# Creating Parallel Forms to Support On-Demand Testing for Undergraduate Students in Psychology

Mark J. GIERL*        Lia DANIELS**        Xinxin ZHANG***

**Abstract**

On-demand testing requires that multiple forms of an exam should be administered to students in each testing session. However, the use of multiple forms raises test security concern because of item exposure. One way to limit exposure is using parallel forms construction. Parallel forms are different versions of a test that measure the same content areas and have the same difficulty level but contain different sets of items. The purpose of this study is to describe and demonstrate how parallel forms can be created from a small item bank using the selected-response item type. We present three unique yet plausible test assembly problems. We also provide a solution for each problem using the results from a free, open-source, software add-in for Microsoft Excel called the Opensolver. Implications for test design and item development are discussed.

*Key Words:* Test development, automated test assembly, fairness in testing

## INTRODUCTION

The principles and practices that guide the design and development of educational tests are undergoing dramatic changes. One of the major catalysts for this change stems from the application of technology in assessment that is best exemplified with the rapid development and application of computer-based testing (CBT) (Drasgow, 2016; Drasgow, Luecht, & Bennett, 2006; Luecht, 2016; Sireci & Zenisky, 2016). A computer-based test is an exam that contains digitally-formatted items that are delivered with a computer. The test administration can be conducted using a computer network or over the Internet. Computer-based tests may contain traditional selected-response (e.g., multiple choice) and constructed-response (e.g., short answer, essays) item types. They may also contain new "innovative" item types—such as drag-and-drop, reordering, and multiple select—that are only made possible with a computer administration because they require technology-based features such as interactivity or multimedia (Sireci & Zenisky, 2016). In short, CBT is dramatically changing educational assessment because the use of expanded item types combined with the growing popularity of digital media and the explosion in Internet use is creating the foundation for a new type of testing system. As a result, many educational tests that were once given in a paper format are now administered by computer as either computer-based or computer adaptive exams. Many common and well-known exams can be cited as examples including ACT Aspire, the Graduate Management Achievement Test, the Graduate Record Exam, the Test of English as a Foreign Language, the American Institute of Certified Public Accountants Uniform CPA examination, the Medical Council of Canada Qualifying Exam Part I, the National Council Licensure Examination for Registered Nurses, and the Canadian Practical Nurse Registration Examination.

### Benefits and Challenges Related to Computer-Based Testing

CBT is growing in popularity not just with large-scale testing companies but also with individual instructors who recognize that paper-based testing is an exceedingly time- and resource-intensive

* Professor of Educational Psychology and Canada Research Chair, University of Alberta, Faculty of Education, Edmonton-Canada, e-mail: mark.gierl@ualberta.ca , ORCID ID: 0000-0002-2653-1761
** Associate Professor of Educational Psychology, University of Alberta, Faculty of Education, Edmonton-Canada, e-mail:lia.daniels@ualberta.ca, ORCID ID: 0000-0001-9202-2538
*** Doctoral Student, University of Alberta, Faculty of Education, Edmonton-Canada, xinxin4@ualberta.ca:, ORCID ID: 0000-0003-4926-7980

process. The printing, scoring, and reporting of paper-based tests require tremendous efforts, expenses, and human interventions. Moreover, as the demand for testing continues to escalate, the cost of developing, administering, and scoring paper-based tests will also continue to increase. One solution that curtails some of these costs is to adopt a CBT system. By administering tests on computers, instructors are liberated from performing the costly and time-consuming administration processes associated with disseminating, scanning, and scoring paper-based tests.

CBT offers many benefits to students and instructors compared to more traditional paper-based testing. Computers permit testing on-demand thereby allowing students to take the exam using a flexible test administration schedule. This benefit means that students can take exams at different locations due to the flexibility of networked computer and Internet access. It also means that instructors can administer tests outside of their scheduled lecture times so that no classes are lost due to a required test administration. Selected-response item types like multiple-choice are scored immediately by the computer. As a result, instructors are not required to manually grade the exams or send the exams to an external facility for optical scoring. If permitted by the instructor, students can even be provided with feedback on their performance immediately upon the completion of the exam because all exams are scored by the computer (Daniels & Gierl, in press). Computer-based tests can be accessed with different types of devices ranging from mobile technologies such as tablets to standalone desktop computers typically found in computer labs or testing centres. Computers also support the development of "innovative" item types that allows teachers to measure more complex performances as well as a broader variety of knowledge and skills using diverse interactive item formats.

Despite these important benefits, two major challenges may prevent instructors initially from attempting to implement CBT. First, instructors may feel that their item bank is too small. A bank is a repository of test items. It contains information about content areas measured by each item as well as statistical information (e.g., difficulty level) about the performance of each item. The items in this repository are used to create forms. On-demanding testing requires that multiple versions or forms of the exam are administered to students in each testing session. Students are then randomly assigned one of the forms. Second, instructors may be concerned that on-demand CBT compromises the security of the exam by allowing students to see the items. These exposed items, in turn, could be disclosed to other students who take a different form or take the exam at a different time. One way to effectively mitigate this concern is to create parallel test forms. Parallel forms are different versions of the test that measure the same content areas and have the same difficulty level but contain a different set of test items (van der Linden, 1998, 2006). The purpose of our study is to describe and demonstrate how instructors can create and implement parallel forms construction using selected-response item types like multiple-choice items. Parallel test forms are considered to be secure exams because each form contains a different set of items thereby minimizing item exposure and maintaining test security so the test administration is fair and equitable for all students[1]. Because parallel forms measure the same content areas and produce tests with the same difficulty levels, instructors can compare students' test results because the scores across the forms are deemed to be equivalent. We begin by providing a description of the construction process using the selected-response item type. Then, we present software that can used to easily implement the construction process using a small bank of test items.

### Test Form Assembly

Producing parallel forms that yield both reliable and valid test scores is complicated because a complex combinatorial problem must be solved. Using a manual test assembly approach, the instructor would first identify the content areas measured by the test, use either the statistical item analysis results from previously administered test items or attempt to predict the expected difficulty

_____

[1]To further enhance security, the order of the items on the parallel forms can also be randomized so students who write the same form receive the same items, but in a different order.

level of the items, and, finally, create multiple forms that each contain items that measure the same content areas and displayed the same item difficulty values. Specifying an item that measures the same content area and displays the same difficulty level from a pool of possible items is a form of constraint programming. As tests increase in specificity with the inclusion of a larger number of content areas and as the number of items increases producing a broader range of item difficulty levels, more constraints are required to produce parallel forms. In addition, multiple content specifications may be required to create truly parallel forms including the use of constraints for variables such as test length, item format, item exposure, date of creation, and source of item. The use of these additional content specifications makes the test assembly process a daunting and, potentially, impossible one to solve using a manual process because items must be selected to meet a statistical requirement while also satisfying two or more content specifications.

Fortunately, efficient computer-based procedures for automated test assembly (ATA) have been developed (see van der Linden, 1998, 2005). Using the optimization algorithms found in a range of readily available software, instructors can quickly solve complex test assembly problems in order to produce parallel test forms, even with relatively small item banks. To date, however, the widespread use of ATA techniques, particularly among university instructors, has been limited. This limitation may be attributed to the instructors lack of knowledge about ATA or it may be attributed to their lack of confidence for implementing the ATA process. The purpose of our study is to address these limitations by describing and illustrating how ATA methods can be used by instructors in psychology to create parallel test forms.

## Overview of Automated Test Assembly

ATA requires the optimization of a test attribute (e.g., overall mean test score) using a unique combination of items so that a feasible solution (e.g., item combinations that meet the content specifications) is produced. The problem of selecting items to meet predefined test specifications can be approached using computerized combinatorial optimization methods (Breithaupt & Hare, 2016, Luecht, 1998; van der Linden, 1998, 2005). Optimization requires the specification of a mathematical model to describe the combinatorial problem. One approach for articulating these models is to specify a system of linear equations that defines the decision variables, an objective function, and the constraints. The system of linear equations contains both integer and string variables. Integers (i.e., discrete variables) are used to describe item attributes such as content area, test length, and item format. Strings (i.e., continuous variables) are used to describe statistical attributes such as mean test score. Once a model is defined, mixed-integer linear programming methods are used to iteratively assess every possible solution relative to the target until the optimal or best combination is identified.

## Program Description with Example

Instructors can solve test assembly problems using optimization algorithms found in specialized software such as IBM ILOG CPLEX, LINGO 16.0, or Premium Solver Pro. In our study, the Opensolver (https://opensolver.org/) add-in was used because it is a freely available, open-source tool than can be run within Microsoft Excel to solve mixed integer optimization problems. Opensolver was created and is supported by Andrew Mason and Iain Dunning in the Department of Engineering Science at the University of Auckland.

Two steps are required to assemble parallel forms using Opensolver. The first step is to describe the test assembly problem as a mathematical model in an Excel spreadsheet. The second step is to specify and solve the test assembly model in the Solver Model interface. To begin, we provide a simple test assembly problem with simulated data from the selected-response item type to illustrate how an ATA problem is defined and specified using the Opensolver. In problem 1, two parallel forms containing 35 selected-response items were assembled from a hypothetical item bank containing 100 items. The bank contained the content code and the difficulty level for each item.

_____

ISSN: 1309 – 6575  _Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi_
_Journal of Measurement and Evaluation in Education and Psychology_

290

Every time an instructor administers a multiple-choice exam, statistics can be computed for each item including the difficulty level. The difficulty level is computed as the proportion of students who answered the item correctly. It ranges from 0.00 to 1.00. Both forms in problem 1 were assembled to meet the same content and statistical specifications. Five content areas were measured on each test. Each form of the exam required 35 items from five different content areas—10 items from content area A, eight items from area B, four items from area C, six items from area D, and seven items from area E. The two forms were also required to be statistically parallel meaning that the items on each form must produce the same mean test score. The requirements for this test assembly problem are summarized in Table 1. Manually selecting two sets of 35 mutually exclusive items from a bank of 100 items that produce the same mean test score across the same five content areas is not a simple task. As an alternative, OpenSolver can be used to solve this test assemble problem very quickly.

Table 1. Summary of Content and Statistical Requirements for Problem 1

| Content Area | Number of Items in Bank | Difficulty Mean (SD) | Number of Items Per Form |
|---|---|---|---|
| A | 20 | 0.70 (0.15) | 10 |
| B | 20 | 0.65 (0.11) | 8 |
| C | 20 | 0.65 (0.15) | 4 |
| D | 20 | 0.74 (0.17) | 6 |
| E | 20 | 0.72 (0.23) | 7 |
| Total | 100 | 0.69 (0.16) | 35 |

*Step 1: Setting up the mathematical model in Excel*

To begin, the content and statistical data for each item in the bank must be specified in an Excel spreadsheet, as shown in the item bank table in Figure 1 (this figure displays 33 of the 100 items). The bank contains the item number, content area, and difficulty level for each item. For example, Cell B3, C3 and D3 indicate item 1 belongs to content A and it has a difficulty level of 0.85 (very easy item). Next, the decision variables (which includes the form difficulty and item overlap tables), the content constraints, and the objective function of the ATA problem must be specified. The decision variables table is used to specify the ATA model for the problem. These cells represent the item decision variable matrix (i.e., item-by-test form). The cells contain the values that the solver algorithm begins with and then updates in order to search for an optimal solution. In our example, the cells are constrained to be binary so that a 1 means that an item was selected for the required form and a 0 means it was not selected. The starting values for these cells are 0, which means that the items were not selected. In step 2 where we specify and solve the test assembly model (shown in next section), these values will change thereby describing the solution for the ATA problem (i.e., Figure 1 and 3 are used together to describe the initial state and the final solution to our problem). For example, Cell F3=0 means that item 1 was not selected for form 1. Cell G4=0 means that item 2 was not selected for form 2. The form difficulty and item overlap tables are defined using the same logic. The starting point for form difficulty is that the difficulty level of each item begins with the value of 0.00. We also define item overlap. In problem 1, we do not want any overlap among the items across the two forms (i.e., the forms will each contain unique mutually exclusive items from the bank). We define item overlap by specifying the starting point for the constraint to be 1 (column M) and the form to be 0 (column N).

| | Item# | content area | Item difficulty | Decision Variables | form 1 | form 2 | Form Difficulty | dif*form 1 | dif*form 2 | Overlap Constraints | Item# | constraints | forms | Content Constrains | content area | constraints | form 1 | form 2 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Item Bank | | | | | | | | | | | | | | | | | | |
| | 1 | A | 0.85 | | 0 | 0 | | 0.00 | 0.00 | | 1 | 1 | 0 | | A | 10 | 0 | 0 |
| | 2 | A | 0.84 | | 0 | 0 | | 0.00 | 0.00 | | 2 | 1 | 0 | | B | 8 | 0 | 0 |
| | 3 | A | 0.85 | | 0 | 0 | | 0.00 | 0.00 | | 3 | 1 | 0 | | C | 4 | 0 | 0 |
| | 4 | A | 0.84 | | 0 | 0 | | 0.00 | 0.00 | | 4 | 1 | 0 | | D | 6 | 0 | 0 |
| | 5 | A | 0.57 | | 0 | 0 | | 0.00 | 0.00 | | 5 | 1 | 0 | | E | 7 | 0 | 0 |
| | 6 | A | 0.61 | | 0 | 0 | | 0.00 | 0.00 | | 6 | 1 | 0 | | | | | |
| | 7 | A | 0.43 | | 0 | 0 | | 0.00 | 0.00 | | 7 | 1 | 0 | | | | | |
| | 8 | A | 0.57 | | 0 | 0 | | 0.00 | 0.00 | | 8 | 1 | 0 | Objective | | | | |
| | 9 | A | 0.61 | | 0 | 0 | | 0.00 | 0.00 | | 9 | 1 | 0 | Function | | | form 1 | form 2 |
| | 10 | A | 0.82 | | 0 | 0 | | 0.00 | 0.00 | | 10 | 1 | 0 | | Mean Test Scores | | 0.00 | 0.00 |
| | 11 | A | 0.85 | | 0 | 0 | | 0.00 | 0.00 | | 11 | 1 | 0 | | Target | | 0.00 | |
| | 12 | A | 0.84 | | 0 | 0 | | 0.00 | 0.00 | | 12 | 1 | 0 | | | | | |
| | 13 | A | 0.85 | | 0 | 0 | | 0.00 | 0.00 | | 13 | 1 | 0 | | | | | |
| | 14 | A | 0.84 | | 0 | 0 | | 0.00 | 0.00 | | 14 | 1 | 0 | | | | | |
| | 15 | A | 0.57 | | 0 | 0 | | 0.00 | 0.00 | | 15 | 1 | 0 | | | | | |
| | 16 | A | 0.61 | | 0 | 0 | | 0.00 | 0.00 | | 16 | 1 | 0 | | | | | |
| | 17 | A | 0.82 | | 0 | 0 | | 0.00 | 0.00 | | 17 | 1 | 0 | | | | | |
| | 18 | A | 0.61 | | 0 | 0 | | 0.00 | 0.00 | | 18 | 1 | 0 | | | | | |
| | 19 | A | 0.43 | | 0 | 0 | | 0.00 | 0.00 | | 19 | 1 | 0 | | | | | |
| | 20 | A | 0.57 | | 0 | 0 | | 0.00 | 0.00 | | 20 | 1 | 0 | | | | | |
| | 21 | B | 0.62 | | 0 | 0 | | 0.00 | 0.00 | | 21 | 1 | 0 | | | | | |
| | 22 | B | 0.57 | | 0 | 0 | | 0.00 | 0.00 | | 22 | 1 | 0 | | | | | |
| | 23 | B | 0.70 | | 0 | 0 | | 0.00 | 0.00 | | 23 | 1 | 0 | | | | | |
| | 24 | B | 0.62 | | 0 | 0 | | 0.00 | 0.00 | | 24 | 1 | 0 | | | | | |
| | 25 | B | 0.83 | | 0 | 0 | | 0.00 | 0.00 | | 25 | 1 | 0 | | | | | |
| | 26 | B | 0.60 | | 0 | 0 | | 0.00 | 0.00 | | 26 | 1 | 0 | | | | | |
| | 27 | B | 0.86 | | 0 | 0 | | 0.00 | 0.00 | | 27 | 1 | 0 | | | | | |
| | 28 | B | 0.53 | | 0 | 0 | | 0.00 | 0.00 | | 28 | 1 | 0 | | | | | |
| | 29 | B | 0.62 | | 0 | 0 | | 0.00 | 0.00 | | 29 | 1 | 0 | | | | | |
| | 30 | B | 0.57 | | 0 | 0 | | 0.00 | 0.00 | | 30 | 1 | 0 | | | | | |
| | 31 | B | 0.62 | | 0 | 0 | | 0.00 | 0.00 | | 31 | 1 | 0 | | | | | |
| | 32 | B | 0.57 | | 0 | 0 | | 0.00 | 0.00 | | 32 | 1 | 0 | | | | | |
| | 33 | B | 0.70 | | 0 | 0 | | 0.00 | 0.00 | | 33 | 1 | 0 | | | | | |

Figure 1. Model Specified for the ATA Problem in Problem 1.

Next, the content constraints table is used to define the content on the parallel forms. The number of items measuring each content category on each test is calculated in the columns (R, S). These cells are defined so they equal the sum of the decision variables assigned to items under each specific category. For instance, cell R3 is equal to sum of F(3:22) because these decision variables are for items measuring content A, which are items 1 to 20 in the bank. The values of the cells in these columns (R, S) must be equal to the values in column Q, as specified in the content constraints table.

Finally, the objective function table is defined. This function specifies the statistical requirement for our problem. In order to make the two forms have the same mean test score, the objective function was formulated to minimize the absolute difference between the mean score on the two forms. Hence the difference between the forms is presented as P13 = 0.00, meaning the forms should have the same mean test score. The mean test score for each form is calculated in Cells Q12 and R12 based on the objective function table which, in turn, is calculated based on the product of the decision variables and form difficulty in columns I and J.

*Step 2: Specifying and solving the test assembly model*

Once the mathematical model is specified in Excel, the Opensolver parameter interface, shown in Figure 2, is used to structure and execute the ATA analysis. The user must specify the objective function by placing the cell containing the objective, Cell $Q$13[2], into the Objective Cell box. Then, the user clicks the appropriate radio button to decide whether the objective function should be maximized (max), minimized (min), or set to a specific target (value of). In our example, we click minimize because the goal is to create two forms with the same mean test score. The decision variables table ($F$3:$G$102[3] from Figure 1) are placed into the Variable Cells box. After the decision variables have been specified, the user adds constraints to the Constraints box using the Add Constraints button on the right side of the interface.

---

[2] A $ in Excel denotes a fixed cell. In this example, $Q$13 means that the fixed value in this cell location is used in the analysis. For problem 1, the value of $Q$13 is 0.00.

3 A colon in Excel describes a range of values. $F$3:$G$102 means all of the fixed values in the range that starts in cell $F$3 and ends in cell G$102.

ISSN: 1309 – 6575 *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*
*Journal of Measurement and Evaluation in Education and Psychology*

292
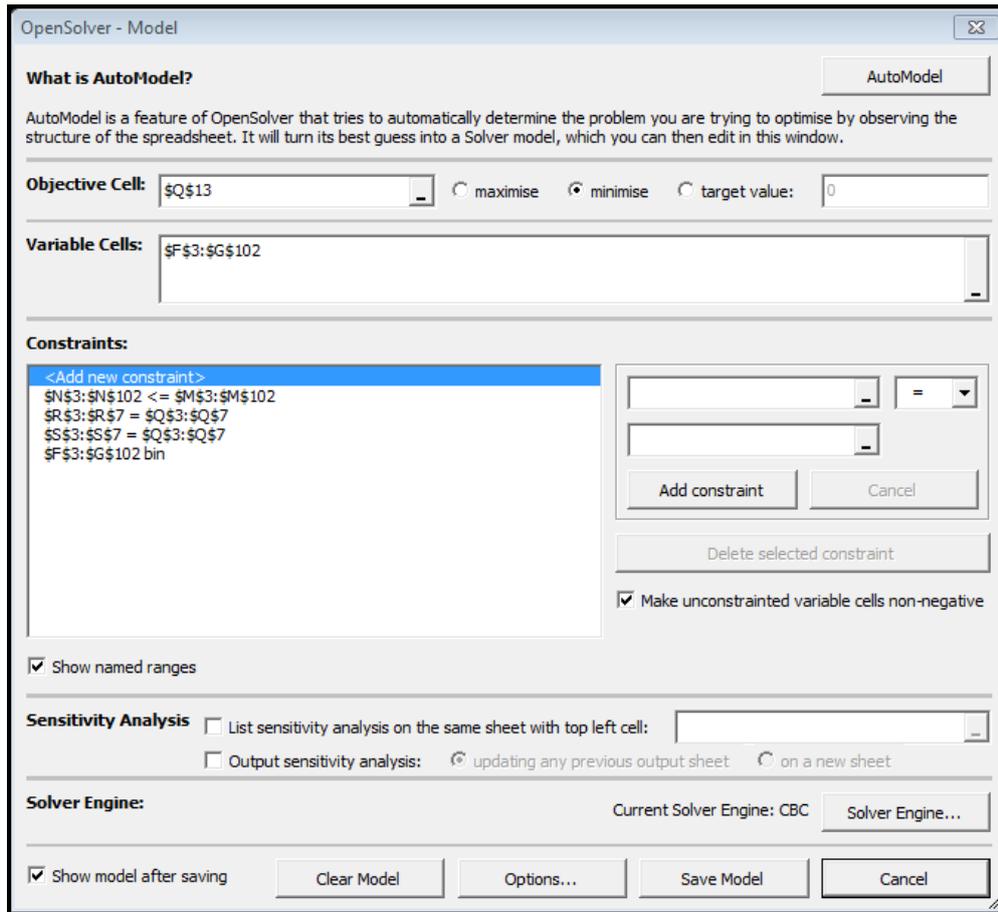
Figure 2. An Example of the Opensolver Parameter Interface.

Problem 1 has three sets of constraints. The first constraint guarantees that items are only used once in each parallel form (i.e., $N$3:$N$102<=$M$3:$M$102 from Figure 1). The second constraint ensures that each form satisfies the content specification (i.e., $R$3:$R$7=$Q$3:$Q$7 and $S$3:$S$7= $Q$3:$M$7 from Figure 1). The third constraint requires the decision variables to be binary (i.e., $F$3:$G$102 bin from Figure 1). The sensitivity analysis provides information about the consequence of changing the objective function for a given constraint and for adjusting a constraint that is currently zero. To ensure our example is relatively simple and straightforward, this optional analysis is not used in our demonstration. Once the model is created, the user clicks the Solver Engine button on the Solver Engine line to run the analysis and solve the optimization problem.

The result from the Opensolver analysis is an Excel spreadsheet that provides a comprehensive summary of the solution, as shown in Figure 3. Recall, Figure 1 defines the ATA problem. Figure 3 presents the solution to the problem specified in Figure 1. Hence, Figures 1 and 3 should be interpreted together. The solution contains the original item bank. The decision variable table defines the assignment of each item to each form. For problem 1, items are assigned to either form 1 or 2. A 1 means the item is assigned to a form and a 0 means the item is not assigned to the form. The form difficulty table identifies the difficulty level for each item on each form. The item overlap table defines the form assignment for the items. For example, items 1 to 24 are assigned to either form 1 or 2. But items 25 and 27 are not selected and hence not assigned to either form in this example (i.e., they receive a 0 in the overlap table). The content constraints table provides a summary of how well each form satisfies the content specification. In our example, all of the content specifications were met. The objective function table provides the mean test score for each

_____

form based on the item statistic information available in the bank. The mean test score in our example for Form 1 and 2 is 0.63 and 0.63, respectively. The means scores are computed from the item difficulty values in the bank. They indicate that the difficulty of the two forms is the same. Hence, the mean test score outcome produced by the ATA algorithm and reported in Figure 3 serves as a summary of the equivalency of the two test forms. The implication of this outcome is important for interpreting student test performance: When form 1 and 2 are randomly administered to two groups of students, the mean score on each forms is expected to be the same (barring some sampling error) because the forms were created and demonstrated to be equivalent to one another. Taken together, the results in Figure 3 identify the items that must be selected from the bank in order to create two forms that satisfy the content and statistical requirements described in our example ATA problem. The results in this figure also reveal that a feasible solution that meets the overlap, content, and statistical requirements was found for problem 1.

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Item | | | | Decision | | | Form | | | Overlap | | | | Content | | | | |
| 2 | Bank | Item# | content area | Item difficulty | Variables | form 1 | form 2 | Difficulty | dif*form 1 | dif*form 2 | Constraints | Item# | constraints | forms | Constrains | content area | constraints | form 1 | form 2 |
| 3 | | 1 | A | 0.85 | | 0 | 1 | | 0.00 | 0.85 | | 1 | 1 | 1 | | A | 10 | 10 | 10 |
| 4 | | 2 | A | 0.84 | | 0 | 1 | | 0.00 | 0.84 | | 2 | 1 | 1 | | B | 8 | 8 | 8 |
| 5 | | 3 | A | 0.85 | | 1 | 0 | | 0.85 | 0.00 | | 3 | 1 | 1 | | C | 4 | 4 | 4 |
| 6 | | 4 | A | 0.84 | | 1 | 0 | | 0.84 | 0.00 | | 4 | 1 | 1 | | D | 6 | 6 | 6 |
| 7 | | 5 | A | 0.57 | | 0 | 1 | | 0.00 | 0.57 | | 5 | 1 | 1 | | E | 7 | 7 | 7 |
| 8 | | 6 | A | 0.61 | | 0 | 1 | | 0.00 | 0.61 | | 6 | 1 | 1 | | | | | |
| 9 | | 7 | A | 0.43 | | 0 | 1 | | 0.00 | 0.43 | | 7 | 1 | 1 | | | | | |
| 10 | | 8 | A | 0.57 | | 1 | 0 | | 0.57 | 0.00 | | 8 | 1 | 1 | Objective | | | | |
| 11 | | 9 | A | 0.61 | | 1 | 0 | | 0.61 | 0.00 | | 9 | 1 | 1 | Function | | | form 1 | form 2 |
| 12 | | 10 | A | 0.82 | | 1 | 0 | | 0.82 | 0.00 | | 10 | 1 | 1 | | Mean Test Scores | | 0.63 | 0.63 |
| 13 | | 11 | A | 0.85 | | 1 | 0 | | 0.85 | 0.00 | | 11 | 1 | 1 | | Target | | 0.00 | |
| 14 | | 12 | A | 0.84 | | 0 | 1 | | 0.00 | 0.84 | | 12 | 1 | 1 | | | | | |
| 15 | | 13 | A | 0.85 | | 0 | 1 | | 0.00 | 0.85 | | 13 | 1 | 1 | | | | | |
| 16 | | 14 | A | 0.84 | | 0 | 1 | | 0.00 | 0.84 | | 14 | 1 | 1 | | | | | |
| 17 | | 15 | A | 0.57 | | 0 | 1 | | 0.00 | 0.57 | | 15 | 1 | 1 | | | | | |
| 18 | | 16 | A | 0.61 | | 1 | 0 | | 0.61 | 0.00 | | 16 | 1 | 1 | | | | | |
| 19 | | 17 | A | 0.82 | | 0 | 1 | | 0.00 | 0.82 | | 17 | 1 | 1 | | | | | |
| 20 | | 18 | A | 0.61 | | 1 | 0 | | 0.61 | 0.00 | | 18 | 1 | 1 | | | | | |
| 21 | | 19 | A | 0.43 | | 1 | 0 | | 0.43 | 0.00 | | 19 | 1 | 1 | | | | | |
| 22 | | 20 | A | 0.57 | | 1 | 0 | | 0.57 | 0.00 | | 20 | 1 | 1 | | | | | |
| 23 | | 21 | B | 0.62 | | 0 | 1 | | 0.00 | 0.62 | | 21 | 1 | 1 | | | | | |
| 24 | | 22 | B | 0.57 | | 1 | 0 | | 0.57 | 0.00 | | 22 | 1 | 1 | | | | | |
| 25 | | 23 | B | 0.70 | | 1 | 0 | | 0.70 | 0.00 | | 23 | 1 | 1 | | | | | |
| 26 | | 24 | B | 0.62 | | 1 | 0 | | 0.62 | 0.00 | | 24 | 1 | 1 | | | | | |
| 27 | | 25 | B | 0.83 | | 0 | 0 | | 0.00 | 0.00 | | 25 | 1 | 0 | | | | | |
| 28 | | 26 | B | 0.60 | | 0 | 1 | | 0.00 | 0.60 | | 26 | 1 | 1 | | | | | |
| 29 | | 27 | B | 0.86 | | 0 | 0 | | 0.00 | 0.00 | | 27 | 1 | 0 | | | | | |
| 30 | | 28 | B | 0.53 | | 1 | 0 | | 0.53 | 0.00 | | 28 | 1 | 1 | | | | | |
| 31 | | 29 | B | 0.62 | | 1 | 0 | | 0.62 | 0.00 | | 29 | 1 | 1 | | | | | |
| 32 | | 30 | B | 0.57 | | 1 | 0 | | 0.57 | 0.00 | | 30 | 1 | 1 | | | | | |
| 33 | | 31 | B | 0.62 | | 1 | 0 | | 0.62 | 0.00 | | 31 | 1 | 1 | | | | | |
| 34 | | 32 | B | 0.57 | | 0 | 1 | | 0.00 | 0.57 | | 32 | 1 | 1 | | | | | |
| 35 | | 33 | B | 0.70 | | 0 | 1 | | 0.00 | 0.70 | | 33 | 1 | 1 | | | | | |

Figure 3. The Solution for Problem 1.


## Two Practical Problems Using Test Data from an Undergraduate Psychology Course

We began with a structured example to illustrate the logic required to build a model and solve a ATA problem using simulated data. Next, we apply this method in two practical testing situations using actual student response data. An item bank containing 144 multiple-choice items was used. The data for the two practical examples in this section of our paper were collected from a midterm and a final multiple-choice exam administered in an introductory undergraduate course focused on adolescent developmental psychology. The content codes for the items were created by the instructor of the course. The item response data were collected from all 162 undergraduate students who completed both exams during the Winter 2016 semester.


## Problem 2: Parallel Forms Construction with Common Overlapping Items

In the second problem, the goal is to create two parallel forms that meet the following requirements: (1) each form should contain 36 items; (2) the forms should have a similar mean test score; (3) the forms must meet five content areas requirements—11 items are from content A, eight items are from

_____

content B, four items are from content C, six items are from content D, seven items are from content E; and (4) the forms must contain some common, overlapping items. Item overlap is included as a constraint in problem 2 because the item bank is relatively small. To address this limitation, a unique set of 20 items that measure key concepts from each content area in the course was first identified by the instructor. These key items were only used once in the test assembly problem to limit their exposure. The remaining items in the bank were free to vary meaning that they could be used in neither form or on either one or both forms. A summary of the item bank and the ATA requirements for problem 2 are presented in Table 2.

Table 2. Summary of Content and Statistical Requirements for Test Assembly Problem 2

| Content Area | Number of Items in Bank | Difficulty Mean (SD) | Number of Items Per Form |
|---|---|---|---|
| A | 14 | 0.75(0.20) | 11 |
| B | 30 | 0.66(0.16) | 8 |
| C | 15 | 0.74(0.17) | 4 |
| D | 53 | 0.72(0.17) | 6 |
| E | 32 | 0.67(0.15) | 7 |
| Total | 144* | 0.71 (0.17) | 36 |

*Items 4, 6, 21, 24, 27, 35, 46, 51, 67, 71, 82, 99, 101, 107, 114, 118, 124, 130, 135 and 140 can only be used in one of the two forms.

The mathematical model for problem 2 is presented in Figure 4 (this figure shows 32 out of 144 items). The structure of problem 2 is comparable to problem 1, except the overlap constraints are define for a specific number of items and the proportion of items across the five content areas is defined to satisfy the problem requirements.



Figure 4. Model Specified for the ATA Problem in Problem 2.

Figure 5 contains the Opensolver parameter interface. This interface contains the input required to structure and execute the ATA analysis. It includes a definition for the objective function, the decision variables, and the user-defined constraints.

_____

ISSN: 1309 – 6575  *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*
*Journal of Measurement and Evaluation in Education and Psychology*

295
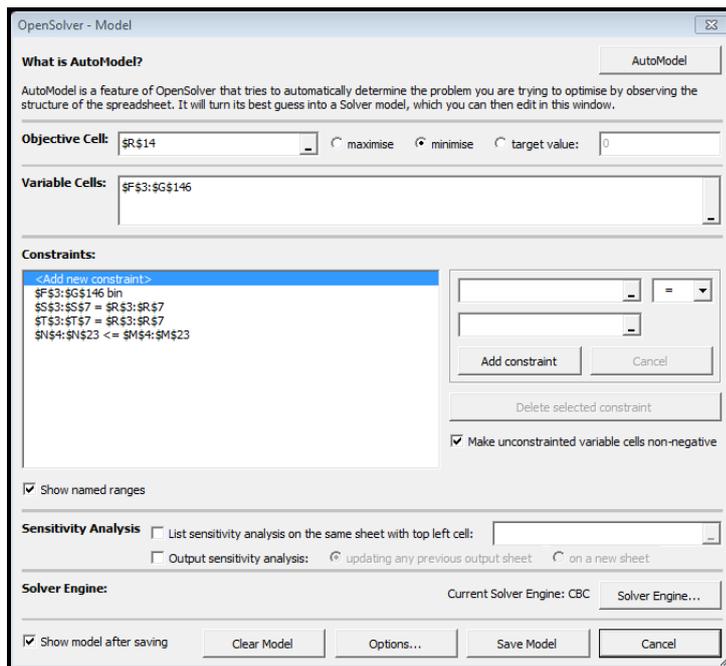
Figure 5. The Opensolver Parameter Interface for Problem 2.

The solution for problem 2 is presented in Figure 6. The solution contains the original item bank with items 1 to 144 (this figure shows 32 out of 144 items). The decision variable table defines the assignment of each item to each form. The form difficulty table identifies the difficulty level for each item. The item overlap table defines the form assignment for the items. It also identifies which items will and will not be used from the bank. The content constraints table provides a summary of how well each form satisfies the content specification. The objective function table provides the mean test score for each form. The mean test score for Form 1 and 2 is 0.55 and 0.54, respectively, using the student response data available in the bank. In other words, for problem 2, the best solution that could be produced given the required constraints is that the mean score is very similar, but not identical, across the two forms (difference between the mean scores across the two forms, as reported in Objective Function summary is 0.01). Form 2 is more difficult than form 1 by 1 score point.

_____

ISSN: 1309 – 6575   _Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi_
_Journal of Measurement and Evaluation in Education and Psychology_                                    296

| Item Bank | Item# | content | Item difficulty | Decision Variables | form 1 | form 2 | Form Difficulty | dif*form 1 | dif*form 2 | Overlap Constraints | Item# | constraints | forms |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | A | 0.88 | | 1 | 1 | | 0.88 | 0.88 | | | | |
| | 2 | A | 0.36 | | 1 | 1 | | 0.36 | 0.36 | | 4.00 | 1.00 | 1.00 |
| | 3 | A | 0.95 | | 0 | 1 | | 0.00 | 0.95 | | 6.00 | 1.00 | 1.00 |
| | 4 | A | 0.94 | | 1 | 0 | | 0.94 | 0.00 | | 21.00 | 1.00 | 0.00 |
| | 5 | A | 0.87 | | 1 | 1 | | 0.87 | 0.87 | | 24.00 | 1.00 | 1.00 |
| | 6 | A | 0.79 | | 0 | 1 | | 0.00 | 0.79 | | 27.00 | 1.00 | 0.00 |
| | 7 | A | 0.63 | | 1 | 1 | | 0.63 | 0.63 | | 35.00 | 1.00 | 1.00 |
| | 8 | A | 0.57 | | 1 | 1 | | 0.57 | 0.57 | | 46.00 | 1.00 | 1.00 |
| | 9 | A | 0.61 | | 1 | 1 | | 0.61 | 0.61 | | 51.00 | 1.00 | 1.00 |
| | 10 | A | 0.97 | | 0 | 0 | | 0.00 | 0.00 | | 67.00 | 1.00 | 0.00 |
| | 11 | A | 0.77 | | 1 | 1 | | 0.77 | 0.77 | | 71.00 | 1.00 | 0.00 |
| | 12 | A | 0.44 | | 1 | 1 | | 0.44 | 0.44 | | 82.00 | 1.00 | 0.00 |
| | 13 | A | 0.81 | | 1 | 1 | | 0.81 | 0.81 | | 99.00 | 1.00 | 0.00 |
| | 14 | A | 0.95 | | 1 | 0 | | 0.95 | 0.00 | | 101.00 | 1.00 | 0.00 |
| | 15 | B | 0.42 | | 1 | 1 | | 0.42 | 0.42 | | 107.00 | 1.00 | 0.00 |
| | 16 | B | 0.74 | | 0 | 0 | | 0.00 | 0.00 | | 114.00 | 1.00 | 0.00 |
| | 17 | B | 0.56 | | 0 | 0 | | 0.00 | 0.00 | | 118.00 | 1.00 | 0.00 |
| | 18 | B | 0.56 | | 1 | 1 | | 0.56 | 0.56 | | 124.00 | 1.00 | 0.00 |
| | 19 | B | 0.84 | | 0 | 0 | | 0.00 | 0.00 | | 130.00 | 1.00 | 0.00 |
| | 20 | B | 0.67 | | 0 | 0 | | 0.00 | 0.00 | | 135.00 | 1.00 | 0.00 |
| | 21 | B | 0.62 | | 0 | 0 | | 0.00 | 0.00 | | 140.00 | 1.00 | 0.00 |
| | 22 | B | 0.57 | | 0 | 0 | | 0.00 | 0.00 | | | | |
| | 23 | B | 0.70 | | 0 | 0 | | 0.00 | 0.00 | | | | |
| | 24 | B | 0.52 | | 1 | 0 | | 0.52 | 0.00 | | | | |
| | 25 | B | 0.83 | | 0 | 0 | | 0.00 | 0.00 | | | | |
| | 26 | B | 0.60 | | 0 | 0 | | 0.00 | 0.00 | | | | |
| | 27 | B | 0.86 | | 0 | 0 | | 0.00 | 0.00 | | | | |
| | 28 | B | 0.53 | | 1 | 1 | | 0.53 | 0.53 | | | | |
| | 29 | B | 0.75 | | 0 | 0 | | 0.00 | 0.00 | | | | |
| | 30 | B | 0.86 | | 0 | 0 | | 0.00 | 0.00 | | | | |
| | 31 | B | 0.78 | | 0 | 0 | | 0.00 | 0.00 | | | | |
| | 32 | B | 0.48 | | 1 | 1 | | 0.48 | 0.48 | | | | |

Content Constraints

| content | Constraints | form 1 | form 2 |
|---|---|---|---|
| A | 11 | 11 | 11 |
| B | 8 | 8 | 8 |
| C | 4 | 4 | 4 |
| D | 6 | 6 | 6 |
| E | 7 | 7 | 7 |

Objective Function

| | form 1 | form 2 |
|---|---|---|
| Test difficulty | 0.55 | 0.54 |
| Target | 0.01 | |

Figure 6. The Solution for Problem 2.

*Problem 3: Parallel forms construction with practice test items*

In the third problem, the goal is to create two parallel forms that meet the following requirements: (1) each form should contain 36 items; (2) the forms should have a similar mean test score; (3) the forms must meet five content areas requirements—11 items are from content A, eight items are from content B, four items are from content C, six items are from content D, seven items are from content E; (4) the forms must contain a specific set of common items (i.e., the two forms contain some overlapping test items), and (5) the forms include nine new items that do not contain item statistics. The fifth constraint is included because the instructor has written nine new items and these items have not been administered to students. Hence, the purpose is to include these new items across the two forms so the final forms satisfy the content and overlap constraints with the added benefit of collecting statistical item analysis data on the new items so they can be used in future test assembly tasks. Constraint 5 is often included in ATA problems when the long-term goal is to increase the size of the item bank. The same bank of 144 test items from problem 2 was used in problem 3, but with the addition of nine new items to produce a bank of 153 items. A summary of the item bank along with the ATA requirements are presented in Table 3.

Table 3. Summary of Content and Statistical Requirements for Test Assembly Problem 3

| Content Area | Number of Items in Bank | Difficulty Mean (SD) | Number of Items Per Form |
|---|---|---|---|
| A | 17(14+3 new) | 0.75 (0.20) | 11 |
| B | 30 (30 + 0 new) | 0.66 (0.16) | 8 |
| C | 17(15+2 new) | 0.74 (0.17) | 4 |
| D | 55(53+2 new) | 0.72 (0.17) | 6 |
| E | 34(32+ 2 new) | 0.67 (0.15) | 7 |
| Total | 153* | 0.71 (0.17) | 36 |

*Items 4, 6, 21, 24, 27, 35, 46, 51, 67, 71, 82, 99, 101, 107, 114, 118, 124, 130, 135 and 140 can only be used in one of the two forms.

ISSN: 1309 – 6575   *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*
*Journal of Measurement and Evaluation in Education and Psychology*

297

The model for problem 3 is presented in Figure 7.  The structure of problem 3 is similar to problem 2, except we add nine new items which have no difficulty statistics (e.g., items 15-17 are new therefore difficulty level is blank).

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Item | | | | Decision | | | Form | | | Content | | | | | |
| 2 | Bank | Item# | content | Item difficulty | Variables | form 1 | form 2 | Difficulty | dif*form 1 | dif*form 2 | Constrains | content | Constraints | form 1 | form 2 | |
| 3 | | 1 | A | 0.88 | | 0 | 0 | | 0.00 | 0.00 | | A | 11 | 0 | 0 | |
| 4 | | 2 | A | 0.36 | | 0 | 0 | | 0.00 | 0.00 | | B | 8 | 0 | 0 | |
| 5 | | 3 | A | 0.95 | | 0 | 0 | | 0.00 | 0.00 | | C | 4 | 0 | 0 | |
| 6 | | 4 | A | 0.94 | | 0 | 0 | | 0.00 | 0.00 | | D | 6 | 0 | 0 | |
| 7 | | 5 | A | 0.87 | | 0 | 0 | | 0.00 | 0.00 | | E | 7 | 0 | 0 | |
| 8 | | 6 | A | 0.79 | | 0 | 0 | | 0.00 | 0.00 | Overlap | | | | | |
| 9 | | 7 | A | 0.63 | | 0 | 0 | | 0.00 | 0.00 | Constraints | | | | | |
| 10 | | 8 | A | 0.57 | | 0 | 0 | | 0.00 | 0.00 | | Item# | constraints | forms | | |
| 11 | | 9 | A | 0.61 | | 0 | 0 | | 0.00 | 0.00 | | 4 | 1 | 0 | | |
| 12 | | 10 | A | 0.97 | | 0 | 0 | | 0.00 | 0.00 | | 6 | 1 | 0 | | |
| 13 | | 11 | A | 0.77 | | 0 | 0 | | 0.00 | 0.00 | | 21 | 1 | 0 | | |
| 14 | | 12 | A | 0.44 | | 0 | 0 | | 0.00 | 0.00 | | 24 | 1 | 0 | | |
| 15 | | 13 | A | 0.81 | | 0 | 0 | | 0.00 | 0.00 | | 27 | 1 | 0 | | |
| 16 | | 14 | A | 0.95 | | 0 | 0 | | 0.00 | 0.00 | | 35 | 1 | 0 | | |
| 17 | | 15 | A | | | 0 | 0 | | | | | 46 | 1 | 0 | | |
| 18 | | 16 | A | | | 0 | 0 | | | | | 51 | 1 | 0 | | |
| 19 | | 17 | A | | | 0 | 0 | | | | | 67 | 1 | 0 | | |
| 20 | | 18 | B | 0.42 | | 0 | 0 | | 0.00 | 0.00 | | 71 | 1 | 0 | | |
| 21 | | 19 | B | 0.74 | | 0 | 0 | | 0.00 | 0.00 | | 82 | 1 | 0 | | |
| 22 | | 20 | B | 0.56 | | 0 | 0 | | 0.00 | 0.00 | | 99 | 1 | 0 | | |
| 23 | | 21 | B | 0.56 | | 0 | 0 | | 0.00 | 0.00 | | 101 | 1 | 0 | | |
| 24 | | 22 | B | 0.84 | | 0 | 0 | | 0.00 | 0.00 | | 107 | 1 | 0 | | |
| 25 | | 23 | B | 0.67 | | 0 | 0 | | 0.00 | 0.00 | | 114 | 1 | 0 | | |
| 26 | | 24 | B | 0.62 | | 0 | 0 | | 0.00 | 0.00 | | 118 | 1 | 0 | | |
| 27 | | 25 | B | 0.57 | | 0 | 0 | | 0.00 | 0.00 | | 124 | 1 | 0 | | |
| 28 | | 26 | B | 0.70 | | 0 | 0 | | 0.00 | 0.00 | | 130 | 1 | 0 | | |
| 29 | | 27 | B | 0.52 | | 0 | 0 | | 0.00 | 0.00 | | 135 | 1 | 0 | | |
| 30 | | 28 | B | 0.83 | | 0 | 0 | | 0.00 | 0.00 | | 140 | 1 | 0 | | |
| 31 | | 29 | B | 0.60 | | 0 | 0 | | 0.00 | 0.00 | | | | | | |
| 32 | | 30 | B | 0.86 | | 0 | 0 | | 0.00 | 0.00 | | | | | | |
| 33 | | 31 | B | 0.53 | | 0 | 0 | | 0.00 | 0.00 | | | | | | |
| 34 | | 32 | B | 0.75 | | 0 | 0 | | 0.00 | 0.00 | | | | | | |
| 35 | | 33 | B | 0.86 | | 0 | 0 | | 0.00 | 0.00 | Objective | | | | | |
| 36 | | 34 | B | 0.78 | | 0 | 0 | | 0.00 | 0.00 | Function | | Form1 | Form2 | | |
| 37 | | 35 | B | 0.48 | | 0 | 0 | | 0.00 | 0.00 | | Mean Test Scores | 0.00 | 0.00 | | |
| 38 | | 36 | B | 0.94 | | 0 | 0 | | 0.00 | 0.00 | | Target | 0.00 | | | |
| 39 | | 37 | B | 0.91 | | 0 | 0 | | 0.00 | 0.00 | | | | | | |

Figure 7. Model Specified for the ATA Problem in Problem 3.

Figure 8 contains the Opensolver parameter interface.  This interface contains the input required to structure and execute the ATA analysis.  It includes a definition for the objective function, the decision variables, and the user-defined constraints.

_____

ISSN: 1309 – 6575   *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*
*Journal of Measurement and Evaluation in Education and Psychology*

298
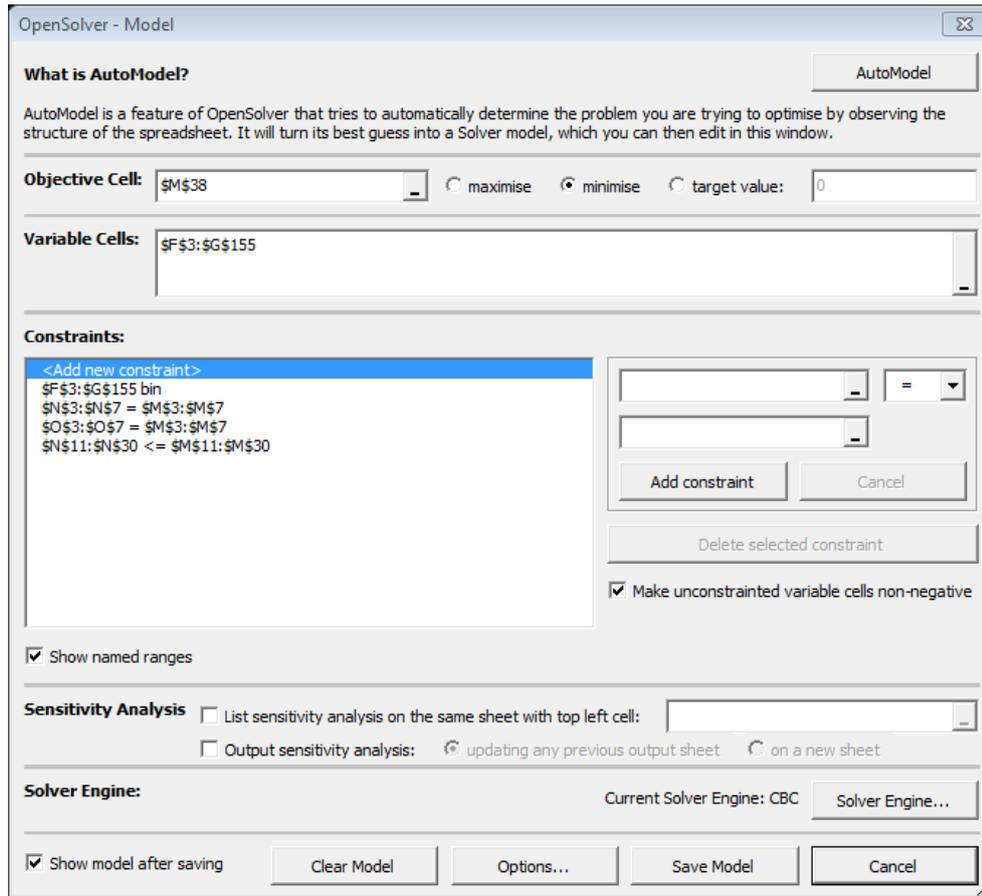
Figure 8.  The Opensolver Parameter Interface for Problem 3.

The solution for problem 3 is presented in Figure 9.  The solution contains the original item bank as well as the decision variable, the form difficulty, the item overlap, the content constraints, and the objective function tables.  For problem 3, the mean test score for Form 1 and 2 is 0.52 and 0.52, respectively, using the student response data available in the bank.  The decision variable table specifies the items to include in each form so the content constrain is satisfied.  These forms will contain both the original and the new items.

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Item | | | | Decision | | | Form | | | Content | | | | |
| 2 | Bank | Item# | content | Item difficulty | Variables | form 1 | form 2 | Difficulty | dif*form 1 | dif*form 2 | Constrains | content | Constraints | form 1 | form 2 |
| 3 | | 1 | A | 0.88 | | 1 | 0 | | 0.88 | 0.00 | | A | 11 | 11 | 11 |
| 4 | | 2 | A | 0.36 | | 1 | 1 | | 0.36 | 0.36 | | B | 8 | 8 | 8 |
| 5 | | 3 | A | 0.95 | | 0 | 0 | | 0.00 | 0.00 | | C | 4 | 4 | 4 |
| 6 | | 4 | A | 0.94 | | 0 | 0 | | 0.00 | 0.00 | | D | 6 | 6 | 6 |
| 7 | | 5 | A | 0.87 | | 1 | 1 | | 0.87 | 0.87 | | E | 7 | 7 | 7 |
| 8 | | 6 | A | 0.79 | | 0 | 1 | | 0.00 | 0.79 | Overlap | | | | |
| 9 | | 7 | A | 0.63 | | 1 | 1 | | 0.63 | 0.63 | Constraints | | | | |
| 10 | | 8 | A | 0.57 | | 1 | 1 | | 0.57 | 0.57 | | Item# | constraints | forms | |
| 11 | | 9 | A | 0.61 | | 1 | 1 | | 0.61 | 0.61 | | 4 | 1 | 0 | |
| 12 | | 10 | A | 0.97 | | 0 | 0 | | 0.00 | 0.00 | | 6 | 1 | 1 | |
| 13 | | 11 | A | 0.77 | | 1 | 1 | | 0.77 | 0.77 | | 21 | 1 | 0 | |
| 14 | | 12 | A | 0.44 | | 1 | 1 | | 0.44 | 0.44 | | 24 | 1 | 1 | |
| 15 | | 13 | A | 0.81 | | 1 | 1 | | 0.81 | 0.81 | | 27 | 1 | 0 | |
| 16 | | 14 | A | 0.95 | | 0 | 0 | | 0.00 | 0.00 | | 35 | 1 | 1 | |
| 17 | | 15 | A | | | 1 | 1 | | | | | 46 | 1 | 0 | |
| 18 | | 16 | A | | | 1 | 1 | | | | | 51 | 1 | 0 | |
| 19 | | 17 | A | | | 0 | 0 | | | | | 67 | 1 | 0 | |
| 20 | | 18 | B | 0.42 | | 1 | 1 | | 0.42 | 0.42 | | 71 | 1 | 0 | |
| 21 | | 19 | B | 0.74 | | 0 | 0 | | 0.00 | 0.00 | | 82 | 1 | 0 | |
| 22 | | 20 | B | 0.56 | | 0 | 0 | | 0.00 | 0.00 | | 99 | 1 | 0 | |
| 23 | | 21 | B | 0.56 | | 1 | 1 | | 0.56 | 0.56 | | 101 | 1 | 0 | |
| 24 | | 22 | B | 0.84 | | 0 | 0 | | 0.00 | 0.00 | | 107 | 1 | 0 | |
| 25 | | 23 | B | 0.67 | | 0 | 0 | | 0.00 | 0.00 | | 114 | 1 | 0 | |
| 26 | | 24 | B | 0.62 | | 0 | 0 | | 0.00 | 0.00 | | 118 | 1 | 0 | |
| 27 | | 25 | B | 0.57 | | 0 | 0 | | 0.00 | 0.00 | | 124 | 1 | 0 | |
| 28 | | 26 | B | 0.70 | | 0 | 0 | | 0.00 | 0.00 | | 130 | 1 | 0 | |
| 29 | | 27 | B | 0.52 | | 1 | 0 | | 0.52 | 0.00 | | 135 | 1 | 0 | |
| 30 | | 28 | B | 0.83 | | 0 | 0 | | 0.00 | 0.00 | | 140 | 1 | 0 | |
| 31 | | 29 | B | 0.60 | | 0 | 0 | | 0.00 | 0.00 | | | | | |
| 32 | | 30 | B | 0.86 | | 0 | 0 | | 0.00 | 0.00 | | | | | |
| 33 | | 31 | B | 0.53 | | 1 | 1 | | 0.53 | 0.53 | | | | | |
| 34 | | 32 | B | 0.75 | | 0 | 0 | | 0.00 | 0.00 | | | | | |
| 35 | | 33 | B | 0.86 | | 0 | 0 | | 0.00 | 0.00 | Objective | | | | |
| 36 | | 34 | B | 0.78 | | 0 | 0 | | 0.00 | 0.00 | Function | | Form1 | Form2 | |
| 37 | | 35 | B | 0.48 | | 1 | 1 | | 0.48 | 0.48 | | Mean Test Scores | 0.52 | 0.52 | |
| 38 | | 36 | B | 0.94 | | 0 | 0 | | 0.00 | 0.00 | | Target | 0.00 | | |
| 39 | | 37 | B | 0.91 | | 0 | 0 | | 0.00 | 0.00 | | | | | |

Figure 9. The Solution for Problem 3.

## DISCUSSION and CONCLUSION

The purpose of our study was to describe and illustrate a test development process that can be used for parallel forms construction using the selected-response item type in order to permit on-demand testing. We presented the logic and highlighted the benefits of parallel forms construction using three different examples. In each example, parallel forms were created that met strict content and statistical constraints. These findings are important because educational testing is undergoing profound changes spurred on, in part, by the application of technology to assessment. As a result, CBT has become commonplace across all levels of education ranging from K-12 to post-secondary levels. This expansion is fueled by the benefits of CBT over paper-based testing that includes the ability to test on-demand, to administer exams at different locations and with different technologies, to provide students with instant feedback while relieving instructors from the monotonous task of manual scoring, and to use dynamic new item types.

Despite these important benefits, exam security remains an important concern. When computer-based tests are administered more frequently, item exposure rates will also increase. To address this concern, parallel test forms are created. Parallel forms are considered to be secure exams because each form contains a different set of items thereby minimizing item exposure in order to enhance test security. But manually assembling parallel test forms is impractical because it requires the solution to a complex combinatorial problem. As an alternative, instructors can implement ATA using existing software that is readily available and relatively easy to implement. In this study we demonstrated the use of the Opensolver, which is a free open-source add-in for Microsoft Excel that can be used to solve a broad range of ATA problems. We began with a simple problem using simulated data. Then we solved two practical problems using real data from an item bank developed

using the student response data from two previously administered exams in a large introductory undergraduate psychology course. The bank contained content codes and statistical indices for each test item. Across all three problems in our study, parallel forms were constructed that met stringent content specifications while satisfying strict statistical targets thereby demonstrating the feasibility and benefits of using this approach for test development.

Instructors may hesitate before attempting to implement ATA due to concerns about the size and quality of their existing item bank. To be sure, item banking is an important requirement for solving test assembly problems. Ideally, a large bank is available with items that span many content areas and include a wide range of difficulty levels. This type of bank was used to solve problem 1. But it is also important to stress that test assembly is a flexible process. In reality, most banks are small or in-development. Problems 2 and 3, which were based on real student data from an undergraduate course in psychology, illustrate how to solve ATA problems under less than ideal item banking conditions.

For problem 2, item overlap rates were needed because the item bank was too small to produce two unique non-overlapping forms that satisfied the objective function of mean score equality. Hence, some common items were included in both forms. The inclusion of common items can be considered both a strength and a weakness in test design. Item overlap has the benefit of providing the instructor with common items across the forms so the performance of students who write each form can be compared directly on the same items. Item overlap also has the drawback of exposure meaning that some items from the test are being viewed by all students in the course. Item exposure has the potential of compromising test security if the same items are used in future exams and if students disclose these items to future test takers. Problems with item exposure can be reduced by decreasing the number of common items used in parallel forms construction.

For problem 3, newly created items that did not contain statistics were included in the construction of the parallel forms. These items were included so data on their statistical performance could be collected. Then, when this statistical information is available, instructors can use these items for future test assembly tasks. The inclusion of practice items can be considered both a strength and a weakness in test design. The benefit of using new items is that statistical information can be collected. This information, in turn, can be used to build a larger item bank. The drawback of using new items is that the difficulty level is unknown and as a result the overall mean score on the parallel forms could be different when the new items are used to compute students' test scores. To resolve this potential problem, new items are often not included in students' mean score calculation. Unfortunately, this introduces the issue of how to explain to students that not all test items actually contribute to their final exam score.

One simple way to address the item banking challenges described in problems 2 and 3 is implement ATA only when a large diverse bank is available. As an alternative, the solutions we presented suggest that there are creative ways to build and replenish banks. These solutions, however, will always require different types of trade-offs and compromises. But it is important to underscore that even with small item banks, ATA can be used to create parallel forms that meet stringent content specifications and statistical targets. In other words, instructors with relatively small banks can use ATA methods to create parallel forms when the goal is to promote flexible on-demand testing while at the same time maintaining security. Because parallel forms measure the same content areas and produce tests with the same difficulty levels, instructors can compare students' test results because the scores across the forms are highly comparable, if not equivalent, to one another.

## REFERENCES

Breithaupt, K., & Hare, D. (2016). Automated test assembly. In F. Drasgow (Ed.), *Technology and testing: Improving educational and psychological measurement* (pp. 128-141). New York: Routledge.

Daniels, L., & Gierl, M. J. (in press). The impact of immediate test score reporting on university students' achievement emotions in the context of computer-based multiple-choice exams. *Learning and Instruction.*

_____

ISSN: 1309 – 6575   *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*
*Journal of Measurement and Evaluation in Education and Psychology*

301

_____

Drasgow, F. (2016). *Technology and testing: Improving educational and psychological measurement.* New York: Routledge.

Drasgow, F., Luecht, R. M., & Bennett, R. (2006). Technology and testing. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed., 471-516). Washington, DC: American Council on Education.

Luecht, R. M. (1998). Computer-assisted test assembly using optimization heuristics. *Applied Psychological Measurement, 22,* 224-236.

Luecht, R. M. (2016). Computer-based test delivery models, data, and operational implementation issues. In F. Drasgow (Ed.), *Technology and testing: Improving educational and psychological measurement* (pp. 179-205). New York: Routledge.

Sireci, S., & Zenisky, A. (2016). Computerized innovative item formats: Achievement and credentialing.  In S. Lane, M. Raymond, & T. Haladyna (Eds.), *Handbook of Test Development* (2nd ed., 313-334). New York: Routledge.

van der Linden, W. J. (1998). Optimal assembly of psychological and educational tests. *Applied Psychological Measurement, 22*, 195-211.

van der Linden, W. J. (2005). *Linear models for optimal test design.* New York: Springer.

_____

ISSN: 1309 – 6575   *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*
*Journal of Measurement and Evaluation in Education and Psychology*

302