Pros 2022-DETERMINE THE NUMBER OF STOPPING RULES AND MINIMUM ITEMS IN THE COMPUTERIZED

by Hari Susanto

Submission date: 07-Jul-2023 03:38PM (UTC+0700)
Submission ID: 2127635135
File name: MBER_OF_STOPPING_RULES_AND_MINIMUM_ITEMS_IN_THE_COMPUTERIZED.pdf (576.14K)
Word count: 4050
Character count: 18125

Proceeding:
International Conference on Learning and Education (ICLE) Vol 1, 2022

DETERMINE THE NUMBER OF STOPPING RULES AND MINIMUM ITEMS IN THE COMPUTERIZED ADAPTIVE TEST

Hari Purnomo Susanto

STKIP PGRI Pacitan.
Email: haripurnomosusanto@gmail.com

Abstract: This research aims to determine the number of stopping rules and the minimum items required by CAT. To achieve this goal, simulation using CAT based on MLE is carried out. Data of 10000 respondents with 173 items were raised using the R program. The calibration of such data results in 152 fit items with the difficulty being at intervals (-3, 3.21). The θ estimation in CAT was determined using the MLE Algorithm. The CAT simulation was set with 15 multiple choice items. Cat simulation results show a stopping rule can occur in each test number except for numbers 1, 3, and 5. Furthermore, the minimum item required for CAT based on MLE is 122 (80.26%) items from 152 items. The selected items consist of 27.05% (33 items) easy, 51.64% (63 items) medium, and 21.31% (26 items) Difficult.

Keywords: Stopping Rule, Minimum Item, CAT, R Program

1. INTRODUCTION

Along with the development of computer technology and artificial intelligence. The computerized Adaptive Test (CAT) is one of the alternative media for measuring the cognitive, affective, and psychomotor domains. CAT-based tests have several advantages, The first. CAT has proven to be able to present items according to the ability of the test taker (Haley et al., 2011). Secondly, the time required can be shorter (Khoshsima et al., 2017). Third, CAT has a good side of efficiency, accuracy, and practical value (Fadzil, 2018). Fourth, shorten the time required for scoring (Samsudin et al., 2019), fifth. Can reduce cheating during the test (Wulandari et al., 2020a).

CAT works based on student ability (θ) estimation. The θ is used as the basis for selecting the next item. In addition, the θ is also used to determine the test taker score. CAT works according to the abilities of the test takers, so each test taker is very likely to get different items. The uniqueness of CAT is that under certain conditions the test will stop by itself. The CAT will stop the test when the stopping rule is working.

The number of items that the CAT uses for the dichotomous test depends on the following things. The first. The maximum number of items that must be done by the test taker. For example, in a multiple-choice test consisting of 10 items, the number of possible answers that can occur is 2^{10} or 1024 possibilities. Furthermore, from these possibilities, the minimum number of items can be determined. Secondly, item bank. The collection of test items in different item banks will result in different minimums items on the same CAT. Third, the item selection process is initiated by the CAT algorithm. At the beginning of the test, all test takers or test takers can work on the same item or different items, depending on the algorithm used in the CAT.

CAT consists of five components (Eggen, 2012; Retnawati, 2014). The first is the Item bank. This component consists of a collection of items that have gone through a Calibration. The second is, starting rule. The initial conditions on the CAT are defined in this step. This starting rule will greatly affect the minimal number of items on CAT. Third, the θ estimation algorithm, in this section contains algorithms to estimate the θ. The θ value is used to select the difficulty level for the next item. Thus, the selected items have been adjusted based on the test takers' from the previous

questions. fourth, Scoring. Scoring can be done by MLE that is optimized using the Newton-Raphson formula (Retnawati, 2014). Fifth. Stopping rule. The test stops if certain conditions have been met. The test will stop based on the stopping rule that occurs in CAT.

Indirectly, the Stoping rule indicates that the number of items worked on by the test taker is enough to promote his ability to the material being tested. The minimum number of items on the stopping rule must be determined, especially the stopping rule that occurs in low numbers. For example. CAT stops at number 3, whether the three questions can already logically explain the ability of the test taker. This information can be used to set up the algorithm for item selection on the few initial numbers of the CAT.

Based on the information above, this study aims to determine 1) the number of stopping rules that may occur in CAT, and 2) determine the number of minimum items required by CAT. The R program is used to determine the number of stopping rules and minimum items in CAT.

2. METHOD

This research is a simulation study. The steps in this study are, 1) generating research data. 2) Build CAT, and 3) determine the stopping rule and minimum items in CAT. The steps are outlined as follows.

2.1 Data

The data used in this study is generated under certain conditions. The Data is generated by using the R program with the MIRT package (Chalmers, 2012). It was conditioned with 173 items and 10000 respondents. This data has been adapted to the 2 PL model.

2.2 CAT

CAT consists of five components (Eggen, 2012; Retnawati, 2014; Ridwan et al., 2020). The five components are discussed in detail as follows.

2.2.1 Calibration of item banks

The item bank contains a complete set of items with item and abilities parameters. The parameters of the test item are obtained from the process of calibration of 10000 generated data above. The calibration results obtained 152 fit items and obtained a level of difficulty from -3 to 3.2 and discriminant between 0 to 1.23.

2.2.2 starting rules

The test begins with a moderate level of difficulty for all participants. The first difficulty level used is 0.0 (Retnawati, 2014) or close to 0.0. in this simulation, the first difficulty level is the parameter of the 58th item in the item bank.

2.2.3 Item selection rule

The Items are selected based on the selected difficulty level. The difficulty level of the item is selected based on the estimated value of the θ. In this study, the MLE Algorithm was used to estimate the θ.

2.2.4 Scoring

The MLE approach with numerical analysis of the Newton-Raphson method (Retnawati, 2014) can be done to carry out the scoring. The formula with this scoring can be seen in the following formula (1). This section is not done in this CAT simulation study.

$$\theta_{n+1} = \theta_n + \frac{\sum_{i=1}^{n} S_i(\theta_n)}{\sum_{i=1}^{n} I_i(\theta_n)} (1)$$

2.2.5 Stopping rule

The stopping rule in this study is carried out $\Delta\theta = \theta_{i+1} - \theta_i = 0$ if (Ridwan et al., 2020) or the test taker has done 15 items.

2.3 MLE algorithm.

Estimation of the test taker's ability parameters can be performed using the maximum Likelihood model within formula (2) Bellow. (Hambleton et al., 1991; Retnawati, 2014).

$$L(u_1, u_2, \dots, u_n | \theta) = \prod_{j=1}^{n} P_j(\theta)^{u_j} . Q_j(\theta)^{1-u_j} \quad (2)$$

With

$$P_j(\theta) = \frac{e^{(\theta - b_j)}}{1+e^{(\theta - b_j)}} \text{ and } Q_j(\theta) = 1 - P_j(\theta)$$

$P_j(\theta)$: The test taker probability with their ability to answer correctly on the j[th] item.

$Q_j(\theta)$: The test taker probability with their ability to answer wrongly on the j[th] item.

this method can estimate the test taker's ability by choosing the maximum θ probability. The θ value that results from the maximum probability will become the next θ estimation. The θ value is then used as a basis for determining the difficulty of the next item in CAT. The theta used to build the MLE model is at intervals (-3, 3.21). The character of this theta is in the form of an arithmetic sequence with initial value = -3, difference = 0.129, and maximum theta = 3.21.

2.4 Stopping rule and minimum item on CAT.

The maximum number of items that must be done by test takers in this study is 15 test items. The number of stopping rules and the minimum item is carried out with the following steps.

a. Determine the number of response patterns that can occur for 15 test items.
b. determine the number of stopping rules in CAT.
c. Determine the minimum number of items in CAT.

Each calculation for determining many stopping rules and minimum items in CAT is carried out with the help of the R program.

3. Results and discussion

Multiple choice tests were used in this simulation and the number of items that the test takers had to do was 15. Many possible response patterns were 2^15 = 32768. The CAT simulation was carried out 32768 times. The device used in calculations has the specification of 11th Gen Intel(R) Core (TM) i3-1115G4 Processor @ 3.00GHz 3.00 GHz 8 GB ram. The time required to perform the simulation is 10 minutes.

The simulation results can be seen in tables 1, 2, and 3. This section discusses the illustration of CAT work, stopping rules, and minimum items in CAT MLE.

3.1 The CAT Illustration

The simulation results of 32768 response patterns can be seen in Tables 1 and 2. Table 1 is the simulation result of the 17055[th] response pattern. The stopping rule process does not occur in this case, because the CAT stops after the test taker has completed the maximum number of items given (there is no delta value equal to zero).

The process that occurs in CAT for the response pattern 17055 can be described. First, the test taker obtains an item with the number 58 and the test taker gives a response of 1 (correctly). Formula (2) processed discriminant and difficulty, obtained θ estimate 3,321. furthermore, the CAT selects items with a difficulty level close to θ=3.321 and obtained item number 152 with a difficulty level of 3.182. secondly, the test taker responds to the second item with 0 (false). Discriminants and difficulties processed by formula (2) obtained the value of θ=2.57. based on this θ value, the third item is obtained, and get 147[th] item for the next item, and so on.

Table 2 describes the performance of CAT with a stopping rule. The test stops at the fourth item. The stop test is marked by Delta value = 0. So the response pattern (0 1 0 0 1 1 1 0 1 0 0 0 1 1 0) is not used entirely.

Table 1. Illustration of CAT results for response pattern to 17055

No.	**level**	Number choose	discriminant	Difficulty	Responses	Θ	Delta
1	Medium	58	0.7670	0.0141	1	3.321	3.321
2	Hard	152	0.0580	3.1822	0	2.547	-0.774
3	Hard	147	0.5985	2.5309	0	1.128	-1.419
4	Medium	108	0.3062	1.1215	0	0.612	-0.516
5	Medium	83	0.4006	0.6301	0	0.096	-0.516
6	Medium	59	0.7072	0.1004	1	0.741	0.645
7	Medium	88	0.5611	0.7420	0	0.354	-0.387
8	Medium	70	0.7756	0.3463	1	0.741	0.387
9	Medium	87	0.3390	0.7282	0	0.612	-0.129
10	Medium	82	0.4430	0.5888	0	0.354	-0.258
11	Medium	71	0.5471	0.3681	1	0.612	0.258
12	Medium	84	0.9229	0.6472	1	0.870	0.258
13	Medium	93	0.5974	0.8641	1	1.128	0.258
14	Medium	107	0.4666	1.1052	1	1.257	0.129
15	Medium	114	0.5913	1.2646	0	0.999	-0.258

Table 2. Illustration of CAT results for 10055^{th} response pattern

No.	**level**	Number choose	discriminant	Difficulty	Responses	Θ	Delta
1	Medium	58	0.7670	0.0141	0	-3.000	-3.000
2	Easy	1	0.1903	-2.9503	1	-1.710	1.290
3	Easy	8	0.3870	-1.7021	0	-3.000	-1.290
4	Easy	2	0.6903	-2.5278	0	-3.000	0.000

3.2 Stopping Rules and Items

Table 3 shows impressive results. the stopping rule in the second item has the most probability of occurring. The response pattern that produces this stopping rule is a response pattern that has the first and second items the same. There are only three items used in the response pattern, namely items 58, 1, and 152. Item 1 in the item bank is the easiest item and item 152 is the most difficult. Furthermore, none of the tests stopped at the third and fifth numbers. In the CAT with 15 multiple choice items, the test automatically stops after the 15th item is completed.

The BI column illustrates that there are several similar items in each response pattern. and the minimum items in the CAT can be determined by the intersection operation of the BI column above and 122 test items are obtained. so that, 30 items in the item bank are not used altogether. The items that are not used at all are 39, 45, 46, 51, 52, 53, 54, 62, 63, 67, 68, 69, 74, 75, 85, 91, 92, 101, 102, 103, 104, 110, 111, 115, 116, 117, 133, 134, 135, and 138.

4. Discussion

The results of the above analysis explain the weakness of CAT based on MLE. The test stops at the second item when the first and second responses are the same. This result is the following (Haryanto, 2013; Suhardi, 2020). As many as 50% (16384) of the 32768 response patterns resulted in the CAT stopping the test. Only two multiple choice items can measure the ability of test takers. What is your opinion? This can be overcome by adding up the difficulty level of the previous item by 0.5 if the first response is correct and vice versa if the first response is wrong (Suhardi, 2020).

This rule is used after the participant has completed the first item only, after that the next number is the MLE algorithm used.

Table 3. Illustration of CAT results for the 10055th response pattern

SR	P	BI
2nd Item	16384	58, 1, 152
4th Item	6144	58, 1, 8, 2, 152, 147, 151
6th Item	1024	58, 1, 8, 2, 9, 3
7th Item	512	58, 152, 147, 108, 126, 148, 146
8th Item	768	58, 8, 27, 15, 4, 2, 3, 30, 42, 64, 152, 147, 108, 126, 93, 114, 125, 148, 149, 151
9th Item	1088	58, 1, 8, 2, 9, 24, 13, 7, 14, 20, 27, 15, 4, 5, 3, 30, 42, 28, 36, 26, 19, 43, 70, 93, 76, 64, 152, 147, 108, 83, 107, 129, 141, 146, 126, 148, 149, 143, 151
10th Item	704	58, 1, 8, 2, 9, 24, 13, 7, 4, 3, 14, 12, 20, 23, 19, 27, 15, 5, 30, 26, 28, 29, 42, 36, 43, 25, 70, 49, 48, 64, 57, 71, 93, 119, 131, 146, 152, 147, 108, 126, 114, 99, 88, 98, 148, 149, 151, 150
11th Item	864	58, 1, 8, 2, 9, 24, 13, 7, 4, 5, 3, 33, 26, 23, 25, 22, 49, 42, 32, 43, 27, 15, 6, 20, 14, 28, 19, 36, 30, 29, 70, 48, 44, 93, 76, 64, 59, 152, 147, 108, 83, 57, 60, 88, 107, 122, 114, 123, 129, 82, 84, 65, 141, 137, 149, 146, 151, 143, 126, 94, 113, 125, 99, 119, 98, 148, 131, 142
12th Item	608	58, 1, 8, 2, 9, 24, 13, 7, 4, 5, 6, 3, 33, 26, 23, 20, 15, 19, 25, 49, 42, 32, 43, 36, 57, 30, 48, 27, 10, 14, 16, 28, 29, 59, 76, 64, 44, 12, 70, 37, 71, 83, 93, 65, 119, 94, 114, 120, 108, 152, 147, 56, 60, 82, 88, 107, 122, 129, 137, 148, 99, 84, 131, 143, 128, 141, 146, 123, 126, 77, 87, 98, 113, 149, 125
13th Item	796	58, 1, 8, 2, 9, 24, 13, 7, 4, 5, 6, 10, 12, 14, 15, 11, 16, 20, 23, 19, 25, 22, 26, 27, 28, 30, 33, 32, 36, 49, 42, 31, 59, 48, 43, 60, 64, 70, 57, 56, 71, 88, 76, 3, 29, 44, 65, 61, 21, 37, 47, 72, 93, 94, 83, 82, 87, 114, 99, 95, 108, 98, 122, 119, 113, 120, 126, 118, 131, 132, 125, 152, 147, 89, 107, 106, 123, 84, 81, 137, 143, 142, 129, 128, 141, 146, 148, 149, 151, 140, 77
14th Item	694	58, 1, 8, 2, 9, 24, 13, 7, 4, 5, 6, 10, 11, 12, 14, 15, 20, 23, 19, 25, 27, 26, 28, 30, 33, 42, 36, 16, 22, 32, 49, 29, 31, 57, 43, 64, 44, 37, 59, 48, 60, 61, 70, 83, 76, 56, 71, 72, 88, 99, 108, 98, 3, 17, 21, 65, 47, 41, 82, 87, 93, 66, 94, 77, 78, 114, 95, 122, 126, 119, 125, 137, 146, 120, 113, 118, 131, 129, 141, 123, 132, 143, 152, 147, 84, 107, 106, 89, 105, 128, 124, 81, 148, 142, 149, 151, 127, 136, 96
15th Item	701	58, 1, 8, 2, 9, 24, 13, 7, 4, 5, 3, 6, 10, 11, 12, 14, 15, 16, 20, 23, 26, 27, 19, 25, 28, 30, 22, 33, 36, 42, 32, 57, 43, 64, 21, 49, 31, 29, 34, 59, 48, 37, 44, 60, 61, 65, 56, 70, 71, 76, 72, 88, 99, 93, 108, 98, 119, 17, 83, 41, 114, 122, 129, 123, 40, 38, 47, 50, 66, 55, 82, 87, 77, 94, 95, 84, 96, 107, 113, 126, 120, 125, 131, 141, 137,

		118, 143, 147, 142, 152, 81, 89, 90, 86, 106, 105, 148, 128, 146, 132, 127, 145, 151, 140, 136, 139, 78, 79, 97, 112, 100, 149
None of SR	2481	58, 1, 8, 2, 9, 24, 13, 7, 4, 5, 3, 6, 10, 11, 12, 14, 15, 16, 17, 18, 20, 23, 19, 26, 25, 27, 28, 21, 22, 30, 33, 42, 29, 36, 32, 57, 43, 59, 31, 64, 37, 70, 49, 34, 35, 44, 41, 38, 48, 47, 60, 61, 65, 56, 83, 76, 71, 77, 88, 82, 87, 99, 93, 108, 84, 94, 107, 114, 122, 55, 72, 89, 98, 119, 129, 126, 137, 78, 40, 66, 62, 113, 120, 146, 50, 54, 73, 81, 131, 79, 95, 96, 97, 106, 112, 125, 136, 118, 121, 123, 132, 128, 143, 141, 142, 147, 149, 148, 151, 152, 90, 86, 105, 109, 124, 127, 100, 80, 115, 130, 140, 139, 145, 144, 135, 138, 101

SR: Stopping Rule Work at i-item in the test

P: The Number of Possibility of stopping rule

BI: The item number from the bank used item

The stopping rule simplifies the number of possible response patterns. The existence of this stopping rule also proves that CAT can shorten the time needed to carry out the test. This finding is following the results of research from (Khoshsima et al., 2017). The many possible response patterns allow students to work on different items for each test taker. this condition can minimize the occurrence of cooperation between test takers in doing the test (Wulandari et al., 2020a).

the minimum number of items required by CAT for all response patterns is 122 (80.3%), this result is different from (Fauzie et al., 2021) which shows that the selected items are 141 (70.5%) from 200 items. In Fauzie the parameters are determined using the CTT model, while in this study the parameters are determined using IRT 2PL. Furthermore, the percentages of difficulty levels were easy, medium, and difficult items from 122 selected items were 27.05% (33 items), 51.64% (63 items), and 21.31% (26 items). This result is almost close to the proportion of difficulty parameters in the item bank, namely 3-4-3 (30% easy, 40% moderate, and 30% difficult) or 3-5-2 (30% easy, 50% moderate, and 20% difficult) stated by Sudjana (Fauzie et al., 2021).

5. Conclusion

The Stopping Rule illustrates that the response pattern can be simplified. An appropriate and logical stopping rule makes it easy to determine the number of items that must be done by test takers. The test taker's ability is measured according to his ability and does not have to work on the maximum number of items that must be done. furthermore, the stopping rule determines the minimum number of items used in the CAT.

In future studies, the refinement of this research can be done by 1) using more than one theta set that is used to build the MLE model. 2) accessing the number of different response patterns for each stopping rule. 3) compare with other CAT-based algorithms.

REFERENCE

Cella, D., Gershon, R., Lai, J.-S., & Choi, S. (2007). The future of outcomes measurement: item banking, tailored short-forms, and computerized adaptive assessment. *Quality of Life Research*, *16*(S1), 133–141. https://doi.org/10.1007/s11136-007-9204-6

Chalmers, R. P. (2012). Mirt : A Multidimensional Item Response Theory Package for the R Environment. *Journal of Statistical Software*, *48*(6). https://doi.org/10.18637/jss.v048.i06

Eggen, T. J. H. M. (2012). Computerized Adaptive Testing Item Selection in Computerized Adaptive Learning Systems. In *Psychometrics in practice at RCEC*. The University of Twente. https://doi.org/10.3990/3.9789036533744.ch2

Fadzil, H. M. (2018). DESIGNING INFOGRAPHICS FOR THE EDUCATIONAL TECHNOLOGY COURSE: PERSPECTIVES OF PRE-SERVICE SCIENCE TEACHERS. *Journal of Baltic Science Education*, *17*(1), 8–18. https://doi.org/10.33225/jbse/18.17.08

Fauzie, M., Pada, A. U. T., & Supriatno, S. (2021). Analysis of the difficulty index of item bank according to cognitive aspects during the Covid-19 pandemic. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *25*(2). https://doi.org/10.21831/pep.v25i2.42603

Georgiadou, E., Triantafillou, E., & Economides, A. A. (2006). Evaluation parameters for computer-adaptive testing. *British Journal of Educational Technology*, *37*(2), 261–278. https://doi.org/10.1111/j.1467-8535.2005.00525.x

Haley, S. M., Coster, W. J., Dumas, H. M., Fragala-Pinkham, M. A., Kramer, J., Ni, P., Tian, F., Kao, Yi.-C., Moed, R., & Ludlow, L. H. (2011). Accuracy and precision of the Pediatric Evaluation of Disability Inventory computer-adaptive tests (PEDI-CAT). *Developmental Medicine & Child Neurology*, *53*(12), 1100–1106. https://doi.org/10.1111/j.1469-8749.2011.04107.x

Hambleton, R., Swaminathan, H., & Rogers, H. J. (1991). *fundamental of item response theory*. SAGE.

Haryanto, H. (2013). PENGEMBANGAN COMPUTERIZED ADAPTIVE TESTING (CAT) DENGAN ALGORITMA LOGIKA FUZZY. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *15*(1). https://doi.org/10.21831/pep.v15i1.1087

Khoshsima, H., Morteza, S., & Toroujeni, H. (2017). COMPUTER ADAPTIVE TESTING (CAT) DESIGN; TESTING ALGORITHM AND ADMINISTRATION MODE INVESTIGATION. *European Journal of Education Studies*, *3*.

Retnawati, H. (2014). *Teori Respons Butir dan Penerapannya*.

Ridwan, W., Wiranto, I., & Dako, R. D. R. (2020). Ability estimation in Computerized adaptive test using Mamdani Fuzzy Inference System. *IOP Conference Series: Materials Science and Engineering*, *850*(1), 012004. https://doi.org/10.1088/1757-899X/850/1/012004

Samsudin, M. . T., SomChut, M. E., & Ismail. (2019). Evaluating Computerized Adaptive Testing Efficiency in Measuring Students' Performance in Science TIMSS. *Jurnal Pendidikan IPA Indonesia*, *8*(4). https://doi.org/10.15294/jpii.v8i4.19417

Suhardi, I. (2020). Alternative item selection strategies for improving test security in computerized adaptive testing of the algorithm. *Research and Evaluation in Education*, *6*(1). https://doi.org/10.21831/reid.v6i1.30508

Wulandari, F., Hadi, S., & Haryanto, H. (2020). Computer-based Adaptive Test Development Using Fuzzy Item Response Theory to Estimate Student Ability. *Computer Science and Information Technology*, *8*(3). https://doi.org/10.13189/csit.2020.080302

Pros 2022-DETERMINE THE NUMBER OF STOPPING RULES AND MINIMUM ITEMS IN THE COMPUTERIZED

ORIGINALITY REPORT

5%	6%	2%	1%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

MATCH ALL SOURCES (ONLY SELECTED SOURCE PRINTED)

3%

★ repository.unpkediri.ac.id

Internet Source

Exclude quotes	On	Exclude matches	< 1%
Exclude bibliography	On		