STANDARDIZING TEST BY INVOLVING OTHER UNIVERSITIES IN DEVELOPING THE ITEM TESTS

By: Endang Wahyuningrum, Suratinah, and Yos Sudarso

Abstract

Universitas Terbuka (UT), as the only distance learning institution in Indonesia, always tries to give the best services for Indonesian Education. One of the strategic efforts to be the best distance learning university in Asia is standardization of procedure and format evaluation to evaluate students' competencies. Their competencies are examinated through the formative and summative evaluation. To get the valid and reliable test, this evaluation has been standardized in Item Test Bank.

The Unit of Test Development (UPS), as a part of the Faculty of Education (FKIP), has jobs to develop and manage the examination. In 2003, this unit developed 1355 sets of item tests for 225 courses in FKIP. The developing of test involved about 340 lecturers. Two hundred sixty lecturers are from others universities and the rest are lecturers from Universitas Terbuka.

Since 2002, UPS have been developing test using standardized procedure. This procedure has advantages and disadvantages. One of the advantages is a validation process for the moduls as well as for the test. There is benchmarking process in this validation process, because of the involvement of lecturers from other universities. This benchmarking can be used as an academic quality assurance.

This paper describes about the procedure in developing the test, advantages and disadvantages of involving others, and also how benchmarking can be gained.

Introduction

Universitas Terbuka (UT), as the only distance learning institution in Indonesia, always tries to give the best services for Indonesian Education. Some strategic efforts have been made in order to serve the best to its students as well as to be one of the best distance universities in Asia. One of the efforts is to improve the quality of the test by developing an Item Test Bank.

The Item Test Bank, started in 2002, has a task as a unit that responsible in compiling, collecting, and managing item tests for all faculties in UT (there are four faculties). Each faculty should develop at least 10 sets of item tests for each course. The item tests are in the form of objective tests. It is a big challenge for the bank since there are more than six hundred courses to be managed. Therefore, in order to reduce the various form of item tests, the Bank developed a standardized form and procedures for the faculties in developing item tests.

Based on the standard that developed by the Bank, Faculty of Education (FKIP), through Test Development Unit (UPS), develops item tests for the need of the Bank. In 2003, UPS developed 1355 sets of item tests for 225 courses and involved

340 lecturers from six universities including UT. This "colossal" development of item tests brought about advantages and disadvantages for FKIP as well as the involved universities.

This paper describes about the experience that UPS went through during the test development. Advantages and disadvantages of involving others is described in detail. In addition, benchmarking that was gained as a result of collaboration is also explained.

Standardized, Validation, and Benchmarking

In UT, item tests are developed based on blueprints that are written by the lecturers. The blueprint consists of the general goal of the course, specific goal and indicator for each item, degree of competencies, and degree of difficulty for each item. The following figure is an illustration of the tests blueprint.

Number of Modul	Specific goals	Subject	p*	C**	a type of an item test	Number of an Item test	Indicator

Figure 1. A Test Blueprint

p* : degree of difficulty
- easy
- hard enough
- very hard

c** : degree of a cognitive competencies (Taxonomy Bloom)
- knowledge (C1)
- enderstanding (C2)
- application (C3)
- analysis (C4)
- synthesis (C5)
- evaluation (C6)

Since each course should have 10 sets to be put in the Bank, there should be only one blueprint for each course, otherwise it will be very difficult for the Bank to manage the item tests. From this point, the Bank asked the lecturers to write one blueprint for one course so that no matter who will write the item tests, they will write

based on the same blueprint and produce parallel item tests. The blueprint that written as required is called standardized blueprint and the item tests are considered as standardized item tests.

Before the blueprint that is written by the lecturer is accepted, it should be reviewed by another colleague and the chairperson of the program. The chairperson decides who will be the reviewer of each course. In the reviewing process, the reviewers examine the validity of the material in the blueprint. Because the students learn only from module, therefore, the material in the blueprint has to be based on the modules. Besides, the reviewers also have to notice whether the specific goals and indicators are accurate so that the item tests that will be written based on these goals are valid. This process is called validation and the blueprints that have been validated are ready to use for developing item tests.

Meanwhile, during the development of item tests, benchmarking process happened. In this paper, benchmarking is a kind of standardizing which is done by people from other institutions. Therefore, while the blueprint is standardized and validated by the colleague and the chairperson in the program, it is also standardized by the lecturer who involved in the item test development. And this kind of standardization is called benchmarked.

Developing the Item Tests

As a result of the banking system, in 2003, FKIP conducted a colossal item test development to provide the need of the Item Test Bank. During that year, FKIP developed 1355 sets of item tests for about 225 courses. With such a huge number of sets, FKIP no longer could develop the tests by itself since it has limited staffs for that. Furthermore, FKIP is used to be involving other universities in providing tests for its students. This development involved about 340 lecturers from six universities including UT. Two hundred sixty lecturers were from five universities and the rest were lecturers from UT. Three public universities and two private ones participated in the test development. They were Universitas Negeri Yogyakarta (UNY), Universitas Negeri Sebelas Maret (UNS), Universitas Sanata Dharma (USD), Universitas Kristen Satya Wacana (UKSW), and Universitas Pendidikan Indonesia (UPI). These five universities were chosen based on several criteria such as programs and staffs they have.

Before asking a help to develop item tests, FKIP usually tries to get information about the intended university from the insider or informant that FKIP knew well. From the informant, FKIP gets the information about the programs and the staffs it has and about the possibility to involve the university in the activity. The connection with the informant is done informally. Later on, the formal correspondence begins. This procedure was also followed during 2003 development. The five universities agreed to help FKIP in developing item tests.

After they agreed to help FKIP, then, UPS began to arrange the activity. Firstly, UPS prepared the materials to be used in the development such as modules, blueprint, and item test cards for each lecturer who was involved in the activity.

The lecturers had to use the provided blueprint in developing the tests. Besides, they had to read the modules of the course in order to understand material. Then, to help them know what they have to do, UPS conducted one-day workshop at each university. During the workshop, it was discussed about how to develop the test, read the blueprint, and fill in and write each item on the card. The workshop is intended to make the lecturers have the same perception of UT's rules in developing item tests. At last, each lecturer had a time to develop item tests individually. The following figure is an illustration of the card.

Figure 2. The Card

4

Advantages and Disadvantages

Involving others in developing item tests has some advantages as well as disadvantages. Before a lecturer develops item tests, he has to read the modules seriously and carefully. Reading the modules gives benefit both for FKIP and the lecturer himself. Because the modules are the only one source the students use to learn, they consist of many important topics from many sources. Besides, they were written by some experts from qualified universities. Therefore, sometimes, the lecturer learns something from the modules. However, some modules were written long ago, then, when the lecturer reads them he may find outdated materials. In addition, some lecturers found some mistakes in the materials discussed in the modules. In this case, their reading activity unintentionally validated the modules because they give judgments about the contents of the modules. As a result, FKIP gets good input about those materials for revising the modules.

Moreover, the lecturers who involved in the test development teach the same or similar course with the course that its item tests were developed. Therefore, when they read the modules, they may find that they also have the same topics in their course. Many of them, then, use the modules for their students. Even, some of them have already used the modules before they involved in the test development. This process can be considered as the process of benchmarking for both modules and item tests.

Furthermore, when the lecturers started to write the item tests, they had to learn the blueprint of the test. They tried to know whether the content in the modules are compatible with the content of the blueprint; whether the blueprint really measures the students' competencies required in the modules. Again, this activity gave benefit for both sides. FKIP got significant feedback for revising the blueprint if any, while the lecturers learn about writing test using blueprint.

Indirect benefit is also gained from involving others in developing the item tests. The lecturers from other universities developed item tests based on the blueprint written by the staffs of FKIP. They learnt and read the blueprint before they started to write the test. In other words, they validated the blueprint as well as the FKIP staffs' competencies. On the other hand, when they finished writing the tests, the FKIP staffs reviewing their item tests – here, validating competencies also happened. Being

validated from others may encourage staffs from both sides to improve their competencies.

The most benefit that FKIP gained from involving others was that a huge number of sets were written in a relatively short time. About 881 sets were written in five universities and most of them have done in less than two months. Without collaboration this was hardly reached.

Although some advantages were gained from the activity, there were some disadvantages aroused during this time. When faculty decided to involve others in developing the tests, UPS had to do a lot of preparation and it needed much time to do. First, it contacted the informants in the intended universities. Then, it sent a permission letter to the universities and waited for the answer. After that, UPS prepared the modules, blueprints, item test cards, and some administration forms.

A lot of budget needed was another disadvantage for involving others in developing tests. A lot of money needed for transportation and accommodation for the staffs during the workshop and picking up the sets.

There were some lecturers discontinued to write the tests in the middle of the time agreed. This forced UPS or Program to find out another person to write the tests and this made the tests late. Furthermore, some lecturers did not attend the workshop and did not ask their colleagues about it. As a result, they made the tests without any understanding of the rules that caused their tests not standardized. Even, it was found that a few lecturers made their own blueprint that obviously different from the one that required. And it was also found, although very few but significant enough to be noticed, that lecturers who came to the workshop with some reasons gave their tasks to other lecturers. This also caused problems because the substitute lecturers did not know the FKIP rules in writing the tests and the former lecturers did not give information to them.

Conclusion

Universitas Terbuka (UT) has some strategic efforts to be one of the best distance universities in Asia. One of the efforts is to improve the quality of the test by developing an Item Test Bank. The Item Test Bank, started in 2002, has a task as a unit that responsible in compiling, collecting, and managing item tests for all faculties in UT (there are four faculties) at least 10 sets of item tests for each course. In order to

reduce the various form of item tests, the Bank developed a standardized form and procedures for the faculties in developing item tests.

Based on the standard that developed by the Bank, in 2003, Faculty of Education (FKIP), through Test Development Unit (UPS), develops 1355 sets of item tests (for 225 courses) for the need of the Bank. This development involved 340 lecturers from six universities including UT. During the development of item tests, benchmarking process happened. Benchmarking is a kind of standardizing which is done by people from other institutions, while the blueprint is standardized and validated by the colleague and the chairperson in the program.

This "colossal" development of item tests brought about advantages and disadvantages for FKIP as well as the involved universities. The advantages that FKIP are validating to test blueprint, getting good input about the materials for revising the modules and using the modules by others lecturers for their students. Moreover, the most benefit that FKIP gained from involving others was that a huge number of sets were written in a relatively short time.

Although some advantages were gained from the activity, there were some disadvantages aroused during this time. UPS had to do a lot of preparation and it needed much time to do. A lot of budget needed for involving others in developing tests. Some lecturers discontinued to write the tests in the middle of the time agreed. Some lecturers did not attend the workshop and caused the item tests that they wrote were not standardized. A few lecturers made their own blueprint that obviously different from the one that required.

References

Asandhimitra, dkk. 2004. *Pendidikan Tinggi Jarak Jauh*. Jakarta: Universitas Terbuka.

Pannen, Paulina, dkk. 1999. *Cakrawala Pendidikan*. Jakarta: Universitas Terbuka.

Tim Penulis Universitas Terbuka, 2003. *Pedoman Pengembangan Tes Objektif*. Jakarta: Universitas Terbuka.

Tim Penulis Universitas Terbuka, 2003. *Pedoman Penilaian Hasil Belajar Mahasiswa*. Jakarta: Universitas Terbuka.

Tuckman, Bruce W., 1975. Measurement Educational Outcomes. New York: Harcourt Press