*МАТЕМАТИКА И МАТЕМАТИЧЕСКО ОБРАЗОВАНИЕ, 2000*
*MATHEMATICS AND EDUCATION IN MATHEMATICS, 2000*
*Proceedings of Twenty Ninth Spring Conference of*
*the Union of Bulgarian Mathematicians*
*Lovetch, April 3–6, 2000*

# MEASUREMENT OF THE CURRICULUM COVERAGE IN INTERNATIONAL STUDIES OF STUDENT ACHIEVEMENT

**Kiril Gueorguiev Bankov**

Before comparing student achievements by countries, all the efforts should be directed to ensure an usable test thst reflects as much as possible the curriculum of the participating countries. Nevertheless, no one test can cover all that is taught or learned in every country. The question arisen is how well the items on the test match the curriculum of each country. The paper deals with this problem in the context of the Third International Mathematics and Science Study (TIMSS), conducted by the International Association for Evaluation of Educational Achievement (IEA).

The test for each international survey usually is developed to represent a set of the agreed-upon content area of the participating countries. Because of the differences among the curricula, some topics of the test are not taught or taught at different grades. One way to compare countries as "fair" as possible is to restrict test items to the common topics of the curriculum of all countries. This approach is not a clever solution to the problem, because it severely limits the test coverage and restricts research questions about international differences. It should be accepted, therefore, that the test should contain some items for measuring topics which are unfamiliar to the most students in some countries.

A technique called Test-Curriculum Matching Analysis (TCMA) was developed to observe the student performance for a particular country based only on the test items that are relevant to its own curricula. It enables also for each country to investigate not only the performance of all other countries on the set of items appropriate for its curricula, but also the performance of its students on the items relevant to the curriculum in other countries.

The TCMA needs information for each country about the set of items from the test that are relevant to the intended curricula of the country. In TIMSS, for example, each National Research Coordinator (NRC) was asked to report, before the administration of the test, whether or not each test item is appropriate or not for the country. The item is considered appropriate if it was in the intended curriculum for more than 50% of the students in the country [1]. This information is presented by numbers $t_{ij}$, being 1 if the item $j$ is appropriate for the country $i$, and 0 if the item $j$ is not appropriate for the country $i$. It is also possible to gather data not limited to a binary choice, but according to the degree of the appropriateness of the item $j$ for the country $I$. In this case, it may have any value between 0 and 1. If $I$ countries participate in a survey and the test has $J$ items, the Test Coverage Matrix has $I$ rows and $J$ columns.

Suppose that the response of each item of the test can have two values, 1 for correct and 0, otherwise. Then the score of each student may be reported in a proportion-correct form, that is, the ratio of the correct answers of the student to all items of the test. For each country $i$, the average of the student proportion-correct responses $p_{ij}$ for the item $j$ in this country is calculated. The Proportion-Correct Matrix $P = (p_{ij})$ has $J$ rows and $I$ columns.

The Test Coverage Index for each country is the ratio of the total possible test points on the appropriate items of the country to the total possible test points in the whole test. The Test Coverage Index indicates the proportion of the score points of the test that is considered appropriate to the curriculum in the country. For example, the Test Coverage Index for TIMSS mathematics test in grade 8 for most countries ranges between 0.7 and 1. For Bulgaria it is 0.73 This means that around 73% of the TIMSS mathematics test items are covered by the intended curricula in Bulgaria. Countries differ more on the science curriculum. The test Coverage Index for TIMSS science test in grade 8 for most countries ranges between 0.5 and 1. For Bulgaria it is 0.77 [2].

To facilitate the comparisons between countries, it is useful to estimate the national proficiency for each country. It has the property that, if the students in a country correctly answered all items that are appropriate for the country, then the country will receive a value of 100; if the students answered all these items incorrectly, then the country will receive a value of 0. Items that are not appropriate for the country are not used in computing these values. If the matrix $T$ contains numbers other than 0 and 1, some country values may exceed 100, which means that students answered more items correctly than it was expected. To avoid such values the normalized weight matrix is computed, where $w_{ij} = \dfrac{t_{ij}}{\sum\limits_{j=1}^{j} t_{ij}^2}$.

The last analytical step is to compute the Country Comparison Matrix $C = (c_{mn})$ by the matrix multiplication $C = 100.(W.P)$. It has $I$ rows and $I$ columns. The elements of $C$ are the national proficiency values. The number $c_{mn}$ indicates how the students in country $n$ scored on the items that are appropriate for country $m$.

The diagonal elements of $C$ show how each country performs on the set of items that are selected based on its own curriculum. TIMSS results [1] show that there is a very small increase in each country's performance on this set of items and performance on the test as a whole. The average percent correct for eight grade students on mathematics in Bulgaria, for example, is 60, while the average percent correct based on the items selected as relevant to the Bulgarian curriculum is also 60. For most countries the difference is no more than 2 average percent points. Only a few countries have an average percent correct on their own selected items more than 3 percentage points higher than their average on the whole test.

Each row of the matrix $C$ shows the performance of the country of the row on the set of items selected as appropriate for other countries. Each column of $C$ shows the relative performance of countries on the set of items selected as appropriate for the country of the column. In TIMSS, for example, the selection of items does not significantly affect the general relationship among countries. Countries that have higher or lower performance on the whole test in comparison to each other also have higher or lower

relative performance on the different subset of items selected as appropriate for different countries. Although there are some changes in the order of countries based on the item selection of the countries, most differences are not statistically significant. For example, based on items that are relevant for Bulgarian mathematics curriculum for the eighth grade in TIMSS, Australia did 56% average correct, and Israel did 57% average correct, i.e. Israel performed better than Australia on Bulgaria's selection of the item set. On the whole test Australia and Israel did 58% and 57% correct responses on average, respectively, i.e. Australia performed better than Israel. But there is not a statistically significant difference between the performances of Australia and Israel on the whole test.

These results are not surprising. On the one hand, they show that the TIMSS mathematics test provides between reasonable basis for comparing achievements of the participating countries. On the other hand, the fact that the countries rejected the items that would be difficult for their own students did not significantly reflect the comparison. These items tended to be difficult also for students in other countries. Omitting such items improves the result of that country, but tends to improve the results for other countries as well. So, the overall order of the countries on their achievement is not significantly affected.

REFERENCES

[1] A. Beaton et al. Mathematics Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study (TIMSS), IEA, 1996.
[2] A. Beaton, E. Gonzalez. TIMSS Test-Curriculum Matching Analysis. In: M. Martin, D. Kelly, TIMSS Technical Report, Volume II: Implementation and Analysis, IEA, 1997.

Kiril Bankov
Faculty of Mathematics and Informatics
University of Sofia
5, James Bouchier Bul.
1164 Sofia
e-mail: kbankov@fmi.uni-sofia.bg

## СРАВНЯВАНЕ НА ПОКРИВАНЕТО НА УЧЕБНИТЕ ПРОГРАМИ ПРИ МЕЖДУНАРОДНИ ИЗСЛЕДВАНИЯ НА УЧЕНИЧЕСКИТЕ ПОСТИЖЕНИЯ

**Кирил Георгиев Банков**

При международни педагогически изследвания сравняването на ученическите постижения често става с помощта на тест. При конструкцията на такъв тест се прави всичко възможно за да се обхване най-добре учебният материал на участващите държави. Учебните програми, обаче, са много разнообразни. Поради това за всяка държава в теста има въпроси, знанията за решаването на които не са изучени от нейните ученици. Така възниква въпросът какво може да се направи, за да може сравняването на ученическите постижения между участващите държави да е колкото е възможно по-равноправно. В този материал се дава един възможен отговор на поставения въпрос в контекста на Третото международно изследване на обучението по математика и природни науки.