

# Computer Assisted Assessment: How many questions are enough?

By

**Rosie Cornish\***

**Charles Goldie\*\***

*and*

**Carol Robinson\***

*\*Mathematics Education Centre  
Loughborough University*

**Web:** <http://mec.lboro.ac.uk>

**Email:** R.Cornish@lboro.ac.uk

C.L.Robinson@lboro.ac.uk

*\*\*Department of Mathematics  
University of Sussex*

**Web:** <http://www.sussex.ac.uk/maths/>

**Email:** C.M.Goldie@sussex.ac.uk

## Abstract

Currently many universities make use of computer assisted assessment (CAA) as part of student assessment. Often these assessments take the form of a test where each question is randomly selected from a bank of alternatives. In this situation, the number of possible tests can be very large. For example, a test may have five questions and for each question there may be a bank of ten alternatives. The number of possible tests would be  $10^5$ .

This paper addresses the issue of how many tests, on average, need to be generated before all available questions have appeared at least once. By means of probability theory and the use of a computer algebra package, results are generated for a number of typical situations. Some surprising results are reported. For the example quoted above, it is found that, on average, only forty-three tests need to be generated before all available questions will have appeared at least once.

The implications of these results, with reference to plagiarism, are discussed. Recommendations on how to reduce the possibility of plagiarism are provided.

## Introduction

Many universities use CAA extensively. Often banks of questions are written on specific topics and, when a student sits a test, questions are selected at random from banks of alternatives. The issue then arises as to how many

questions one should put in each bank of alternatives. Clearly there are many factors to consider here, such as the resources available, the number of students sitting the test and whether or not all the students sit the test simultaneously. With large class sizes and small computer laboratories, it is often not possible to ensure the latter and the possibilities for plagiarism then need to be considered. We use the word ‘plagiarism’ in a broad sense to cover various forms of computer-based cheating. For instance a student who sits a test may take screen shots of questions, paste these into a document and then email it to a friend who has yet to sit the test. Or, for unsupervised tests with 24-hour access, a student may generate so many as to see all questions in the banks of alternatives before finally taking the test. If either of these can occur it is important to consider how many tests need to be taken, on average, before all questions in the banks of alternatives have appeared at least once. If this number is unfeasibly large then one can safely assume that a student logging on to take a test will not have seen all the questions selected. Otherwise plagiarism is a possibility and more questions may need to be written to lessen the danger. This paper addresses this issue and determines, for a number of typical scenarios, the average number of tests that need to be taken before all questions have appeared at least once.

Firstly two examples are presented to demonstrate the approach and then probability theory is used and general results derived. Following on from this recommendations for practitioners writing CAA questions are made and, finally, conclusions are presented.

## Examples

1. Consider a test with just one question and a bank of ten alternatives. Altogether there are ten possible tests. To calculate the average number of tests that need to be taken until all questions have appeared at least once, the classic coupon collector problem (see e.g. Ross (2003)) can be used. This shows that, on average, after

$$1+10/9+10/8+\dots+10/2+10 = 29 \text{ tests (approximately),}$$

all questions have appeared at least once.

2. Now consider a test with five questions, with banks of ten alternatives for each. Altogether there are now  $10^5$  possible tests. However, this time the classic coupon collector problem cannot be used. Computer simulation, using Excel, was used and led to the following results:

<b>Test with a bank of 10 alternatives for each question</b>					
Number of questions in test	1	2	3	4	5
Number of tests required	<b>29</b>	<b>35</b>	<b>39</b>	<b>41</b>	<b>43</b>

Table 1 – Average number of tests required for all questions in a test to have appeared at least once. (The test has 1-5 questions and banks of 10 alternatives.)

From Table 1 it can be seen that, for a test with 5 questions, after an average of approximately 43 tests all questions will have appeared at least once.

These two examples yield some surprising results. Although the number of possible tests has increased from 10 to  $10^5$ , the average number of tests required before all questions have appeared at least once has only increased from 29 to 43. With class sizes in many universities of the order of one or two hundred students, it can be seen that plagiarism could potentially be a problem if the number of questions and banks of alternatives presented here were to be used.

## Probability Theory

Obtaining predictions via computer simulation, as in Example 2, is time consuming and only yields approximate answers. Thus a method, based on probability theory, for calculating the expected number of tests required was sought. The number of tests,  $N$ , that need to be generated in order that all questions have appeared at least once, is a random quantity. Its probability distribution can be obtained by extending well-known calculations for the coupon-collector's problem, or by specialising the results for a more general setting that are given in Adler and Ross (2001). From it one can obtain equation (1), which gives the mean value or expectation of  $N$ . In this equation,  $q$  is the number of questions in the test and  $a$  is the number of alternatives in a bank.

$$E(N) = \sum_{r=0}^{\infty} \left[ 1 - \left( 1 - \sum_{j=1}^a (-1)^{j+1} \binom{a}{j} \left( 1 - \frac{j}{a} \right)^r \right)^q \right] \quad (1)$$

## Results

Using Equation (1), results can now be generated for any test with a given number of questions,  $q$ , and a given number of alternatives,  $a$ , for each question. Figure 1 shows the results obtained for the situation where there are ten alternatives in each bank and the number of questions in the test varies from 0 to 20. Matlab (<http://www.mathworks.com/>) was used to produce the results; the Matlab code for Figure 1 is available in Appendix A. Truncating the infinite series in Equation (1) to the first few hundred terms leads to results which are accurate enough for the purposes of this investigation. (In fact the authors converted the above infinite series to a finite series and checked results using this exact answer. However it was found to be easier in practice to use the truncated infinite series of Equation (1) and hence this is used in the Matlab code of Appendix A.)

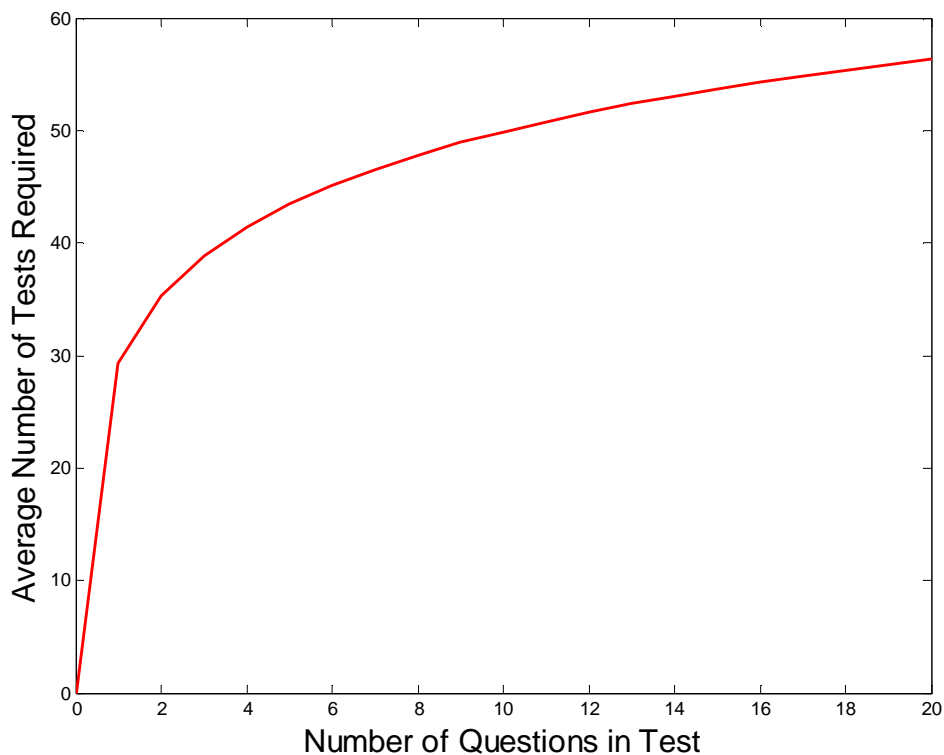


Figure 1 – Average number of tests required for all questions in a test to have appeared at least once. (The tests have up to 20 questions and banks of 10 alternatives.)

The results in Figure 1 confirm the results obtained in Examples 1 and 2 that, for tests with 1-5 questions and banks of ten alternatives for each question, the average number of tests required ranges from 29 to 43. In fact, for a test with 20 questions and banks of ten alternatives for each question, the number of tests required is found to be only 56. These results led the authors to investigate further: Figure 2 extends the results obtained by considering tests with up to 200 questions.

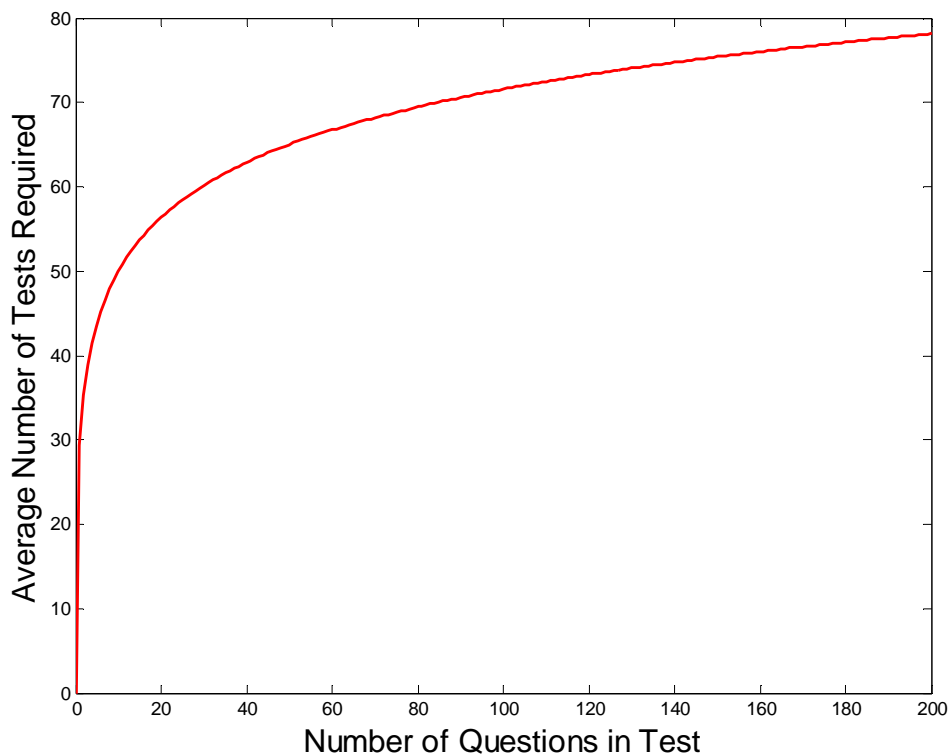


Figure 2 – Average number of tests required for all questions in a test to have appeared at least once. (The tests have up to 200 questions and banks of 10 alternatives.)

Figure 2 demonstrates that as the number of questions in a test is increased, with the number of alternatives for each question remaining constant, the average number of tests required increases very slowly. For a test with 200 questions, each having a bank of 10 alternatives, the number of tests required is 78. Thus although the number of questions which have been written is 2000, on average all questions will have appeared at least once by the time 78 students have taken a test.

Clearly many different scenarios could be investigated. Figures 1 and 2 cover the situation where the number of questions in a test is varied. Figure 3 shows the results obtained when the number of questions in each bank of alternatives is also varied. As would be expected, it is found that when there are more alternatives in each bank then more tests are required before all questions have appeared at least once in test.

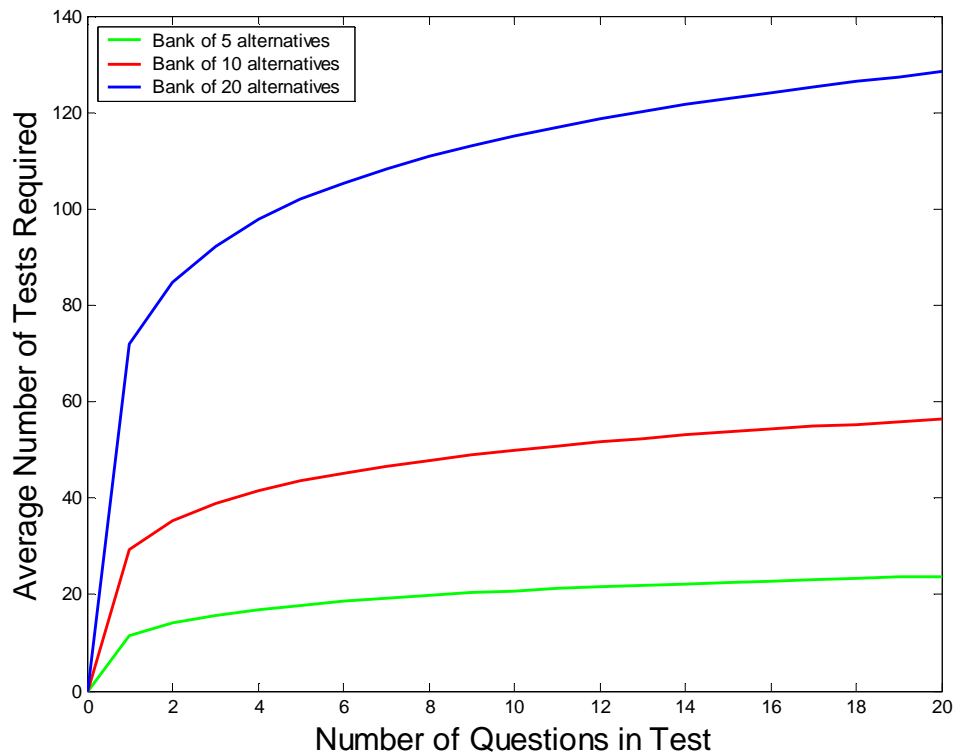


Figure 3 – Average number of tests required for all questions in a test to have appeared at least once. (The tests have up to 20 questions and banks of 5,10 and 20 alternatives.)

Table 2 summarises the results of Figure 3 for the case of 20 questions in a test with each question having 5, 10 or 20 alternatives. It can be seen that when the number of alternative questions in a bank is increased to 20, the number of tests required increases to 129. The likelihood of plagiarism in this situation is remote. It is also interesting to note that as the number of test questions written doubles (from 100 to 200 to 400) the number of tests required increases by a factor of just over two whereas the number of available tests increases enormously (from  $5^{20}$  to  $20^{20}$ ).

Test with 20 questions			
Alternatives for each question	5	10	20
Total number of questions	100	200	400
Number of possible tests	$5^{20}$	$10^{20}$	$20^{20}$
Number of tests required	24	56	129

Table 2 – Average number of tests required for all questions in a test to have appeared at least once. (The test has 20 questions and banks of 5,10 or 20 alternatives.)

## Recommendations

It is clear from the previous section that the possibilities of plagiarism are reduced if the number of questions in a test is increased and, more significantly, if the number of alternatives in each question bank is increased. However, practitioners writing questions for CAA clearly need to decide how many questions they should put in each bank of alternatives. In this section two different situations will be considered to illustrate how this work can be applied.

1. If a practitioner has time to write 100 questions for a test, how many questions should the test contain and how many alternatives should be in each bank? Table 3 presents findings from the cases where the number of alternatives in each bank is 5, 10 or 20 and the test thus has 20, 10 or 5 questions, respectively. The average number of tests required before each question has appeared at least once increases from 24 to 102 as the number of alternatives in a bank increases from 5 to 20. Thus, considering the three possibilities presented here, it is clear that to increase the number of tests needed before each question has appeared at least once, the practitioner should write 5 questions, each with 20 alternatives. Clearly, one must balance the benefits of reducing the risk of plagiarism against the fact that a test with only 5 questions may not meet the assessment requirements of the course.

<b>100 questions available</b>			
Alternatives for each question	<b>5</b>	<b>10</b>	<b>20</b>
Number of questions in test	<b>20</b>	<b>10</b>	<b>5</b>
Number of possible tests	<b><math>5^{20}</math></b>	<b><math>10^{10}</math></b>	<b><math>20^5</math></b>
Number of tests required	<b>24</b>	<b>50</b>	<b>102</b>

Table 3 – Average number of tests required for all questions in a test to have appeared at least once. (For each test 100 questions are available.)

2. Now consider the situation of a class of 150 students. It is decided that a test will consist of 15 questions and that the average number of tests required before each question has appeared at least once should be at least 40. Figure 3 can be used to determine how many alternatives should be in each bank. It can be seen that banks of 5 alternatives would not be enough, but a bank of 10 alternatives would satisfy this requirement. Clearly Figure 3 could be refined to consider cases where the number of banks of alternatives differs from 5, 10 or 20. However, it is sufficient for the purposes of illustration to consider these three cases only.

The above two situations illustrate the application of this work in a practical setting and demonstrate how the solution of Equation (1) can be used to aid decision making regarding the number of questions to include in banks of alternatives.

## Conclusions

This paper has addressed the issue of how many tests, on average, need to be generated before all available questions in banks of alternatives have appeared at least once.

Using probability theory and a computer algebra package, results were generated for a number of typical situations. Some rather surprising results were reported: although the number of available tests increases enormously as the number of questions in the banks of alternatives increases, the number of tests that need to be generated before all available questions have appeared at least once grows very slowly. For example, for a test with 20 questions, each having 10 alternatives, there are  $10^{20}$  available tests, yet it was found that the average number of tests which need to be generated before all available questions have appeared at least once is just 56.

With the situation of large class sizes and small computer laboratories in many universities, it is often not possible to ensure that all students sit a computer based test simultaneously. As students may be able to take screen shots of questions and send these to their peers, there is the potential for plagiarism. This work demonstrates that the number of tests that need to be generated before all available questions have appeared at least once can be quite low. It is advisable that practitioners consider this issue when deciding how many questions to put in the banks of alternatives. In the recommendations section specific advice is given as to how a practitioner might use the results in this paper to decide how many questions to include in a bank of alternatives. Generally, it is advisable to write banks of questions with as large a number of alternatives as practicable. If a choice has to be made it would be better to write fewer questions with a larger number of alternatives.

## References

Adler, I. and Ross, S.M. (2001), *The coupon subset collection problem*. *Journal of Applied Probability*, **38**, pp 737-746.

Ross, S. M. (2003), *Introduction to Probability Models*, 8<sup>th</sup> edition. Academic Press, New York.

## Appendix A

Matlab code used to produce Figure 1. The infinite series has been truncated to 200 terms.

```
format short
q = 0
while q < 21
r=0
while r < 200
```



```

j = 1;
while j < 11
E(j) = (-1).^(j+1)./factorial(j).*factorial(10)./factorial(10-j).*(1-
j./10).^r;
j = j+1;
end
e = sum(E);
F(r+1) = 1 - (1-e).^q;
r = r+1;
end
numtest(q+1) = sum(F);
Q(q+1)= q;
q = q+1;
end
numtest
A = [Q',numtest']
plot(Q,numtest, 'r','LineWidth', 2)
xlabel('Number of Questions in Test','FontSize',16)
ylabel('Average Number of Tests Required','FontSize',16)

```