Test Construction For Average Score Maximization

Peter M. Ellis, College of Business, Utah State University, Logan, UT, 84322-3510, peter.ellis@usu.edu
Ann Ellis, College of Education, Weber State University, Ogden, UT, 84408-1304, aellis4@weber.edu

ABSTRACT

This work shows how items can be selected from a test bank so that an examination which seeks attainment of a given outcome goal might be constructed. The model presented here develops a general model which is easily adapted to the needs of a particular user. It permits consideration of separate test sections and the required inclusion of a certain number of items from each section. It shows how selection of competing items might be accomplished. It also provides for the generation of parallel tests, so that the identical examination is not used on consecutive test administrations.

INTRODUCTION

We envision the existence of a test bank of a set of items which have been pretested. During test development, item response theory (IRT) is applied in order to predict the probability that an examinee with a specific ability level will correctly answer a given item. A large pool of potential test items is field tested and item responses are analyzed for such indices as difficulty, discrimination power, reliability and validity. Items not meeting desired values are eliminated. Tests are then constructed by selecting items from the remaining pool.

For any item there might exist differing percentages of correct responses between participants who receive differing overall test scores. There will be a cohort of high achievers, a cohort of average achievers and another cohort of low achievers. For example, the high achieving group might be the top quartile, the middle achievers would come from the second and third quartiles and the low achievers would then come from the lowest quartile. One simple parameter is the index of discrimination **D**. It is determined as

$$\mathbf{D} = p_u - p_l \tag{1}$$

where p_u is the proportion in the upper group who answer the item correctly and p_l is the proportion of the lowest group who answer the item correctly.

This work will provide for the existence of C cohorts, and the numerical example will set C = 3. Also, the model includes K test sections. These might be thought of as different topics being examined. In the application shown here K was set to 3. For any test item i the average percentage of correct response from cohort c is given as D_{ic}. This value will be available because of the pretesting of the items in the test bank.

The Mathematical Model

Several variables and parameters are needed in the model. They are:

D_{ics} = average correct response percentage on item i from cohort c in section s of the test Ns = number of items to include from section s Ns = specified value of Ns ns = number of items available in the test bank for section s X_{is} = 1 if item i in section s is of the test bank is included, and = 0 if not C = number of cohorts to include S = number of test sections to have m = number of items from a restricted set that may be permitted for inclusion in the test S_{cs} = aggregate score from cohort c in section s

Because several objectives are possible, the general objective function of the formulation will be designated as f(X), where X is the vector of the X_{is} values. With that, the general mixed bivalent integer formulation of the item selection problem is:

Problem (1)

$$\text{maximize or minimize } Z = f(\underline{X}) \qquad \text{(Ia)}$$

subject to:

$$\sum_{i=1}^{ns} X_{is} = NS \qquad \text{(Ib)}$$

$$\sum_{i=1}^{ns} D_{ics}X_{is} - S_{cs} = 0 \qquad \text{(Ic)}$$

$$Ns = \underline{Ns} \qquad \text{(Id)}$$

$$\text{all } Ns,\ S_{cs} \geq 0,\ \text{all } X_{is} = \{0,1\}$$

Constraints (Ib) establish the number of items that are to be chosen from each of the S test sections. The constraints of (Ic) calculate the expected total score for each cohort c. The limits provided in (Id) dictate the number of items that are to come from each section c. Note that the total number of items included on the test is the sum of the Ns values. Two testing goals will be portrayed here. The first one seeks the maximization of the average test score for all participants. The second seeks to maximize the spread in expected scores highest and lowest cohorts.

An Example Of The Model

Let there exist a test bank with 99 questions. Further, there are S = 3 test sections and C = 3 cohort groups. These cohort groups represent the top, average and low achievers. Each test item has been pretested so that expected percentages of correct scores D_{ics} on each question i in section s by cohort c is known. There are 33 items in each of the three test sections and the test is to use exactly 10 questions from each section (N1 + N2 + N3 = 10 + 10 + 10 = 30). A Monte Carlo simulation has been used to generate the D_{ics} values. For any item then expected percentage correct for cohort 3 was generated randomly with the formula D3 = INT(100*(RN +(.6)*(1-RN))), where RN is a uniformly distributed pseudo-random number on (0,1). For cohort 2 the expected correct percentage is D2 = INT(100*(.9*D3 + RN*(D3-.9*D3))), and D1 is generated in the same way. The values obtained from the simulation are the ones found in constraints 5) – 13) of listings 1 and 3.

Average Test Score Maximization

To formulate this problem in the framework of integer programming, for cohort i let the expected score on test section j be given as Sij. The number of items available in section i is N_i. Then the expected aggregate score for cohort i is

$$S_i = \sum_{j=1}^{Ni} S_{ij} \quad \text{(Ie)}$$

The number of cohorts is C. Let C_i be the percentage of cohort j in the entire student population. The average score for the entire population is then given as

$$\sum_{i=1}^{C} C_i S_i \quad \text{(Ia1)}$$

The complete formulation of the problem thus becomes the maximization of (Ia1), subject to the constraints given in (Ia) – (Id), as well as the new equations in (Ie). Finally, the same nonnegativity and integrality restrictions are retained. An example is given here. Let there again be three cohorts and three test sections. The percentages of each cohort in the general student population are, respectively, 40%, 40% and 20%. Listing 3 presents the complete integer programming formulation for this example. Listing 4 then shows the optimal solution to this problem. The aggregate student average score is maximized by including items 5, 8, 9, 15, 16, 19, 20, 21, 26 and 29 from section 1. From section 2 the included items are 3, 4, 6, 7, 18, 19, 21, 26, 29 and 30. From section 3 the selected items are 1, 3, 5, 7, 8, 10, 11, 14, 17, 19 and 21. The resulting cohort aggregate expected scores are S1 = 2442, S2 = 2593 and S3 = 2758.

Interested readers are invited to contact the authors to receive the complete article, including references, tables and computer output.