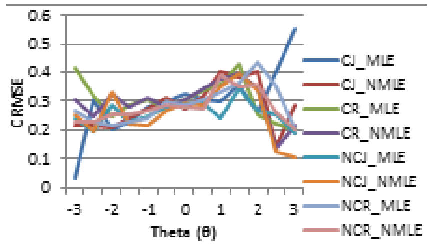

Figure 8c: CRMSE (Random-Final Averaged θ range)

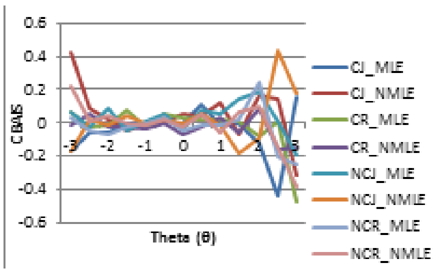

Figure 3c: CBAIS (Data-Final Averaged θ range)

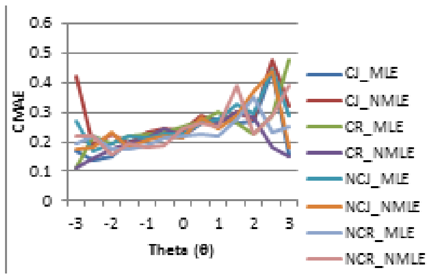

Figure 6c: CMAE (Data-Final Averaged θ range)

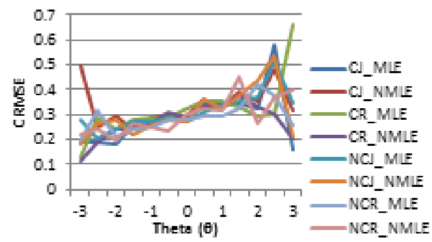

Figure 9c: CRMSE (Data-Final Averaged θ range)

*C: Control; NC: No Control; J: Jump; R: Range MLE: Maximum Likelihood Estimate; NMLE: No Maximum Likelihood Estimate

Discussion

The simulation results at the start of CAT reveal that the CBIAS (Figs. 1a; 2a and 3a respectively) indicated that the observed systematic error was almost zero as fixed better achieved with item exposure control, closer to -1 with random values between -0.5 and +0.5 and tightly around zero using data. The observed errors based on CMAE were also consistent across the theta areas of -2 to +2 short-ranged with fixed and random values (Fig. 4a and 5a respectively), and widest ranged with data (Fig. 6a). The CRMSE was tightly controlled to be lower than 0.3 across the theta areas of -2 to +2 tightly controlled around zero with maximum values of 0.4 as fixed and random (Figs. 7a and 8a respectively) and best at data with the maximum value of 0.3 (Fig. 9a). This finding deviates from that of Han (2018) which revealed that CMAE and CBAIS are consistent across all θ areas of -3 to +3 strictly using random values using the randomesque item exposure control items five best.

van der Linden and Pashley (2000) stressed that errors with an initial ability estimate are only acute for short tests, such as a 10-item tests battery. With more extended tests, of more than 20-30 items, ability estimator generally stands the chance of recovering despite a poor start. Furthermore, empirical initialisation of adaptive tests based on a fixed item results in item overexposure as the first items in the test is always chosen from the same subset in the pool. This finding further approves of using data over other methods. Veldkamp and Matteucci (2013) reiterate that collateral information such as scores resulting from earlier tests can be useful at the initial stage of the candidate's ability estimation during a CAT. Doing this ensures that only informative items at the candidate's ability level will be selected which lessens the problem of an item overexposure.

At the interim of CAT reveal, the simulation results revealed that the θ ranged between -2 to +2, similar to the observation starting CAT, with CBAIS best tightly controlled around zero while applying item exposure control with jumps limited to less than 1 for the first five items (Figs 1b; 2b and 3b) with similar outcomes with the CMAE (Figs 4b; 5b and 6b), and CRMSE (Figs. 7b, 8b and 9b). The inconsistency recorded at the extreme θ level is in line with that of Han (2018), which also limited θ estimate jump to less than 1 for the first five items. The author proffered relaxing the constraint for the θ estimate jump as a solution to this.

At the final stage of CAT ability estimation, the simulation results CBAIS (Figs 1c; 2c and 3c) with similar outcomes with the CMAE (Figs 4c; 5c and 6c), and CRMSE (Figs. 7c, 8c and 9c); similar to the findings at initial and interim stages also reveal that the θ ranged between -2 to +2, deviating from the initial scale of -3 to +3 (Han, 2016). The maximum likelihood estimator with the EAP method results in a shrinkage in the theta scale resulting in an optimal condition where each item matches the examinee trait-level parameter (Segall, 2005). Weiss (2011) further stressed that the maximum likelihood procedure aids the modification, necessary when several items have been administered. They have all been answered either correctly or incorrectly) and temporarily assume that theta for a group of examinees is normally distributed with Bayesian methods. Seo and Choi (2018) also stressed that the Bayesian methods might become more biased as θ approaches the extremes due to regression toward the mean of the prior. According to Weiss and McBride (1984), fixed prior performed better with variable length test as to when it was used with a fixed test length, severe bias was observed using Bayesian methods. In contrast, data priori performed better with fixed length test. However, he stresses the need for an accurate apriori for adaptive tests to yield unbiased thetas estimates. Furthermore, Veldkamp and Matteucci (2013) stressed that CAT converges faster, leading to shorter test being a primary advantage of CAT while reducing cost germane for cost analysis. This submission strengthens Bayesian methods using existing data on examinees with the EAP method of score estimation for CAT. However, Caution should be taken in ensuring accurate prior as a condition for obtaining unbiased thetas estimates with Bayesian methods (Weiss & McBride, 1984; Segall, 2005).

Conclusion

Based on this study's findings, the conclusion is made for using data as a priori for CAT while applying jump at the interim of CAT and using maximum likelihood estimation for maximum CAT performance. Item exposure should also be tightly controlled. Background information of examinees as impacting the scoring of candidates' responses is also germane to CAT performance. CAT as an assessment innovation is gaining more relevance in the 21st century where technology is progressively tailored for solving educational problems like in other fields. Adaptively test administration would enhance authentic assessments in teacher education whose data capture true learning, as well as cater for the lacking assessment component of the Integrated Strategic Planning Framework for Teacher Education and Development in South Africa.

Acknowledgment

The author give due acknowledgement to Dr. Kyung (Chris) T. Han of the Graduate Management Admission Council who authored SimulCAT, a free software package used to generate the data for this study.

REFERENCES

Butterfield, M. S. (2016). *Comparing item selection methods in computerised adaptive testing using the rating scale model.* Doctoral dissertation: the University of Texas at Austin. https://bit.ly/3pMhPTq

Chen, S. K., & Cook, K. F. (2009). SIMPOLYCAT: An SAS program for conducting CAT simulation based on polytomous IRT models. Behaviour research methods, 41(2), 499-506.

Davey, T. (2011). A Guide to Computer Adaptive Testing Systems. *Council of Chief State School Officers*. https://files.eric.ed.gov/fulltext/ED543317.pdf

Dooley, K. (2002), "Simulation research methods," Companion to Organisations, Joel Baum (ed.), London: Blackwell, p. 829-848. https://www.researchgate.net/publication/228986412_Simulation_research_methods

Eignor, D. R., Stocking, M. L., Way, W. D., & Steffen, M. (1993). Case Studies in Computer Adaptive Test Design Through Simulation 1, 2. ETS Research Report Series, 2, i-41.

Ejima, O. S. (2012). Teacher professional development in the 21st Century, Nigeria. *An international journal by Global educators for all initiative,* 1-6. http://globaleducators.org/downloads/global_voice_of_educators/Journal%202012%20-%2013.pdf

Fulljames, T. (2009, July 9-10). A career in tertiary management – a transformative experience! [Paper presentation]. ATEM Aotearoa Conference, Christchurch, New Zealand. https://bit.ly/2Gbd0hG

Gibbons, R. D., Weiss, D. J., Kupfer, D. J., Frank, E., Fagiolini, A., Grochocinski, V. J., Bhaumik, D., K., Stover, A., Bock, R. D., Immekus, J. C. (2008). Using computerised adaptive testing to reduce the burden of mental health assessment. *Psychiatric Services, 59*(4), 49-58.

Han, K. C. T. (2018). Conducting simulation studies for computerised adaptive testing using SimulCAT: An instructional piece. *Journal of Educational Evaluation for Health Professions, 15*(20). https://doi.org/10.3352/jeehp.2018.15.7

Han, K. T. (2016). Maximum likelihood score estimation method with fences for short-length tests and computerised adaptive tests. *Applied psychological measurement, 40*(4), 289-301. https://doi.org/10.1177/0146621616631317

Hofmeyer, J. (2015). Teachers in South Africa: Supply and Demand, 2013-2025. Centre for Development and Enterprise, CDE Technical Report.

International Association for Computer Adaptive Testing (n.d.) CAT Software https://www.iacat.org/content/cat-software

Jibril, M. O. (2007). Teacher Education in Nigeria: An Overview. African Research Review, 1(2), 130-140.

Kachelhoffer, P. M. (1995). Teacher training models for a new South Africa. Teacher Education Quarterly, 49-57. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1048.807&rep=rep1&type=pdf

Le, D. T. (2013). Applying item response theory modeling in educational research. Published Graduate Theses and Dissertations, Lowa State University Digital Repository. https://lib.dr.iastate.edu/cgi/viewcontent.cgi?article=4417&context=etd

Linacre, J. M. (2000). Computer-adaptive testing: A methodology whose time has come (No. 69, p. 58). MESA memorandum. https://www.cehd.umn.edu/EdPsych/C-Bas-R/Docs/Linacre2000_CAT.pdf

Nakpodia, E. D., & Urien, J. (2011). Teacher education in Nigeria: Challenges to educational administrators in the 21st century. The Social Sciences, 6(5), 350-356.

Ogunjimi, M.O., Ayanwale, M.A., Oladele, J. I., Daramola, D. S., Jimoh, M. I., & Owolabi, H.O. (2021). Simulated Evidence of Computer Adaptive Test Length: Implications for High-stake Assessment in Nigeria. *Journal of Higher Education Theory and Practice* (ISSN# 2158-3595). https://nabpress.com/jhetp-indexing-and-impact

Oladele, J. I., Ayanwale, M. A. & Owolabi, H. O. (2020). Paradigm Shifts in Computer Adaptive Testing in Nigeria in Terms of Simulated Evidences. *Journal of Social Sciences, 63*(1-3): 9-20. https://doi.org/10.31901/24566756.2020/63.1-3.2264

Oladele, J. I., & Ndlovu, M. (2021). A Review of Standardised Assessment Development Procedure and Algorithms for Computer Adaptive Testing: Applications and Relevance for Fourth Industrial Revolution *International Journal of Learning, Teaching and Educational Research,* 20, 5, 1-17. https://doi.org/10.26803/ijlter.20.5.1

Olea, J., Barrada, J. R., Abad, F. J., Ponsoda, V., & Cuevas, L. (2012). Computerised adaptive testing: The capitalisation on the chance problem. The Spanish journal of psychology, 15(1), 424. https://repositorio.uam.es/handle/10486/665572

Petersen, J. (2016). Innovative assessment practices. https://freshgrade.com/wp-content/uploads/2020/07/FG-Innovative-Assessment-Whitepaper-1.pdf

Privitera, G. J., & Delzell, L. A. (2019). Quasi-Experimental and Single-Case Experimental Designs. Research methods for education, 333-370. https://us.sagepub.com/sites/default/files/upm-binaries/89876_Chapter_13_Quasi_Experimental_and_Single_Case_Designs.pdf

Sederevičiūtė-Pačiauskienė, Ž., & Vainorytė, B. (2015). The concurrent and consecutive models of initial teacher training: problematics and tendencies. Rural Environment. Education. Personality, 347-354. https://llufb.llu.lv/conference/REEP/2015/Latvia-Univ-Agricult-REEP-2015proceedings-347-354.pdf

Segall, D. O. (2005). Computerised adaptive testing. *Encyclopedia of social measurement, 1,* 429-438. http://www.iacat.org/sites/default/files/biblio/se04-01.pdf

Seo, D. G., & Choi, J. (2018). A 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

Simkins, C. (2015). Technical Report: Teacher Supply and Demand in South Africa: 2013 to 2025. Centre for Development and Enterprise. https://www.cde.org.za/wp-content/uploads/2018/07/Teacher-Supply-and-Demand-2013-2025-Technical-Report-March2015-CDE.pdf

Thompson, N. A. & Weiss, D. J. (2009). Computerised and adaptive testing in educational assessment. In: F Scheuermann, J Björnsson (Eds.): The Transition to Computer-Based Assessment. Italy: Office for Official Publications of the European Communities.

Thompson, N. A. (2011). Advantages of Computerized Adaptive Testing (CAT). Assessment Systems (White Paper). https://assess.com/docs/Advantages-of-CAT-Testing.pdf

Thompson, N. A., & Weiss, D. A. (2011). A framework for the development of computerised adaptive tests. *Practical Assessment, Research, and Evaluation, 16*(1), 1.

University of Maimi (2020). MANOVA/MANCOVA using SPSS. Statistical Supporting Unit. https://sites.education.miami.edu/statsu/wp-content/uploads/sites/4/2020/07/MANOVAMANCOVA-Webinar.pdf

van der Linden, W. J. & Pashley, P. J. (2000). Item selection and ability estimation in adaptive testing. In W. J. Linden, W. J. van der Linden, & C. A. Glas (Eds.). Computerised adaptive testing: Theory and practice. Springer Science & Business Media. https://doi.org/10.1007/0-306-47531-6_1

van der Linden, W. J. & Pashley, P. J. (2009). Item Selection and Ability Estimation in Adaptive Testing. W.J. van der Linden and C.A.W. Glas (eds.), *In Elements of Adaptive Testing, (pp. 3-30). Springer, New York, NY.* https://doi.org/10.1007/978-0-387-85461-8_1

Veldkamp, B. P., & Matteucci, M. (2013). Bayesian computerised adaptive testing. Ensaio: Avaliação e Políticas Públicas em Educação, 21(78), 57-82. https://doi.org/10.1590/S0104-40362013005000001

Wang, T. & Kolen, M. J. (2001). Evaluating Comparability in Computerized Adaptive Testing: Issues, Criteria and an Example. *Journal of Educational Measurement, 1*(1), 19-49.

Weiss, D. J. (2011). Better data from better measurements using computerised adaptive testing. Journal of Methods and Measurement in the Social Sciences, 2(1), 1-27. https://journals.uair.arizona.edu/index.php/jmmss/article/viewFile/12351/11641

Weiss, D. J., & McBride, J. R. (1984). Bias and information of Bayesian adaptive testing. Applied Psychological Measurement, 8(3), 273-285. https://doi.org/10.1177/014662168400800303

Zuzovsky, R., & Donitsa-Schmidt, S. (2017). Comparing the effectiveness of two models of initial teacher education programmes in Israel: concurrent vs consecutive. European Journal of Teacher Education, 40(3), 413-431. https://doi.org/10.1080/02619768.2017.1318377