

Development of Science Test Items

Tanyakarn Yuentrakulchai
IPST, Bangkok, Thailand
kampa@ipst.ac.th
wkamt@ipst.ac.th

Wanpen Kamtet
IPST, Bangkok, Thailand

Precharn Dechsri
IPST, Bangkok, Thailand
pdech@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 8 content areas of the Thai Basic Education Curriculum: Living Things and Processes of Life, Life and the Environment, Substances and Properties of Substances, Forces and Motion, Energy, Change Process of the Earth, Astronomy and Space, and Nature of Science and Technology. 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; Science 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.

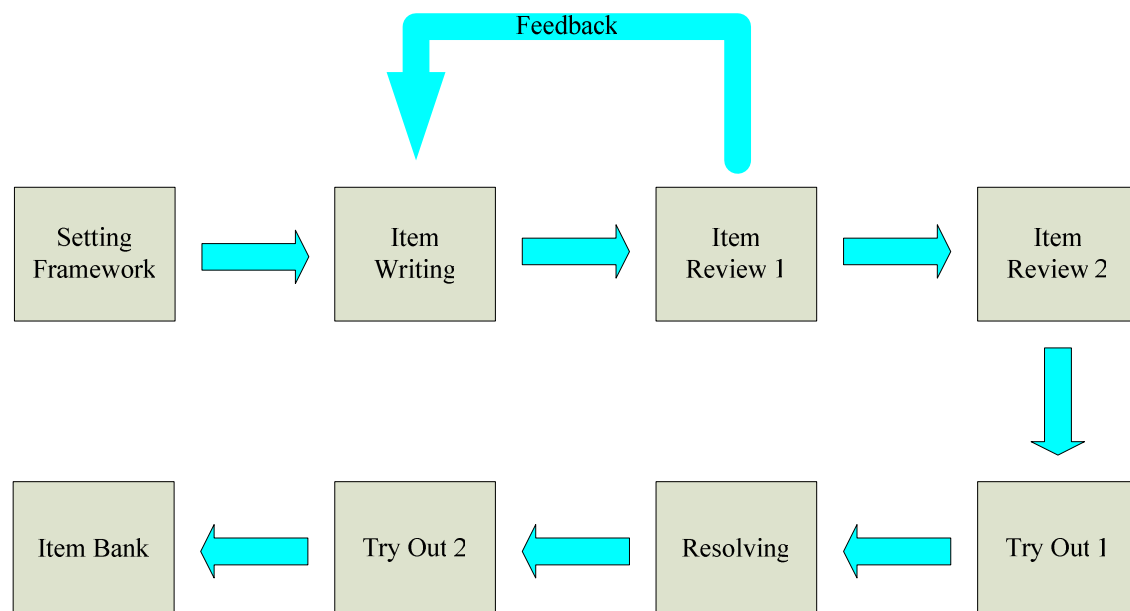


Fig 1: A development process of test items

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 specializes in both content knowledge and item development.

Step 3: Reviewing items

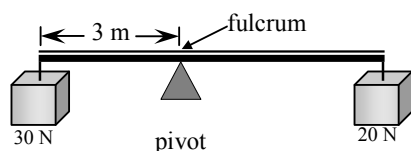
Once the test items have been written, the next important step is to review them. The item review committee comprises subject matters experts and measurement experts. The test items should be reviewed all aspects such as content, grammar and spelling, distracters, and correct answers. In addition, the 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: Two objects were hung at the end of a lever as shown in the picture.



If the lever was balanced, find the distance between 20 newton object and the fulcrum?

A 3.0 m

B 4.5 m

C 7.5 m

D 9.0 m

Round 1	Selected choice	A	18.2 %	<input checked="" type="radio"/> B	45.4 %	C	23.2 %	D	13.2 %
	Difficulty	0.45			Discrimination	0.57			
Round 2	Selected choice	A	14.2 %	<input checked="" type="radio"/> B	59.4 %	C	16.8 %	D	9.6 %
	Difficulty	0.59			Discrimination	0.42			

After using example 1 in the first round of try-out, the statistics meet the requirement. In the second round, the results were similar. Consequently, this item is stored in the item bank which means it can be used for measuring the student achievement.

Example 2: Materials used in toys' production were shown in the table below.

Toys	Materials
type 1	wood and rubber
type 2	metal, plastic, and rubber
type 3	plastic
type 4	cloth and rubber

In the table, which material can be used with the most toys' production?

A wood

B metal

C rubber

D plastic

Round 1	Selected choice	A	22.2 %	B	19.4 %	C	19.8 %	D	38.6 %
	Difficulty	0.20			Discrimination	0.03			

After using example 2 in the first round of try-out, the result of Difficulty 0.20 and Discrimination 0.03 can be inferred that this test item was difficult and cannot discriminate between high and low performance students. Experts comment that students cannot interpret data from a given table due to those data caused them confuse. Therefore, the alteration has been made before conducting the second round of try-out.

Materials used in toys' production were shown in the table below.

Material	Toys			
	Type 1	Type 2	Type 3	Type 4
cloth	×	×	✓	×
wood	×	✓	×	×
rubber	✓	×	✓	×
plastic	✓	✓	✓	×

In the table, which material can be used with the most toys' production?

A wood

B metal

C rubber

D plastic

Round 2	Selected choice	A	7.7 %	B	9.3 %	C	7.1 %	D	75.9 %
	Difficulty	0.76			Discrimination	0.55			

After conducting the second round of try-out, a slightly change in given data made the statistics increase and meet the requirement.

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.
- Klopfer, Leopold, E. (1971). Evaluation of Learning in Science. *Handbook on Formative. Handbook on Formative and Summative Evaluation of Student Learning*. New York: McGraw-Hill.
- 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.