

# ITEM STEM FORMAT AND ITEM ANSWER FORMAT INFLUENCE ON THE RESULTS OF THE ITEM

Jolita Dudaite  
National Examinations Centre,  
Kaunas University of Technology, Lithuania

## Abstract

The article deals with item stem format and item answer format influence on the results of the item. Lithuanian case is under consideration. The TIMSS 2003 8<sup>th</sup> grade national booklets results are analyzed (out of 150 schools, 806 students participated). In TIMSS 2003 Lithuania added 2 national booklets for mathematics, in which the TIMSS items were changed in such a way that it would be possible to reach the following goals – to verify the effect of the answer format and the stem format of the item on the results. Analysis of the results showed, that the 8<sup>th</sup> graders in Lithuania solve multiple-choice answer format items statistically significantly better and omit them less than open-ended answer format items. There was no difference found between the genders. According to the results of the analysis, the distractors given in multiple-choice answer format items influence their choice. If exactly the same items would be presented, only without the answers, considerably less students would get the answers that are given in the distractors. There was a hypothesis, that if the TIMSS item's stems would be formulated more in "Lithuanian style", the more familiar format of the stem would help students to solve the items better, but it did not prove to be so. However, analysis of the results shows, that in general the formulation of the stem has an effect on the results of the item.

**Keywords:** *multiple-choice, open-ended, item format.*

## Introduction

When educational research is conducted, multiple-choice items as well as open-ended ones are used. It is constantly debated on how much percent of the test must multiple-choice items take and how much percent of it must be in open-ended items. The usage of various answer types of items has its benefits and its faults. By using multiple-choice items it is possible to cover a larger scale of material in a shorter period of time; this type of items guarantees an easy, absolutely objective and cheaper way of marking; if these items are well made, they correlate well with the results received from solving other types of items; multiple-choice items allow to easily identify standart mistakes; it is especially appropriate when the choices of possible answers are clearly determined. However, for multiple-choice items, it is quite difficult to choose appropriate distractors; with these items it is difficult to examine higher level of ability (problem raising, argumentation, etc.); they can be easily "copied"; there is a chance of guessing. In case of open-ended items (including short-answer, short-solution, structured) there is no need to think of the chance of guessing; there is no need to choose appropriate distractors; these items are suitable for evaluation of various levels of abilities; they allow to involve in a more particular, or problematic situation, by giving a list of items that become harder as the list progresses. However, the solving of open-ended items takes a lot of time; therefore not as many topics may be covered as with the multiple-choice items; they are more difficult to evaluate, more subjective and more expensive; usually, with open-ended items the most typical mistakes are not identified.

Due to the fact that the making and the administration of the different answer types of items has its pluses and minuses, it is important to know the impact of the different answer types of items on the actual results – how much of it, judging by the answer type, influences the difficulty of the item and how much of it – not answering (omitting) the item. With the intention of clarifying this question, many studies have been conducted. However their results

are not the same. It is claimed, that the answer format has a different effect on different teaching subjects as well as on different grades.

According to Elley and Mangubhai (1992), on large scale reading ability tests, the answer format does not have a statistically significant difference on the results. Hastedt (2004), by analyzing international reading ability study (PIRLS 2001) results, shows the opposite, that the answer format has an effect on the results of the given item. Multiple-choice items solved on average statistically significant better than open-ended items. Abrahamson (2000) comes to the conclusion that in the tests of physics, different answer format items' results do not differ. Nasser (internet) claims that in the tests of statistical literacy, multiple-choice items are answered worse than the open-ended items. According to Gadalla (1999), in the field of mathematics computation, there is no statistically significant difference in results of items of the different type of answer format in the 4<sup>th</sup>, 5<sup>th</sup> and 6<sup>th</sup> grade, but in the 2<sup>nd</sup> and 3<sup>rd</sup> grades students answer the multiple-choice items statistically significantly better than the open-ended items. Traub (1993) claims that for the quantitative domain, the different answer format has no effect for the results. Zabulionis (1997), by analyzing TIMSS 1995 7<sup>th</sup> and 8<sup>th</sup> grade students of Eastern Europe Mathematics and Science results, comes to the conclusion that in some countries students are more likely to omit without solving a multiple-choice format item than an open-ended format item, but in most countries the answer format doesn't have an effect on the omitting of the item.

## Methodology

Due to the fact that the results of various researches differ, it is interesting to analyze, how the item answer format, as well as the item stem format influences the 8<sup>th</sup> grade students mathematics results in Lithuania. The TIMSS 2003 8<sup>th</sup> grade national booklets results will be analyzed (out of 150 schools, 806 students participated). In TIMSS 2003 Lithuania added 2 national booklets for mathematics, in which the TIMSS items were changed in such a way that it would be possible to reach the following goals:

1. To verify the effect of the **answer format** of the item on the item results:
  - 1.1. Whether the multiple-choice answer format items' results differ from the open-ended answer format items' results;
  - 1.2. How does the multiple-choice answer format items' given distractors influence their choice;
2. To verify the effect of the **stem format** of the item on the item results:
  - 2.1. Whether TIMSS items, rephrased in the "Lithuanian style", were solved better (by "Lithuanian style" we mean the formulations more commonly used in Lithuanian Mathematics textbooks);
  - 2.2. Whether the results of the TIMSS items differ when the item stem phrasing is changed in a certain way.

Because of it, the national booklets were arranged in the following way:

1. Multiple-choice answer format TIMSS items were changed to open-ended answer format items and vice versa;
2. Rephrased TIMSS items:
  - 2.1. The stems were rephrased in "Lithuanian style";
  - 2.2. The stems were rephrased in other different ways;
  - 2.3. The fractions in the stems were written in words and vice versa;
3. A few extra "TIMSS style" items were created, in which the ideas previously mentioned were verified;
4. A part of unchanged TIMSS items was taken, so that it would be possible to check, whether the achievements of the students who solved the national

booklets are identical to the achievements of the students who solved the TIMSS booklets.

**The main goal** of this analysis is to verify the effect of the answer format and stem format of the item on the item results.

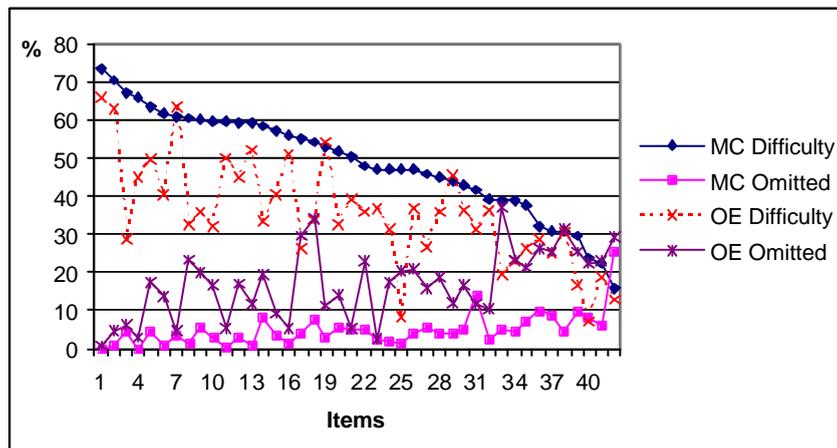
The following hypotheses were made:

1. Students solve the multiple-choice answer format items better than the open-ended answer format items;
2. Students are more likely to omit the open-ended answer format items than the multiple-choice answer format items;
3. The distractors of the multiple-choice items have an influence on choosing them (solving open-ended answer format items less students get a certain wrong answer, then if they would choose it in the case of a multiple-choice answer format);
4. The item's solving results are influenced by the phrasing of the stem of the item;
5. "Lithuanian style" items are solved better;
6. The items are solved worse when in the item stem used fractions are written in words.

## Results

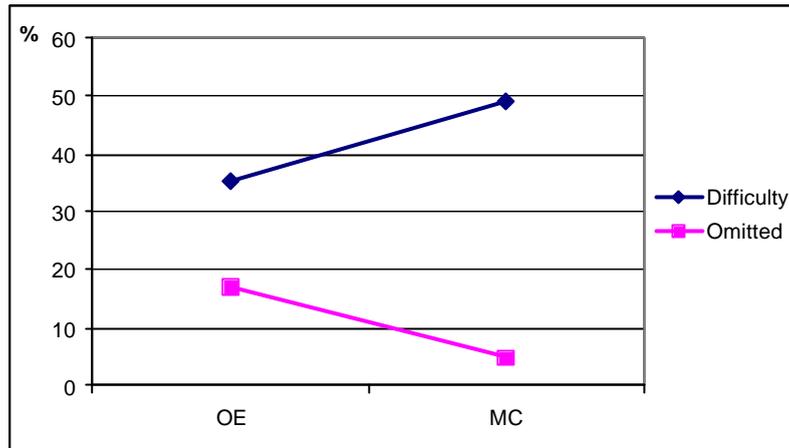
Analysis of the results showed, that the 8<sup>th</sup> graders in Lithuania, solve multiple-choice answer format items statistically significantly better and omit them less than open-ended answer format items (see Exhibit 1). 42 items were analysed.

Exhibit 1. The difference between the results of a multiple-choice and an open-ended answer format items.



As we can see, the multiple-choice answer format items' difficulty and omitting curves cover the open-ended answer format items' difficulty and omitting curves. It means, that if the same item was presented in a multiple-choice answer format, students would solve it in an average better and will omit it less, comparing with the case when it was presented in an open-response answer format. A general difference between a different answer format difficulty and the omitting of an item is presented in Exhibit 2.

Exhibit 2. Average difference between the multiple-choice answer format and open-ended answer format items' results.



There was no difference found between the genders in solving the different answer format items – neither in the difficulty of the item nor in the omitting of the item.

Analysis showed, that in multiple-choice answer format items, the given distractors influence their choice. Should exactly the same items be presented, only without the multiple-choice answers, considerably less students get the answers that are given in the distractors in the multiple-choice format case. Because of that, the distractors, which are considered as typical mistakes, need to be evaluated with much care. One of such item is presented in the Exhibit 3.

Exhibit 3. Difference in the results, between the multiple-choice and the open-ended format answer format items.

What fraction of an hour has passed between 1:10 a.m. and 1:30 a.m.?	MC	OE
A $\frac{1}{5}$	12,6 %	1.8 %
<b>B <math>\frac{1}{3}</math></b>	<b>50.0 %</b>	<b>39.3 %</b>
C $\frac{1}{2}$	17.5 %	1.3 %
D $\frac{2}{3}$	10.1 %	0.3 %
E $\frac{3}{4}$	3.1 %	28.5 %
Other	-	23.5 %
Omitted	5.8 %	5.5 %

Chi square test shows that the chosen answers depend on the item format ( $\chi^2=481.368$ ,  $df=6$ ,  $p=0.000$ ).

It can be seen in the table that in the multiple-choice format case the most students choose the wrong answers C and A. Those answers might be considered as a typical mistakes. Choosing the distractor A students “forgot” that an hour consist of 60 minutes, but not of 100. It is a common mistake in a Basic School. Those students who choose the distractor C maybe were misleded by the “similarity” of 20 minutes and 1/2.

Analysing the results of the open-ended answer format case we see that almost nobody got wrong answers 1/5 and 1/2 (given in the distractors A and C in the multiple-choice format case), but a lot of students got answer 3/4 (which was the distractor E). These results indicate the different typical mistake.

So, this example shows that the multiple-choice answer format items’ the given distractors influence their choice.

Analysis shows that the phrasing of the stem has an effect on the results of the item. However it is difficult to guess, what change in the stem may ease or harden the item. For example, if in one item the worded stem is taken away (“calculate the expression and the answer write down to decimal”) and having left only the expression, which needs to be calculated (add two simple fractions), 20 percent less students solve such item, than in the case which includes the worded stem by the expression. But in another identical item (to subtract two simple fractions), after having taken off the worded stem, 20 percent more students solve the item.

By trying to rephrase TIMSS items in different ways it was found some natural results: students solve an item worse, when instead of the number the unknown quantity (x) is written; students solve an item better, when instead of the part of an hour the minutes is written.

By making the TIMSS item’s stem more Lithuanian-like, it was hoped that the more familiar format of the stem would help the students solve the item better. But it did not prove to be so. Some TIMSS items, which were rephrased in style, more common for the Lihuanian students, were solved better, some items were solved worse, and in some items the difference between the results before and after trying to formulate an item in Lithuanian style were not found. For example , in Exhibit 4 is seen, that after rephrase the item in more common style for Lithuanian students, it was solved better (the answers depend on the item format;  $\chi^2=19.521$ ,  $df=2$ ,  $p=0.000$ ). So, the item’s solving results are influenced by the phrasing of the stem of the item.

Exhibit 4. Example of the TIMSS item,  
which after rephrase in “Lithuanian style” was solved better.

If $4(x+5)=80$ , then $x=$	TIMSS stem	Solve the equation: $4(x+5)=80$ .	Lithuanian stem
<b>True</b>	<b>50.9 %</b>	<b>True</b>	<b>59.3 %</b>
False	30.9 %	False	27.9 %
Omitted	16.6 %	Omitted	10.1 %

The other example (see Exhibit 5) shows the item, which after rephrasing into “Lithuanian style” was solved worse (the answers depend on the item format;  $\chi^2=20.43$ ,  $df=4$ ,  $p=0.000$ ). Despite the fact, that the item stem became more common for Lithuanian students, maybe after rephrase the item came to be more difficult (the item stem became much longer and complicated).

Those two examples show different effects of item’s rephrase into “Lithuanian style”.

Exhibit 5. Example of the TIMSS item,  
which after rephrase in “Lithuanian style” was solved worse.

Which of these is equal to 370·998+370·2?	TIMSS stem	Carry out the number before the parentheses: 370·998+370·2. Which expression will you get after the calculation?	Lithuanian stem
A 370·1000	47.5 %	A 370·1000	35.9 %
B 372·998	4.5 %	B 372·998	5.2 %
C 740·998	17.1 %	C 740·998	12.0 %
D 370·998·2	28.2 %	D 370·998·2	37.0 %
Omitted	2.4 %	Omitted	3.7 %

Rephrasing the item stem by writing down the fraction in words most often changed the difficulty of the item: in the most cases – the item was solved worse, in some cases – there was no difference in the results, yet in very few cases – the item was solved better. For example (see Exhibit 6), students solve the item much worse, where the fraction (1/4) is written in words (a quarter) (the answers depend on the item format;  $\chi^2=102.358$ ,  $df=5$ ,  $p=0.000$ ).

Exhibit 6. Example of the “TIMSS style” item,  
in which item stem was rephrased by writing a fraction in words.

On the first day Andrew read a quarter of a book. During the second day he read $\frac{1}{2}$ of the book, and during the third day – $\frac{1}{5}$ of the book. Which part of the book he read through these three days?	Stem with a fraction, written in words	On the first day Andrew read $\frac{1}{4}$ of a book. During the second day he read $\frac{1}{2}$ of the book, and during the third day – $\frac{1}{5}$ of the book. Which part of the book he read through these three days?	Stem with a fraction
A $\frac{2}{7}$	19.0 %	A $\frac{2}{7}$	5.2 %
B $\frac{7}{10}$	9.8 %	B $\frac{7}{10}$	4.4 %
C $\frac{3}{11}$	15.5 %	C $\frac{3}{11}$	31.0 %
D $\frac{3}{20}$	17.5 %	D $\frac{3}{20}$	4.9 %
E $\frac{19}{20}$	32.0 %	E $\frac{19}{20}$	48.0 %
Omitted	6.0 %	Omitted	5.9 %

In the last example it can be noticed that there are differences in choosing distractors between the case with “a quarter” and “1/4”. In the first case the main mistake was distractor A, which means that students for their calculations used just two fractions: 1/2 and 1/5 (those, written like fractions). Those students did not recognize “a quarter” as a number. Also almost 10 percent of students chose distractor B, which also means that students for their calculations did not use “a quarter”.

In the second case it is seen that less students chose distractors A and B. Now the main mistake was distractor C, which clearly shows the usage of all three fractions.

### Conclusions

1. The difficulty of the open-ended answer format items is higher than of the multiple-choice answer format items.
2. Students more frequently omit the open-ended answer format items than the multiple-choice answer format items.
3. The distractors of the multiple-choice items have an influence on choosing them (solving open-ended answer format items less students get a certain wrong answer, then if they would choose it in the case of a multiple-choice answer format).
4. The item’s solving results are influenced by the phrasing of the stem of the item.
5. “Lithuanian style” items are not solved better than original “TIMSS style” items.
6. The items are mostly solved worse when in the item stem used fractions are written in words.

### References

1. Abrahamson, Marty (2000). KSU Studies the effects of Multiple Choice format on the FCI, [www.bedu.com](http://www.bedu.com).
2. Elley, W. R., Mangubhal, F. (1992). Multiple-choice and Open-ended items in Reading Tests: Same or Different?, *Studies in Evaluation* 18, p.191-199.
3. Gadalla, Tahany M. (1999). Multiple-Choice versus Constructed-Response Tests in the Assessment of Mathematics Computation Skills, ERIC.
4. Hastedt, Dirk (2004). Differences Between Multiple-Choice and Constructed Response Items in PIRLS 2001 in Papanastasiou C. (2004). *Proceedings of the IRC-2004 PIRLS*. Cyprus University Press.
5. Nasser, Fadia (in Internet). On the Relationship between Test Format, Attitudes Towards and Performance in a Statistics Test, [www.stat.auckland.ac.nz](http://www.stat.auckland.ac.nz).
6. TIMSS 2003 Grade 8 Students’ Database.
7. Traub, R. E. (1993). On the equivalence of the traits assessed by multiple-choice and constructed-response tests. In Bennet, R. E. & Ward, W. C. (1993). *Construction versus choice in cognitive measurement: Issues in constructed response, performance testing, and portfolio assessment*, Hillsdale, New Jersey: Lawrence Erlbaum Associates, p.29-44.
8. Zabulionis, Algirdas (1997). Student Achievement. In Vari, Peter (1997). *Are We Similar In Math And Science? A Study Of Grade 8 In Nine Central And Eastern European Countries*, IEA, Hungary, p.99-146.