International Journal of Evaluation and Research in Education (IJERE)
Vol. 11, No. 3, September 2022, pp. 1392~1399
ISSN: 2252-8822, DOI: 10.11591/ijere.v11i3.21986

Simulated computer adaptive testing method choices for ability estimation with empirical evidence

Jumoke I. Oladele, Mdutshekelwa Ndlovu, Erica D. Spangenberg
Department of Science and Technology Education, Faculty of Education, University of Johannesburg, Johannesburg, South Africa

Article Info

Article history:

Received Apr 6, 2021
Revised May 28, 2022
Accepted Jun 17, 2022

Keywords:

Ability estimation
CAT
Item response theory
Methods
Simulation

ABSTRACT

Computer adaptive testing (CAT) is a technological advancement for educational assessments that requires thorough feasibility studies through computer simulations to ensure strong testing foundations. This advancement is especially germane in Africa being adopters of technology, and this should not be done blindly without empirical evidence. A quasi-experimental design was adopted for this study to establish methodological choices for CAT ability estimation. Five thousand candidates were simulated with 100 items simulate through the three-parameter logistic model. The simulation design stipulated a fixed-length test of 30 items, while examinee characteristics were drawn from a normal distribution with a mean of 0 and a standard deviation of 1. Also, controls for the simulation were set not to control item exposure or to use the progressive restricted method. Data gathered were analyzed using descriptive statistics (mean and standard deviation) and inferential statistics (Two-way multivariate analysis of variance: MANOVA) for testing the generated hypotheses. This study provided empirical evidence for choosing ability estimation methods for CAT as part of the efforts geared towards designing accurate testing programs for use in higher education.

Corresponding Author:

Jumoke I. Oladele
Department of Science and Technology Education, Faculty of Education, University of Johannesburg
Kingsway and University Road, Aukland Park 2006, Johannesburg, South Africa
Email: jumokeo@uj.ac.za

1. INTRODUCTION

The origin of Computer adaptive testing (CAT) can be traced to a French psychologist Alfred Binet starting in the early 19th century, who was strictly interested in accurate ability measurements and the basis for which CAT is built [1], [2]. CAT has been widely employed in the developed world as technological advancement for educational assessments with over two decades of applications [3]. Being in the second generation of computer-based testing, the world of assessment focuses on Artificial Intelligence at the fourth industrial revolution [4]. A clarion call is on for the African continent still predominantly in the first generation of utilizing the fixed-form of computer-based assessments to make a forward march to the foremost generation while canvassing for a start with the adaptive forms of testing [5]. While this call for a jolt forward is well overdue, care must be taken to ensure that moving to the adaptive forms is well-founded and carefully researched. Simulations are carried out for feasibility studies to ensure that a CAT testing program is built on empirically proven foundations. The legitimacy of CAT further gains importance for the sub-Saharan African continent as late technology adopters [6], [7], with the majority of the region lagging in the bottom half of the networked readiness index rankings; a measure of the propensity for countries to exploit and benefit from the opportunities offered by information and communications technology [8].

CAT is premised on item response theory (IRT), which is classified as either dichotomous or polytomous models based on how responses are scored [9], [10]. Extensive CAT research has been carried out in the military, health, and education sectors for ability measurements [2], [3], [11]–[13]. Research reveals that adaptive form tests have psychometric properties equal to or greater than fixed forms while reducing test lengths by up to 50% [13]–[15]. Furthermore, adaptive tests hold superior statistical properties than traditional tests [16]. Research has been done on various statistical methods for CAT ability estimation, being the essential aspect of CAT performance evaluation of a testing program, premised on the model fit and response patterns adopted [13], [17], compared the Bayesian modal (BM) estimator with Jeffreys' prior distribution and the weighted likelihood (WL) estimator under the three-parameter logistic model. Ability estimation can be approached using maximum likelihood estimation with or without fences (MLEF or MLE) or the Bayesian maximum and expected posteriori (MAP and EAP) method. Previous researches [3], [18], [19] observed that when Bayes theorem (1763) is applied, the conditional probability of the item and person parameters given the data can be modeled as a combination of prior beliefs about them and a parametric model about what the data should look like, conditional on the item and person parameter values.

MLE which relies heavily on the quality of the items is gauged by its parameters and is commonly used [1], explained that the ML estimation method does not produce finite estimates for response patterns with all items correct or all incorrect. This ability estimation method does not work with the specific dichotomous response [13], which constitutes challenges at the early stages of CAT administration with short test lengths. A similar school of thought indicating that maximum likelihood methods treat person abilities as fixed effects which results into undesirable skewness which can be circumvented using bias correction methods [10]. This challenge has been taken care of using MLEF, whereby lower and upper bounds of theta estimation are set while truncating the score estimation to be one of those bounds when the log-likelihood function fails to yield a peak with the dichotomous response pattern. Alternatively, Bayesian procedures strengthen ability estimation using prior information on the target population's distribution while reducing the item parameters' estimation errors with small sample sizes, especially for the discrimination parameter [18], [20]. CAT without adequate feasibility studies through simulation research in each stage of the development process runs the risk of inefficiency, rendering its advantages worthless and legally indefensible [15].

Research has been carried out on the performance of CAT regarding test forms [12], item selection procedures [5], [21] and methodological choices [9], [22] and test performance with Bayesian methods [23]. The performance of these ability estimation methods can be researched through simulations for CAT feasibility studies using computer software such as SimulCAT [24], CATSIM [25], FireStar [26] or SimulMCAT [27]. While these ability estimation methods are readily available using computer software, choices made should be empirically proven. Establishing empirical evidence is especially necessary for technology adopters; countries who accept, integrate, and use new technology in society; a category in which African researchers belong. This study established the precision of various methodological choices for ability estimation achieved through three research questions: i) What is the precision of ability estimation of CAT using fixed values, randomly chosen values at ±0.5 and mean performance values using MLEF, MAP and EAP with or without progressively restricted item exposure controls?; ii) What is the precision of interim ability estimation of CAT while limiting the range of estimation and estimates by jumps using MLEF, MAP and EAP with or without progressively restricted item exposure controls?; iii) What is the precision of the final ability estimation of CAT using MLEF, MAP and EAP with or without progressively restricted item exposure controls? Hence, there are three hypotheses of this research: i) There is no significant effect of ability estimation precision starting CAT using fixed mean performance values, randomly chosen values at ±0.5 using MLEF, MAP and EAP with or without progressively restricted item exposure controls (H_{01}); ii) There is no significant effect of interim ability estimation precision of CAT limiting the range of estimation and estimates by jumps using MLEF, MAP and EAP with or without progressively restricted item exposure controls (H_{02}); iii) There is no significant effect of the final ability estimation precision of CAT using MLE or MLEF, MAP and EAP with or without progressively restricted item exposure controls (H_{03}).

2. RESEARCH METHOD

2.1. Design

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. A quasi-experimental design was adopted for this study. Using a factorial design of 3x3x2x2, the ability estimation methods of MLEF, Bayesian MAP and EAP were contained in the first factorial level. The second factorial level was used to start the CAT at three levels (fixed values, randomly chosen values at ±0.5 and mean performance values). The third factorial level was used at the interim of CAT occurring at two levels (limiting the range of estimation and estimates by jumps). The fourth factorial level was used at the final stage of CAT occurring at two levels (using maximum likelihood and not using maximum likelihood). Across all these levels, Item exposure controls had two levels of applying control and not applying control using the progressive restricted method across all these levels. The simulation used the three ability estimation methods as a treatment in the quasi-experiment with this design while varying item exposure as control. The experimental design is shown in Table 1.

Table 1. The experimental design

Groups	Treatment	Moderating variable	Post-test
Experimental Group I (EG-I)	X_1	Start: fixed values, randomly chosen values at ±0.5 and using mean performance values Interim: limiting the range of estimation and estimates by jumps Final: Maximum Likelihood/not with Maximum Likelihood	O_1
Experimental Group II (EG-II)	X_2	Start: fixed values, randomly chosen values at ±0.5 and using mean performance values Interim: limiting the range of estimation and estimates by jumps Final: Maximum Likelihood/not with Maximum Likelihood	O_1
Experimental Group III (EG-III)	X_3	Start: fixed values, randomly chosen values at ±0.5 and using mean performance values Interim: limiting the range of estimation and estimates by jumps Final: Maximum Likelihood/not with Maximum Likelihood	O_1
Control		Item exposure (With or without applying the progressively restricted)	O_1

X_1=Treatment for EG-I: MLEF; X_2=Treatment for EG-II: MAP; X_3=Treatment for EG-III: EAP; Control=Item exposure; O_1=Simulation

2.2. Simulation protocol

The Monte Carlo simulation method was used to generate data for this study using SimulCAT. SimulCAT is deemed appropriate for being a specialized, Monte-Carlo based simulation software [24]. The a-Stratification with b-Blocking item selection criteria method was used [5] with a fixed-length test of 30 items [28], and the choice of the progressive restricted item exposure method [20] was used for all simulations with 500 simulees. An item pool of 100 dichotomously scored items was created using 3-parameter logistic (3PL) item response with item discrimination (a), the difficulty (b), and the guessing (c) drawn from a normal distribution with a mean of 0 and standard deviation of 1. The descriptive statistics for the item parameter estimate for a pool of 100 items used for the simulated CAT are shown in Table 2.

As shown in Table 2, the mean of a, b, and c parameters of 0.44, -2.34 and 0.00, respectively, for the fixed-length simulated computer-adaptive test show that the generated data fell within the specified ranges to guarantee adequate discrimination between the low and high-ability students, moderate difficulty and pseudo guessing required for maximal functioning of CAT [5]. The researchers simulated the ability estimation for CAT using three Maximum Likelihood Estimation methods with fences, Bayesian maximum a posteriori and Bayes expected posteriori. The simulation design stipulated a fixed-length test of 30 items specified for 500 simulates “taking” the adaptive test at time slot 1. Also, controls for the simulation were set not to control item exposure or to use the progressive restricted method.

Data gathered were analyzed in two stages. In the first stage, descriptive statistics (mean and standard deviation) and inferential statistics (Two-way multivariate analysis of variance: MANOVA) were used to test the generated hypotheses to establish the precision of methodological choices for CAT ability estimation. MANOVA was deemed appropriate for this study with three dependent variables (MLEF, MAP, and EAP), having two independents (simulations across fixed, random and data initial score values and item exposure control/no control) [29], [30].

Table 2. Descriptive statistics for item pool

Parameters	Mean	Std. Deviation
a	.4428	1.26245
b	-2.3383	2.71670
c	.0000	.00000

3. RESULTS

3.1. Hypothesis 1

There is no significant effect of ability estimation precision starting CAT using fixed values, randomly chosen values at ±0.5 and using mean performance values using MLEF, MAP, and EAP with no and progressively restricted item exposure controls. To test Hypothesis 1, the ability estimation precision using MLEF, MAP, and EAP while controlling for item exposure (varying between the use of no control and progressive restrictions) with fixed, random and data initial score estimation methods premised on the conditional BIAS (CBIAS), conditional maximum mean absolute error (CMAE) and conditional root mean square error (CRMSE) SimulCAT outputs were analyzed using two-way MANOVA at 0.05 level of significance. The multivariate tests reported using Wilks' Lambda are displayed, as shown in Table 3.

Table 3 shows that the calculated values of F (6, CBIAS=.381; CMAE=.246; CRMSE=.119) tested at 0.05 alpha level. The first null hypothesis is accepted since all the p-values are greater 0.05 alpha level (.89; .96; .99 are >.05). This result connotes that premised on the conditional statistics, the initial score estimation methods (fixed, random and data) with or without applying progressively restricted item exposure controls for CAT have no significant effect on the ability estimation precision of starting CAT by using MLEF, MAP or EAP.

Table 3. Multivariate tests on initial ability estimation precision using MLEF, MAP and EAP

	Effect	Value	F	df	Error df	Sig.
CBIAS	Intercept	.908	2.355^b	3.000	70.000	.079
	SIM	.895	1.337^b	6.000	140.000	.245
	Exposure	.987	$.302^b$	3.000	70.000	.824
	SIM * Exposure	.968	$.381^b$	6.000	140.000	.890
CMAE	Intercept	.050	443.866^b	3.000	70.000	.000
	SIM	.952	$.579^b$	6.000	140.000	.746
	Exposure	.973	$.636^b$	3.000	70.000	.594
	SIM * Exposure	.979	$.246^b$	6.000	140.000	.960
CRMSE	Intercept	.043	517.991^b	3.000	70.000	.000
	SIM	.965	$.418^b$	6.000	140.000	.866
	Exposure	.967	$.801^b$	3.000	70.000	.498
	SIM * Exposure	.990	$.119^b$	6.000	140.000	.994

a. Design: Intercept + SIM + Exposure + SIM * Exposure; b. Exact statistic

3.2. Hypothesis 2

There is no significant effect of interim ability estimation precision of CAT while limiting the range of estimation and estimates by jumps using MLEF, MAP and EAP with or without progressively restricted item exposure controls. To test Hypothesis 2, ability estimation precision using MLEF, MAP and EAP while controlling for item exposure (varying between the use of no control and progressive restrictions) while limiting the range of estimation and estimates by jumps at the interim of CAT were analyzed using two-way MANOVA at 0.05 level of significance. The multivariate tests are reported using Wilks' Lambda, as shown in Table 4.

Tables 4. Multivariate tests on interim ability estimation precision using MLEF, MAP and EAP

Method	Effect	Value	F	df	Error df	Sig.
CBIAS	Intercept	.892	5.720^b	3.000	142.000	.001
	SIM	.959	1.005^b	6.000	284.000	.422
	Method	.995	$.226^b$	3.000	142.000	.878
	Control	.992	$.383^b$	3.000	142.000	.765
	SIM * Methods	.970	$.718^b$	6.000	284.000	.635
	SIM * Control	.956	1.068^b	6.000	284.000	.382
	Methods * Control	.995	$.238^b$	3.000	142.000	.869
	SIM * Methods * Control	.982	$.442^b$	6.000	284.000	.850
CMAE	Intercept	.063	706.642^b	3.000	142.000	.000
	SIM	.989	$.262^b$	6.000	284.000	.954
	Method	.994	$.277^b$	3.000	142.000	.842
	Control	.994	$.270^b$	3.000	142.000	.847
	SIM * Methods	.974	$.639^b$	6.000	284.000	.699
	SIM * Control	.976	$.567^b$	6.000	284.000	.757
	Methods * Control	.982	$.863^b$	3.000	142.000	.462
	SIM * Methods * Control	.987	$.307^b$	6.000	284.000	.933
CRMSE	Intercept	.050	901.566^b	3.000	142.000	.000
	SIM	.988	$.288^b$	6.000	284.000	.942
	Method	.994	$.298^b$	3.000	142.000	.827
	Control	.994	$.274^b$	3.000	142.000	.844
	SIM * Methods	.973	$.657^b$	6.000	284.000	.684
	SIM * Control	.977	$.566^b$	6.000	284.000	.758
	Methods * Control	.981	$.932^b$	3.000	142.000	.427
	SIM * Methods * Control	.987	$.321^b$	6.000	284.000	.926

a. Design: Intercept + SIM + METHODS + CONTROL + SIM * METHODS + SIM * CONTROL + METHODS * CONTROL + SIM * METHODS * CONTROL
b. Exact statistic

Table 4 shows calculated value of F(6, CBIAS: .442; CMAE: .307; CRMSE: .321) tested at 0.05 alpha level. The null hypothesis one is accepted since the P-values are greater 0.05 alpha level (.85; .93; .93 all >.05). This result connotes that premised on the conditional statistics, the initial score estimation methods (fixed, random and data) while applying no or using Progressively restricted item exposure controls for CAT has no significant effect on ability estimation precision at the interim of CAT while applying jumps and ranges using MLEF, MAP and EAP methods.

3.3. Hypothesis 3

There is no significant effect of the final ability estimation precision of CAT using MLE or MLEF, MAP and EAP with or without progressively restricted item exposure controls. To test Hypothesis 3, ability estimation precision using MLEF, MAP and EAP while controlling for item exposure (varying between the use of no control and progressive restrictions) using maximum likelihood or not using maximum likelihood estimation at the final stage of CAT were analyzed using two-way MANOVA at 0.05 level of significance. The multivariate tests are reported using Wilks' Lambda, as shown in Table 5.

Table 5. Multivariate tests on final ability estimation precision using MLEF, MAP, and EAP

	Effect	Value	F	df	Error df	Sig.
CBIAS	Intercept	.872	13.984	3.000	286.000	.000
	Start (SIM)	.982	.885	6.000	572.000	.505
	Interim (Methods)	.992	.812	3.000	286.000	.488
	Final	.983	1.659	3.000	286.000	.176
	Control	.990	.964	3.000	286.000	.410
	Start * Interim	.986	.692	6.000	572.000	.656
	Start * Final	.979	1.002	6.000	572.000	.423
	Start * Control	.987	.639	6.000	572.000	.699
	Interim * Final	.987	.639	6.000	572.000	.699
	Interim * Control	.998	.165	3.000	286.000	.920
	Final * Control	.996	.405	3.000	286.000	.749
	Start * Interim * Final	.997	.160	6.000	572.000	.987
	Start * Interim * Control	.990	.491	6.000	572.000	.815
	Start * Final * Control	.996	.209	6.000	572.000	.974
	Interim * Final * Control	.983	1.631	3.000	286.000	.182
	Start * Interim * Final* Control	.994	.279	6.000	572.000	.947
CMAE	Intercept	.061	1473.679	3.000	286.000	.000
	Start (SIM)	.990	.480	6.000	572.000	.823
	Interim (Methods)	.994	.549	3.000	286.000	.649
	Final	1.000	.019	3.000	286.000	.996
	Control	.977	2.210	3.000	286.000	.087
	Start * Interim	.988	.590	6.000	572.000	.739
	Start * Final	.986	.652	6.000	572.000	.689
	Start * Control	.985	.716	6.000	572.000	.637
	Interim * Final	.998	.188	3.000	286.000	.904
	Interim * Control	.991	.863	3.000	286.000	.461
	Final * Control	.999	.050	3.000	286.000	.985
	Start * Interim * Final	.975	1.196	6.000	572.000	.307
	Start * Interim * Control	.995	.222	6.000	572.000	.970
	Start * Final * Control	.997	.123	6.000	572.000	.994
	Interim * Final * Control	.991	.822	3.000	286.000	.483
	Start * Interim * Final* Control	.966	1.687	6.000	572.000	.122
CRM	Intercept	.052	1736.550	3.000	286.000	.000
	Start (SIM)	.985	.703	6.000	572.000	.647
	Interim (Methods)	.995	.490	3.000	286.000	.689
	Final	.999	.124	3.000	286.000	.946
	Control	.973	2.641	3.000	286.000	.050
	Start * Interim	.984	.790	6.000	572.000	.578
	Start * Final	.984	.763	6.000	572.000	.599
	Start * Control	.479	6.000	572.000	.824	.479
	Interim * Final	.999	.106	3.000	286.000	.957
	Interim * Control	.993	.635	3.000	286.000	.593
	Final * Control	1.000	.039	3.000	286.000	.990
	Start * Interim * Final	.973	1.291	6.000	572.000	.259
	Start * Interim * Control	.995	.244	6.000	572.000	.962
	Start * Final * Control	.997	.163	6.000	572.000	.986
	Interim * Final * Control	.991	.832	3.000	286.000	.477
	Start * Interim * Final* Control	.979	1.034	6.000	572.000	.402

a. Design intercept + START + INTERIM + FINAL + CONTROL + START * INTERIM + START * FINAL + START * CONTROL + INTERIM * FINAL + INTERIM * CONTROL + FINAL * CONTROL + START * INTERIM * FINAL + START * INTERIM * CONTROL + START * FINAL * CONTROL + INTERIM * FINAL * CONTROL + START * INTERIM * FINAL * CONTROL

b. Exact statistic

Table 5 shows calculated value of F(6, CBIAS:.279; CMAE: 1.69; CRMSE: 1.034)tested at 0.05 alpha level. The null hypothesis one is accepted since the P-values are greater 0.05 alpha level (.95; .12; .40 all >.05). This result connotes that premised on the conditional statistics, the initial score estimation methods (fixed, random and data) while applying no or using Progressively restricted item exposure controls for CAT has no significant effect on ability estimation precision at the interim of CAT while applying jumps and ranges with maximum likelihood or not using maximum likelihood estimation at the final stage of CAT using MLEF, MAP and EAP methods.

4. DISCUSSION

The simulated study shows no significant effects using MLEF, MAP, and EAP methods at the initial, interim and final stages of CAT ability estimation methods. This finding is supported by previous researchers [15], [21] stated that the comparison between maximum likelihood and Bayesian methods produces little difference in observed results but not without some implications [17], also reported an insignificant in the differences observed between weighted likelihood and Bayesian methods. This finding implies that the method used concerning ability estimation is essentially not an end but a means to an end.

According to previous study [3], the maximum likelihood method (MLEF) can be used only when there is a mixed response pattern. On the other, Bayesian methods (MAP and EAP) can be used for any response pattern with less dependence on the item pool's optimality but rather on existing data from students with Bayesian methods [23]. This outcome strengthens the fact that a factor such as response pattern [3], and item pool [21] are determinant factors on the method chosen for ability estimation in designing a CAT program for educational testing. Further stressed was that a CAT requires additional considerations for ability estimation such as adaptivity, dimensions, consistency, and standards with implications for personalization in user environments and artificial intelligence [2]. Also worthy of note is the fact that Bayesian estimator are convenient with small-scaled tests and when ability levels are not extremely low [17]. This shows that CAT can be applied to school-based assessments rather than only standardized [31].

Despite the non-significance among the methods [10], maximum likelihood method treat person abilities as fixed effects which results into an undesirable estimation inconsistency which can be circumvented using bias correction methods. Furthermore, maximum likelihood estimates are deficient when estimates for response patterns occurs with items correct or all incorrect [1]. Bayesian ability estimations may be preferable over maximum likelihood estimators in CAT. They rely less on the selected item's item information, except a prior distribution obtained during the test [21]. Another drawback with maximum likelihood estimation method is that all item and person parameters are regarded as unknowns to be estimated resulting in the occasional non-existence of estimates and the bias of item parameter estimates [32].

Worthy of note, the non-significance across ability estimation methods recorded is also premised on the conditional statistics due to the equivalent design employed in terms of theta ranges and fixed test length. The non-significance recorded shows the need for consistency across CAT designs and the ease of replicating designs once conditions remain constant [13]. The discussions reveal that methodological choices for constructing CATs based on simulation procedures with empirical evidence allow test experts to identify the necessary characteristics of the CAT before actual administration to real examinees.

5. CONCLUSION

It can be concluded that while non-significance in ability estimation methods were recorded across the ML and Bayesian methods were recorded, Bayesian methods with a preference for EAP could be the right choice considering its flexibility with response patterns irrespective of the availability of an optimal item pool typical with early CAT programs as determinant factors. There are several recommendations stemming from this study. Bayesian methods with a preference for the EAP method should be used in designing early CAT programs with paper and pencil alternatives from which a prior distribution can be obtained. The peculiarity of the testing situation should inform methods chosen for the ability estimation of CAT. Equivalence of designs should be ensured when replicating CAT for a testing program. Hence, CAT design should be based on results from the simulation as furtherance research and analyses.

ACKNOWLEDGEMENTS

The authors acknowledge 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

[1] W. J. van der Linden and P. J. Pashley, "Item Selection and Ability Estimation in Adaptive Testing," in *Computerized Adaptive Testing: Theory and Practice*, Springer, Dordrecht, 2000, pp. 1–25, doi: 10.1007/0-306-47531-6_1.

[2] E. Georgiadou, E. Triantafillou, and A. A. Economides, "Evaluation parameters for computer-adaptive testing," *British Journal of Educational Technology*, vol. 37, no. 2, pp. 261–278, 2006, doi: 10.1111/j.1467-8535.2005.00525.x.

[3] D. G. Seo, "Overview and current management of computerized adaptive testing in licensing/certification examinations," *Journal of Educational Evaluation for Health Professions*, vol. 14, p. 17, 2017, doi: 10.3352/jeehp.2017.14.17.

[4] C. Redecker and Ø. Johannessen, "Changing Assessment - Towards a New Assessment Paradigm Using ICT," *European Journal of Education*, vol. 48, no. 1, pp. 79–96, 2013, doi: 10.1111/ejed.12018.

[5] J. I. Oladele, M. A. Ayanwale, and H. O. Owolabi, "Paradigm Shifts in Computer Adaptive Testing in Nigeria in Terms of Simulated Evidence," *Journal of Social Sciences*, vol. 63, no. 1–3, pp. 9–20, 2020, doi: 10.31901/24566608.2020/63.1-3.2264.

[6] O. C. Eneh, "Technology transfer, adoption and integration: A review," *Journal of Applied Sciences*, vol. 10, no. 16, pp. 1814–1819, 2010, doi: 10.3923/jas.2010.1814.1819.

[7] J. W. Richardson, "Challenges of adopting the use of technology in less developed countries: The case of Cambodia," *Comparative Education Review*, vol. 55, no. 1, pp. 8–29, 2011, doi: 10.1086/656430.

[8] B. Dutta, S. Geiger, and T. Lanvin, "The global information technology report," *The World Economic Forum*, vol. 1, no. 1, p. 28, 2015, doi: 10.54666/2334-000-003-002.

[9] D. G. Seo and J. Choi, "Post-hoc simulation study of computerized adaptive testing for the Korean Medical Licensing Examination," *Journal of Educational Evaluation for Health Professions*, vol. 15, p. 14, 2018, doi: 10.3352/jeehp.2018.15.14.

[10] A. Robitzsch, "A Comprehensive Simulation Study of Estimation Methods for the Rasch Model," *Stats*, vol. 4, no. 4, pp. 814–836, 2021, doi: 10.3390/stats4040048.

[11] D. J. Weiss and G. G. Kingsbury, "Application of Computerized Adaptive Testing To Educational Problems," *Journal of Educational Measurement*, vol. 21, no. 4, pp. 361–375, 1984, doi: 10.1111/j.1745-3984.1984.tb01040.x.

[12] S. M. H. T. Hooshang Khoshsima, "Computer Adaptive Testing (CAT) Design; Testing Algorithm And Administration Mode Investigation," *European Journal of Education Studies*, vol. 3, no. 5, pp. 796–807, 2017, doi: 10.46827/ejes.v0i0.725.

[13] K. C. T. Han, "Conducting simulation studies for computerized adaptive testing using SimulCAT: an instructional piece," *Journal of Educational Evaluation for Health Professions*, vol. 15, p. 20, Aug. 2018, doi: 10.3352/jeehp.2018.15.20.

[14] S. M. Čisar, D. Radosav, B. Markoski, R. Pinter, and P. Čisar, "Computer Adaptive Testing of Student Knowledge," *Acta Polytechnica Hungarica*, vol. 7, no. 4, pp. 139–152, 2010, [Online]. Available: http://acta.uni-obuda.hu/.

[15] N. A. Thompson and D. A. Weiss, "A framework for the development of computerised adaptive tests," *Practical Assessment, Research, and Evaluation*, vol. 16, no. 1, p. 1, 2011, doi: 10.7275/wqzt-9427.

[16] E. Latu and E. Chapman, "Computerised adaptive testing," *British Journal of Educational Technology*, vol. 33, no. 5, pp. 619–622, 2002, doi: 10.1111/1467-8535.00296.

[17] D. Magis and G. Raîche, "On the Relationships Between Jeffreys Modal and Weighted Likelihood Estimation of Ability Under Logistic IRT Models," *Psychometrika*, vol. 77, no. 1, pp. 163–169, 2012, doi: 10.1007/s11336-011-9233-5.

[18] D. Cella, R. Gershon, J. S. Lai, and S. Choi, "The future of outcomes measurement: Item banking, tailored short-forms, and computerized adaptive assessment," *Quality of Life Research*, vol. 16, no. SUPPL. 1, pp. 133–141, 2007, doi: 10.1007/s11136-007-9204-6.

[19] B. P. Veldkamp and M. Matteucci, "Bayesian computerized adaptive testing," *Ensaio*, vol. 21, no. 78, pp. 57–82, 2013, doi: 10.1590/S0104-40362013005000001.

[20] J. Olea, J. R. Barrada, F. J. Abad, V. Ponsoda, and L. Cuevas, "Computerized Adaptive Testing: The Capitalization on Chance Problem," *The Spanish journal of psychology*, vol. 15, no. 1, pp. 424–441, 2012, doi: 10.5209/rev_sjop.2012.v15.n1.37348.

[21] T. H. Ho, "A Comparison of Item Selection Procedures Using Different Ability Estimation Methods in Computerized Adaptive Testing Based on the Generalized Partial Credit Model," Dissertation, The University of Texas, 2010, [Online]. Available: http://hdl.handle.net/2152/ETD-UT-2010-05-780.

[22] T. Wang and W. P. Vispoel, "Properties of ability estimation methods in computerized adaptive testing," *Journal of Educational Measurement*, vol. 35, no. 2, pp. 109–135, 1998, doi: 10.1111/j.1745-3984.1998.tb00530.x.

[23] J. I. Oladele, "Computer-Adaptive-Testing Performance for Postgraduate Certification in Education as Innovative Assessment," *International Journal of Innovation, Creativity and Change*, vol. 15, no. 9, p. 2021, 2021, [Online]. Available: https://www.ijicc.net/images/Vol_15/Iss_9/15929_Oladele_2021_E1_R.pdf.

[24] K. T. Han, "User's Manual for SimulCAT: Windows Software for Simulating Computerized Adaptive Test Administration," *Applied Psychological Measurement*, vol. 36, no. 1, pp. 64–66, 2012, doi: 10.1177/0146621611414407.

[25] D. D. R. Meneghetti and P. T. A. Junior, "Application and Simulation of Computerized Adaptive Tests Through the Package catsim," *arXiv: Applications*, 2017, doi: 10.48550/arXiv.1707.03012.

[26] S. W. Choi, "Firestar: Computerized adaptive testing simulation program for polytomous item response theory models," *Applied Psychological Measurement*, vol. 33, no. 8, pp. 644–645, 2009, doi: 10.1177/0146621608329892.

[27] H. Sie, "A Review of SimuMCAT: A Simulation Software for Multidimensional Computerized Adaptive Testing," *Applied Psychological Measurement*, vol. 39, no. 3, pp. 241–244, 2015, doi: 10.1177/0146621615569503.

[28] M. O. Ogunjimi, M. A. Ayanwale, J. I. Oladele, D. S. Daramola, I. M. Jimoh, and H. O. Owolabi, "Simulated evidence of computer adaptive test length: Implications for high stakes assessment in Nigeria," *Journal of Higher Education Theory and Practice*, vol. 21, no. 2, pp. 202–212, 2021, doi: 10.5209/rev_SJOP.2012.v15.n1.37348.

[29] C. J. Huberty and S. Olejnik, "Factorial MANOVA," in *Applied MANOVA and Discriminant Analysis*, John Wiley & Sons, 2009, pp. 131–162, doi: 10.1002/9780471789475.ch8.

[30] N. Hasan, "MANOVA/MANCOVA using SPSS," *Webinars in Psychometrics and Statistics*, 2020, doi: 10.13140/RG.2.2.19846.40004.

[31] C. Demir and B. F. French, "Applicability and Efficiency of a Computerized Adaptive Test for the Washington Assessment of the Risks and Needs of Students," *Assessment*, 2021, doi: 10.1177/10731911211047892.

[32] T. Lando and L. Bertoli-Barsotti, "A modified minimum divergence estimator: Some preliminary results for the Rasch model," *Electronic Journal of Applied Statistical Analysis*, vol. 7, no. 1, pp. 37–57, 2014, doi: 10.1285/i20705948v7n1p37.

BIOGRAPHIES OF AUTHORS

Jumoke I. Oladele holds a permanent faculty position at the University of Ilorin, Nigeria and currently a Postdoctoral Research Fellow at the University of Johannesburg, South Africa. The fellowship has been awarded to support her research in "Computer Adaptive Testing Item Bank Development for Postgraduate Certification in Education in Mathematics". Her specialization is in educational research, measurement, and evaluation. Her research interests are in item response theory-based psychometrics in educational assessments with computer adaptive testing (CAT); applicable to a wide range of subjects with focus on improving standardized assessment in Africa, educational evaluation, and innovation studies. So far, her research has generated a good number of publications. She can be contacted at email: jumokeo@uj.ac.za.

Mdutshekelwa Ndlovu is an Associate Professor of Mathematics Education at the University of Johannesburg. He was awarded a research grant to investigate mathematics pre-service teachers' beliefs about technology integration in the teaching of mathematics. During this time, he was appointed Vice-Dean: Research and Postgraduate Studies in the Faculty of Education, a post that he still holds. Overall, he completed 17 years of teaching and educational leadership in basic education and have now completed 18 years of lecturing and management in higher education. He has accumulated more than 40 publications in mathematics and science education and graduated 6 PhDs, 7 Masters and 10 Honors students. To date, he has examined 22 doctoral and 25 masters' theses and peer-reviewed over 130 pieces of work. He has also delivered several keynote addresses, invited lectures, panel discussions. He can be contacted at email: mndlovu@uj.ac.za.

Erica D. Spangenberg is an Associate Professor in Mathematics Education at the University of Johannesburg (UJ), South Africa. Erica Dorethea Spangenberg currently works at the Department of Science and Technology Education, University of Johannesburg. She lectures preservice teachers specializing in mathematics and supervising postgraduate students in the field of Mathematics Education. Her research niche centers on affective constructs in the teaching and learning of mathematics via technology, and she has also and interest in pedagogy/didactics, assessment, visualization, and ethno-mathematics. She can be contacted at email: ericas@uj.ac.za.