Development of Mathematics Test Items

Nantachat Wongpanya
IPST, Bangkok, Thailand
nanw@ipst.ac.th

Supattra Pativisan
IPST, Bangkok, Thailand
spati@ipst.ac.th

Jatupon Ngamman
IPST, Bangkok, Thailand
jngam@ipst.ac.th

IPST has responsible for improving quality of teaching and learning in Science, Mathematics and Technology; emphasis on K-12 levels. To achieve this objective, Standard and Assessment Department has been working on a standard achievement test. The purpose of this project is to design and develop test items for measuring mathematical content knowledge and skills. This paper presents the process of item development and some examples of modified items. The items cover 6 content areas of the Thai Basic Education Curriculum: Number and Operations, Measurement, Geometry, Algebra, Data Analysis and Probability, and Mathematical Skills and Processes. The process consists of (1) setting an assessment framework; (2) constructing items (3) reviewing items and (4) conducting two rounds of try-out with sample groups. After try-out processes, the item analysis is conducted to determine quality of the items, and provide feedback for modification. Item difficulty indices and item discrimination indices are considered when selecting to the item bank. Results also indicate what knowledge and skills students could not achieve and need more understanding.

Keywords: Development Test; Mathematics Test Items; Multiple-Choice Test Items

Introduction

IPST is an autonomous organization under the direction of the Ministry of Education that has a vision to increase Thai's quality of science, mathematics and technology education to the international standards. Measurement and assessment processes of education are the core factor which can reflect the quality of education management and student learning. Furthermore, the implementation of measurement and assessment processes could improve the development of science and mathematics learning.

Standard and Assessment Department as a part of IPST is responsible for development test items to assess students' achievement in the area of science and mathematics. In this paper, a development process of multiple-choice test items is discussed in details.

A development process of test items

The procedure of multiple-choice test items development consisted of four steps: setting an assessment framework, constructing items, reviewing items, and conducting try-out.

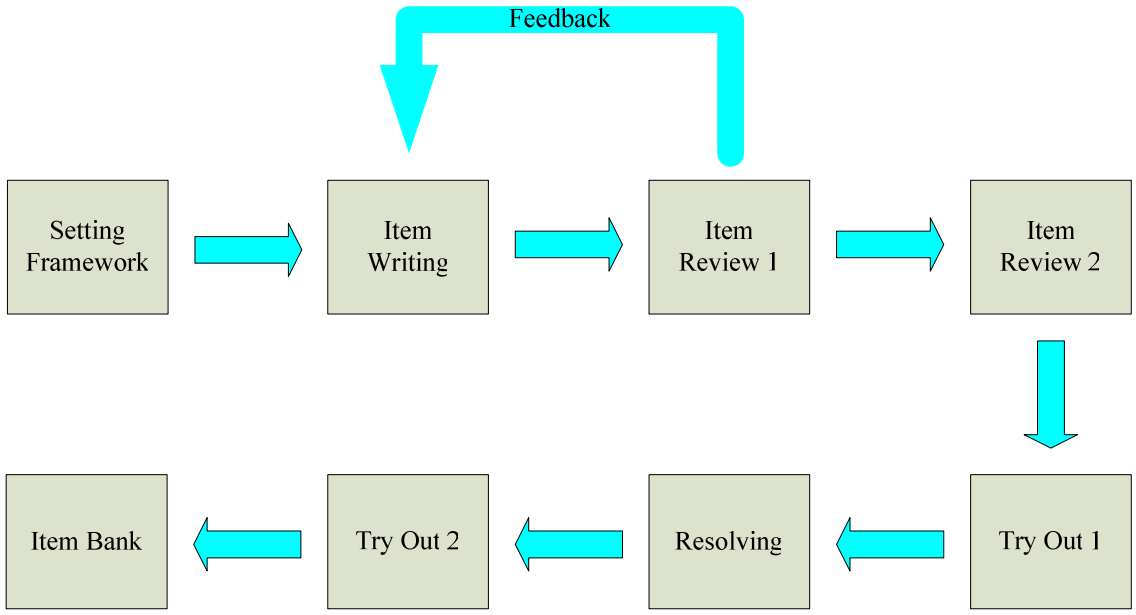


Fig 1: A development process of test items

Step 1: Setting an assessment framework

First of all, the assessment framework should be defined, and it consists of two parts: test descriptions and a table of specification. Generally, test descriptions comprise many aspects of the exam, such as test purposes, target examinee, and test length. The second part is a table of specification which identifies major content areas, cognitive levels, numbers of items within each major content areas and cognitive levels. In addition, the test specification needs to be consistent with the curriculum standards.

Step 2: Constructing items

After the assessment framework has been set, it is important for test writers to develop the test items by following the table of specification. The item writer should be the person who is specializes in both content knowledge and item development.

Step 3: Reviewing items

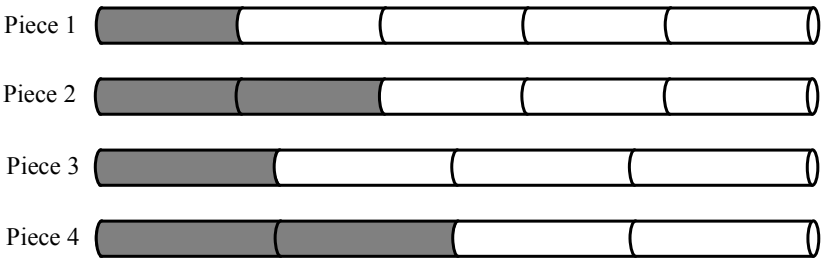
Once the test items have been written, the next important step is to review them. The item review committee comprises subject matters experts and measurement experts. The test items should be reviewed all aspects such as content, grammar and spelling, distracters, and correct answers. In addition, the fairness, e.g. gender, region, and religion should be considered to ensure that the test items will not be likely to disadvantage any test taker. The first reviewer is the academic staff from IPST and afterwards the test items will be reviewed by the outside reviewing committee. Feedbacks from reviewers will be discussed to identify issues of test items and making corrections before conducting the first round of try-out.

Step 4: Conducting try-out

After revising test items, a try-out process is conducted for quality testing. There are two rounds of try-out. The first round begins with a group of 200-300 students. Then, the first round results are statistically analyzed for difficulty, discrimination and distracter. The modified items from the first round are used with another group of 800-1,000 students in the second round. After that, the second round results are statistically analyzed again.

Examples of test items

Example 1: Four equal length of bamboo pieces which are painted in black and white, as shown below.



Which one has 20% of black color?

A. Piece 1
B. Piece 2
C. Piece 3
D. Piece 4

Round 1	**Selected choice**	**A**	55.4 %	**B**	20.4 %	**C**	9.2 %	**D** 15.0 %
	Difficulty	0.55		**Discrimination**	0.44			
Round 2	**Selected choice**	**A**	46.4 %	**B**	25.0 %	**C**	10.4 %	**D** 18.2 %
	Difficulty	0.46		**Discrimination**	0.53			

From the first example, it is a mathematics item for sixth grade students that measured ratio and percent concepts. The statistics of two rounds meet the requirement. Consequently, this item is stored in the item bank for measuring the student achievement.

Example 2:

Container A has water as shown in Figure 1, if we pour one liter of water from container A to container B, then water level of A and B are shown in Figure 2.

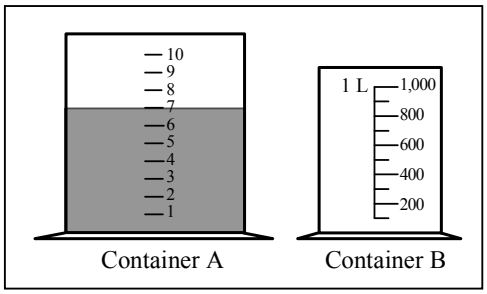


Figure 1

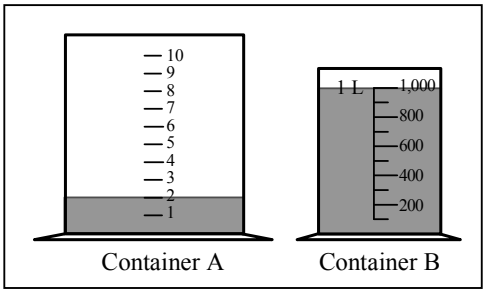


Figure 2

In Figure 2, how much water left in container A?

A. 100 milliliter
B. 200 milliliter
C. 300 milliliter
D. 400 milliliter

Round 1	**Selected choice**	**A**	17.2 %	**B**	51.6 %	**C**	18.3 %	**D** 12.9 %
	Difficulty	0.13				**Discrimination**	0.03	

From the second example, it is a mathematics item for sixth grade students that measured a volume concept. The result of the first round try-out shows Difficulty 0.13. It means that this item is very difficult. Furthermore, the Discrimination 0.03 indicates that item can not be discriminated between high and low performance students. The distracter analysis indicates that the students choose B most (51.60%). It means that the students may be confused between the height of container A and the level of water of container B.

After revising example 2 and try-out again, the results of the test are as shown in example 3. We can see that the difficulty and discrimination are increase and meet the requirement.

Example 3:

Container A has water level as shown in figure 1, if pouring one liter of water from A into B, the water level of A and B are shown as in figure 2.

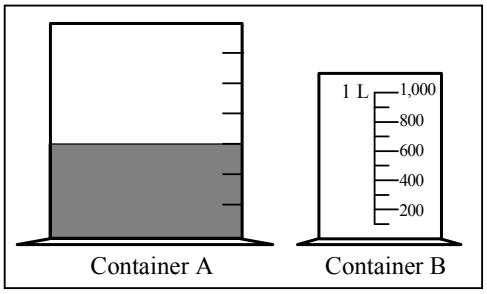


Figure 1

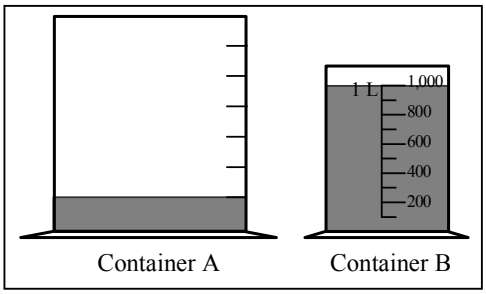


Figure 2

From figure 2, how much water left in container A?

A. 500 milliliter
B. 300 milliliter
C. 200 milliliter
D. 100 milliliter

Round 1	Selected choice	A 31.3 %	B 17.3 %	C 32.6 %	D 18.8 %
	Difficulty	0.31	Discrimination	0.35	

Summary

After finish four steps of development, the qualified items with 0.2-0.8 difficulty, 0.2 up discrimination and each distracter must be selected by 5% of respondents are selected for the item bank. These qualified items are ready for the national assessment process that can be evaluated the quality of student learning. The test results also provide evidence of students' understanding that would be feedbacks for curriculum development and textbook revisions.

References

Hedges, William D. (1966). *Testing and evaluation for the sciences on the secondary school*. Belmont: Wadsworth Publishing.

Haladyna Thomas, M. (1999). *Developing and validating multiple-choice test items*. New Jersey: Lawrence Erlbaum Associates, Inc.

The Institute for the Promotion of Teaching Science and Technology, Ministry of Education (2003). *Science Assessment Manual*. Bangkok: S.P.N. Printing Co.,Ltd.

The Institute for the Promotion of Teaching Science and Technology, Ministry of Education (2008). *Basic Education Curriculum B.E. 2008*. Bangkok: The Agricultural Co-operative Federation of Thailand Printing Co.,Ltd.