VOL. 11, NO. 3, FEBRUARY 2016 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

THE MODEL FOR CALCULATING THE LIFETIME OF ASSESSMENT ITEM BANKS AND REPOSITORIES

Dmitry Ivanovich Popov and Elena Dmitrievna Popova
Moscow State University of Printing Arts, Pryanishnikova 2A, Moscow, Russian Federation
E-Mail: damitry@mail.ru

ABSTRACT

The model which can be used to make a prediction about how long you need to replace the assessment item bank (AIB) or make significant changes to it (or update it) is considered in the paper. The target function is proposed to predict the number of students' testing sessions, on the assumption that students purposefully are making copies and exchanging each other of test items and answers to them, so actually after a while they have a large number of templates with ready-made answers. In this case, the test does not give an objective assessment of knowledge and an item bank has to be updated. The paper provides a formula for calculating the cardinality of assessment item bank, length of single test or amount of tasks (items) per one session of testing, and the amount of testing sessions during the entire period of use of the item bank.

Keywords: knowledge processing system, computer testing, computer-based teaching, item banks, knowledge assessment.

1. INTRODUCTION

Using computer-based testing in order to assess the level of student's knowledge in different universities regularly confronted with attempts of the assessment item banks and repositories [1] hacking during attestation procedure [2]. The task of increasing the AIB confidentiality level is rather important and can be solved by different methods of the information security theory. But, despite of the great amount of items in AIB, it cannot be protected in any means from copying and saving test materials that appear on the computer screen during the computerized testing process. For example, the copying and sharing test materials from a screen between students occurs. They can make screenshots and send the items from the AIB to each other and so on. That's mean in the next assessment session test items becomes easy to access and probability of getting correct conclusion on test estimation with a part of bias is highly increased. Using of such "cribs" during the testing process fully discredits the whole idea of computer assessment and, as a result, wrongly increases the real result of test as well as assessment of the knowledge level of student.

According to this statement the task appears, in which have to predict the optimal amount of sessions which need to spend for recognizing all items from AIB (Figure-1). At this picture blue area is a full AIB; each session covers just a little part of AIB. That's mean the number of used items per session much less than AIB cardinality. But after several sessions all blue area will covered. So all the items (test tasks) were used from the item bank [3]. We should figure out how many sessions will cover entire item bank. On the other hand, it would be interesting to know the best AIB cardinality with exact terms of use, as well as to count the recommended test length (number of items per testing session) in view of multiple use.

2. METHODOLOGY

To solve there should be added next parameters:
N = AIB cardinality - amount of items in AIB,
L = length of single test - the amount of items (tasks) per one session of testing,
M = the amount of testing sessions during of all period using AIB.

Also k = coefficient with value between 0 and 1 called "Coefficient of Aggressive Environment" (CAE) need to be added. The meaning of "Aggressive Environment" is the contingent of testees that seeks to share all known test tasks and right answers on them. In this way CAE is 1 (i.e. the percent of test recognition is maximum). If we take the example when testees copy and share a half of submitted test items so coefficient k will be 0,5.

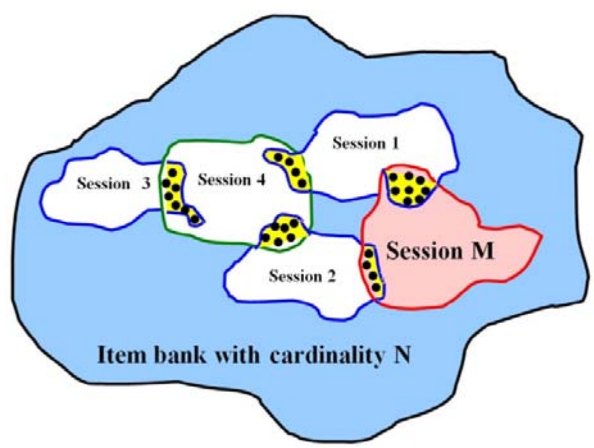

Figure-1. Item bank and testing sessions.

Next it is necessary to find the probability p that all items from AIB are known and copied. This probability may be interpreted as a part of items from AIB that testees will know after passing M sessions of testing. Note than term "known" in this content don't mean "learned", but mean that student has an answer to the item even without understanding the meaning of the item from bank. Below we also will use term "recognized" [4].

VOL. 11, NO. 3, FEBRUARY 2016 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

Now consider how the k-coefficient can be used. Imagine the situation when the session of testing held in “Full Aggressive Environment”, i.e. k = 1. Then within one session every L of test items from bank become known to all testee. If $k \neq 1$ so, taking into account the physical sense of k – coefficient, then kL of test items become available (recognized), otherwise, the same situation taking place, when all tasks are remembered when test length us kL. Thus k-coefficient may be used for correcting the length of test L and in the further the corrected amount kL will be used.

Assume the test items are randomly taken from the bank during the test process. The probability that certain test item would not be chosen during one session is:

$$p_1 = \frac{N - kL}{N}.$$

The same probability $p_2 = \frac{N - kL}{N}$ will also be during the second testing session and in M-session of testing.

Since every session of testing is an independent event so the general probability that certain task will not be chosen after M sessions of testing can be counted by multiplication of probabilities p_i. Thus

$$p_0 = \left(\frac{N - kL}{N}\right)^M.$$

Then probability p can be calculated as

$$p = 1 - p_0 = 1 - (1 - kL/N)^M. \quad (1)$$

From this expression (1) we can find M – number of sessions that is required, to find pN test tasks from AIB:

$$M = \frac{log(1-p)}{log\left(\frac{N-kL}{N}\right)} = \frac{log(1-p)}{log(1-kL/N)}. \quad (2)$$

Similarly, from expression (1) we can find minimal length of the test, at given values M, N, k, p:

$$L = N\left(1 - \sqrt[M]{1-p}\right)/k. \quad (3)$$

And finally, from (3) known parameters M, N, k, p help us to determine the recommended size of the AIB:

$$N = kL/\left(1 - \sqrt[M]{1-p}\right). \quad (4)$$

3. RESULTS AND DISCUSSIONS

Calculation tables and graphs

For ease of calculations, which can be used for examining the quality of item banks, we built tables following formulas (1), (2), (3), (4). In this section is given an example of the table for the calculation of the share of p known items in the AIRB after M testing sessions with different length of single test L and fixed k = 0,1 and N = 1000 (Table-1).

Table-1. The calculation of the share of the known items in AIB after M testing sessions with different length of single test L and fixed k = 0,1, N = 1000.

M	**L, length of single test**				
	30	**40**	**50**	**60**	**70**
20	0,0583	0,0770	0,0954	0,1134	0,1311
40	0,1132	0,1481	0,1817	0,2139	0,2450
60	0,1650	0,2138	0,2597	0,3031	0,3439
80	0,2137	0,2743	0,3304	0,3821	0,4299
100	0,2595	0,3302	0,3942	0,4522	0,5046
120	0,3027	0,3818	0,4520	0,5143	0,5696
140	0,3434	0,4294	0,5043	0,5694	0,6260
160	0,3817	0,4734	0,5516	0,6182	0,6750
180	0,4177	0,5140	0,5943	0,6615	0,7176
200	0,4517	0,5514	0,6330	0,6999	0,7546
220	0,4837	0,5859	0,6680	0,7339	0,7868
240	0,5138	0,6178	0,6997	0,7641	0,8147
260	0,5421	0,6473	0,7284	0,7908	0,8390
280	0,5688	0,6745	0,7543	0,8146	0,8601
300	0,5940	0,6995	0,7777	0,8356	0,8784
320	0,6177	0,7227	0,7989	0,8542	0,8944
340	0,6400	0,7440	0,8181	0,8708	0,9082
360	0,6610	0,7638	0,8354	0,8854	0,9203
380	0,6807	0,7820	0,8511	0,8984	0,9307
400	0,6993	0,7988	0,8653	0,9099	0,9398
420	0,7169	0,8143	0,8782	0,9201	0,9477

It means, for example, if we use test with length L = 60, and already have 400 testing sessions at fixed item bank cardinality N = 1000 and CAE k = 0, 1 then share of known items from bank is 0, 9099. So, more than 0, 9099*1000 ≈ 909 items from bank are open for students.

On Figure-2 is shown a three-dimensional graph of function for variable M from N and L at given probability p = 0, 99 and k = 0, 1.

ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

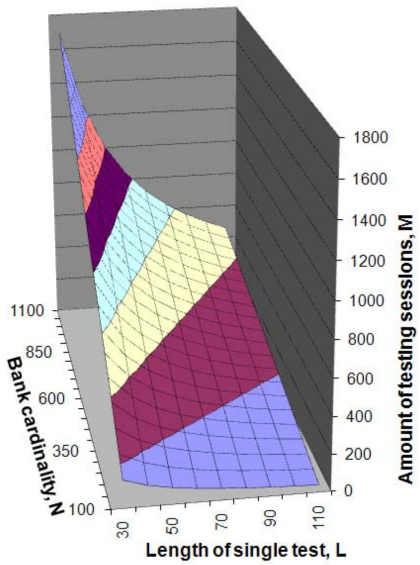

Figure-2. Graph of dependence M (N,L).

Let us fix $k = 0,1$, $p = 0,99$ and select different discrete values for N, next calculate dependence of amount of testing sessions - M from length of single test L with different values of N (Figure-3). At these examples and below we use $p = 0, 99$ by meaning, that almost all items from the bank are recognized. Precisely this case is of greatest interest.

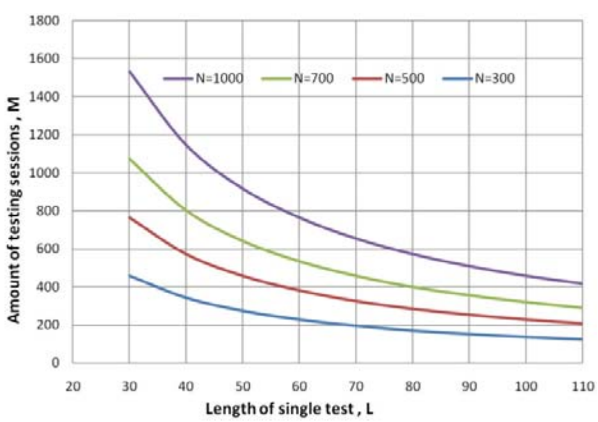

Figure-3. Dependence M(L) at fixed k = 0,1 and p = 0, 99 and discrete N.

At last, dependence M(N) for fixed amounts $k = 0,1$, $p = 0,99$ and discrete values L is shown on Figure-4.

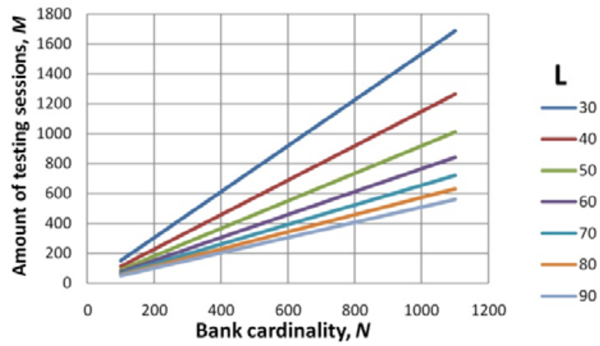

Figure-4. Session amount M dependence of N for fixed k = 0, 1 and p = 0, 99.

Development of algorithms for simulation

To check the correctness of the above formulas conduct simulation of testing process [5]. We will use modeling algorithms based on the testing system functionality taking into account the random [6] design presentation of testing items from bank. Parameters p, M, L, N are modulating in algorithms and matched with values that are calculated in appropriated formula (1), (2), (3), (4). As an example, consider the main steps of the algorithm simulation proportion of recognized tasks p, which can be used for verify basic formula (1):

0. Begin.
1. Set N – AIB cardinality.
2. Set L – length of single test.
3. Set M – amount of testing sessions.
4. Set k – coefficient of aggressiveness.
5. Clear array Used – used items from bank.
6. Amount of testing sessions modeling, number of current testing session j=1.
7. Cycle on amount of testing sessions: While j < M :
7.1. The serial number of the current test item i=1.
7.2. Cycle on L: while i <= L :
7.2.1. Generate random test item from AIB.
7.2.2. Check using k: is it really current test item is shared.
7.2.3. If it is shared, then save the number of task in array Used.
7.2.4. Take next item: i=i+1.
7.3. And of cycle on L.
7.4. Modeling next testing session: j=j+1.
8. End of cycle on amount of testing sessions.
9. Print a share of used test items: p = length (Used) / N.
10. End.

As a result of this algorithm we obtain a data from Table-2 to initial parameters M, L, with fixed N = 1000 and k = 0, 1.

VOL. 11, NO. 3, FEBRUARY 2016 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

Table-2. Modeling a share of known tasks from AIB after M testing sessions with fixed k = 0, 1, N = 1000.

M	L, length of single test				
	30	40	50	60	70
20	0,061	0,068	0,100	0,092	0,131
40	0,110	0,116	0,174	0,224	0,237
60	0,171	0,240	0,250	0,302	0,329
80	0,191	0,273	0,330	0,377	0,423
100	0,263	0,328	0,376	0,448	0,509
120	0,287	0,387	0,462	0,519	0,555
140	0,337	0,452	0,499	0,552	0,632
160	0,385	0,456	0,565	0,614	0,669
180	0,440	0,522	0,606	0,643	0,726
200	0,431	0,542	0,620	0,691	0,776
220	0,513	0,577	0,657	0,741	0,809
240	0,530	0,608	0,709	0,762	0,806
260	0,530	0,637	0,730	0,774	0,837
280	0,575	0,661	0,739	0,827	0,867
300	0,573	0,702	0,776	0,834	0,889
320	0,610	0,727	0,800	0,865	0,892
340	0,621	0,740	0,809	0,874	0,908
360	0,657	0,769	0,838	0,886	0,930
380	0,669	0,796	0,831	0,901	0,920
400	0,701	0,810	0,859	0,908	0,948
420	0,714	0,797	0,875	0,925	0,949

Comparing theoretical results and experimental data

Let's compare results made with formula (1) (Table-1) with experimental results (Table-2). In Table-3 cells put $|x_i - y_i|$ – the module of difference of corresponding cells in Table-1 and Table-2 [7].

Table-3. Comparing theoretical and experimental results.

M	L, length of single test				
	30	40	50	60	70
20	0,0027	0,0090	0,0046	0,0214	0,0001
40	0,0032	0,0321	0,0077	0,0101	0,0080
60	0,0060	0,0262	0,0097	0,0011	0,0149
80	0,0227	0,0013	0,0004	0,0051	0,0069
100	0,0035	0,0022	0,0182	0,0042	0,0044
120	0,0157	0,0052	0,0100	0,0047	0,0146
140	0,0064	0,0226	0,0053	0,0174	0,0060
160	0,0033	0,0174	0,0134	0,0042	0,0060
180	0,0223	0,0080	0,0117	0,0185	0,0084
200	0,0207	0,0094	0,0130	0,0089	0,0214
220	0,0293	0,0089	0,0110	0,0071	0,0222
240	0,0162	0,0098	0,0093	0,0021	0,0087
260	0,0121	0,0103	0,0016	0,0168	0,0020
280	0,0062	0,0135	0,0153	0,0124	0,0069
300	0,0210	0,0025	0,0017	0,0016	0,0106
320	0,0077	0,0043	0,0011	0,0108	0,0024
340	0,0190	0,0040	0,0091	0,0032	0,0002
360	0,0040	0,0052	0,0026	0,0006	0,0097
380	0,0117	0,0140	0,0201	0,0026	0,0107
400	0,0017	0,0112	0,0063	0,0019	0,0082
420	0,0029	0,0173	0,0032	0,0049	0,0013

As it is shown in results of comparing (Table-3), experimental data differs in the third decimal digit from calculated using formula (1). This means that we achieved high accuracy in our calculations [8].

Next, let's construct the graphs of experimental and theoretical results dependency p(M) with fixed k = 0,1, N = 1000 (Figure-5) and various L. As it is shown on this picture theoretical and experimental results are very close.

Finally, let's compare theoretical and experiment results using the Pearson correlation coefficient:

$$r_{xy} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - \left(\sum x_i\right)^2}\sqrt{n\sum y_i^2 - \left(\sum y_i\right)^2}}, \quad (5)$$

where x_i – data from table 1, y_i – data from Table-2, n=105 – quantity of data values in Table-1.

So, substituting the data in the formula (5), the Pearson correlation coefficient for tables 1 and 2 will be 0,9988, that says about the complete relationship between theoretical and experiment results. Similar results were obtained [9] through for calculation formulas (2), (3), (4). Thus results obtained through the analysis suggest that formulas (1), (2), (3), (4) are valid and can be used for counting parameters of AIB.

Restrictions

As it shown in classical test theory [10] the length of single test must be between 40 and 70 on two reasons: small length of tests cause a little reliability of the test and very large length of test cause weariness of testee. From another hand, a big size of item bank will cause high financial expenses for it development. So you need to take into account the financial possibilities of the organizers of the item bank development and plan the possible cardinality N.

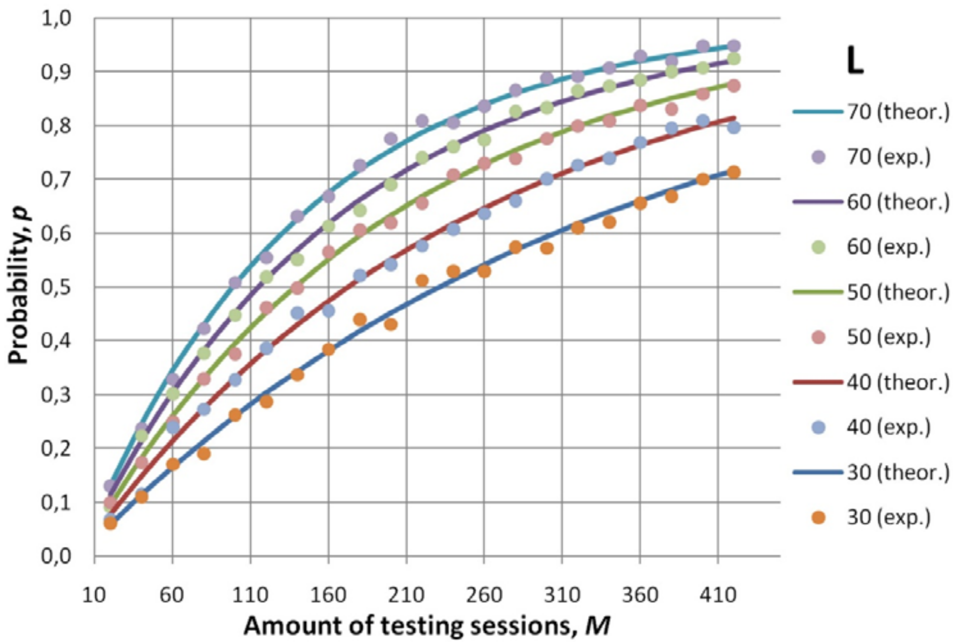

Figure-5. Theoretical and experimental results in compare.

4. CONCLUSIONS AND FUTURE WORK

In this paper we proposed the model which can be used to make a prediction about how long you need to replace AIB or make significant changes to it, for example - update it. Also we found dependences between bank cardinality N, length of single test L and the amount of testing sessions M during of all period using AIB. All given results were verified by experimentally. For the first time it was proposed to use Coefficient of Aggressive Environment for best setting of parameters the model. On the basis of results the following protection ways for AIB can be suggested:

- Regular update of AIB, using suggested model for calculating the optimal interval of updating. Assessment items banks should be usually increased by new items (tasks);
- Increasing the control level of testing process and process of sharing items between testees;
- Timely improvement criterion score of test. Depending on the students knowledge level where the testing process is taking place it is necessary to improve the criterion score after a certain time, using suggested model.

Thus, received results allow predicting the update time of AIB, making a cardinality of bank calculation with current parameters of aggressive environment and required amount of testing sessions, as well as correct conditions of AIB functionality and fix criterion score.

For future work it is necessary to research values of coefficient *k* taking into account the different conditions of educational process, features of computer-based testing, and the students' motivation. It may be expert method with possible base of rules [11] or method using fuzzy logic, but as a result it has to give reasonable value of the CAE *k* for different cases and situations. Another interesting way of research is checking the model by using different values of *p* in relation with various *k*.

REFERENCES

[1] Sarah Currier. 2007. Assessment item banks and repositories. Centre for Educational Technology and Interoperability. Available: http://wiki.cetis.ac.uk/Assessment_item_banks_and_repositories.

[2] Popov D.I. 2001. Designing the intelligent distance learning systems. D.I. Popov. Proceedings of Southern Federal University. Engineering science, T. 22. № 4. pp. 325-332.

ARPN Journal of Engineering and Applied Sciences

[3] Eggen T. J. H. M. 1999. Item selection in adaptive testing with the sequential probability ratio test. Applied Psychological Measurement. 23, pp.249-261

[4] Mardberg B. and Carlstedt B. 1998. Swedish Enlistment Battery: Construct validity and latent variable estimation of cognitive abilities by the CAT-SEB. International Journal of Selection and Assessment. 6, pp. 107-114.

[5] Popov D.I. 2011. Adaptive training strategy of enterprise staff. D.I.Popov, D.G.Demidov. In the World of Scientific Discoveries. (9): 65-71.

[6] Popov D.I. 2011. Choice of algorithms for the estimation the test results. D.I.Popov, D.A.Arsentiev. In the World of Scientific Discoveries. (9): 60-65.

[7] Popov D.I. 2012. Using interactive forms of education in the formation of adaptive abilities of students. T.I.Komolova, D.I.Popov, E.D.Popova. Proceedings of the Higher Educational Institutions. Problems of Printing Arts and Publishing. (4): 65-71.

[8] Popov D.I. 2009. Information Technology. Databases. D.I. Popov, E.D. Popova. - Moscow: MSUPA. p. 117.

[9] Popov D.I. 2007. Automation of control staff certification processes at industrial enterprises. D.I. Popov. - Moscow: MSUPA. p. 178.

[10] Allen M.J. and Yen W. 2002. M. Introduction to Measurement Theory. Long Grove, IL: Waveland Press.

[11] Ostroukh A.V., Popov D.I, Demidov D.G. 2015. Development of the rules base for an expert system choice adaptive learning strategy. ARPN Journal of Engineering and Applied Sciences 10(10): 4430-4435.