PROBLEMS
OF EDUCATION
IN THE 21st CENTURY
Volume 28, 2011
114

ON OPEN-ENDED AND CLOSED-ENDED QUESTIONS IN DIDACTIC TESTS OF MATHEMATICS

Bronislava Štěpánková, Petr Emanovský
Palacký University, Olomouc, Czech Republic
E-mail: bronislava.stepankova@upol.cz, emanovsky@inf.upol.cz

Abstract

Since educational assessment of pupils is one of the most important activities in educational process the topic of didactic tests is currently very discussed by teachers and pedagogical experts. There exist a lot of factors that have effect on quality of the didactic test. One of them is a kind of test questions (open-ended or closed-ended). In the case of closed-ended questions a pupil has to choose an answer from given variants. These questions are the most frequented form in the world and they are also quite commonly used at Czech schools. There are described the methods as well as process and results of research connected with this problem in mathematical education at elementary school there. The aim of the research was to answer the question: Are pupils more successful solving the test with open-ended or closed-ended questions? Using statistic test of significance there was validated hypothesis that the kind of questions does not play any significant role in the fruitfulness of solving the tests.
Key words: *closed-ended question, didactic test, mathematical education at elementary school, open-ended question, pedagogical research*.

Introduction

Questions about didactic tests are very discussed topic today. Particular problems of the topic are creation and using the tests as form of investigation of knowledge and skills level of pupils as well as objectivity and reliability of the tests. The discussion is consequence of early implementation of state school-leaving exam and using the tests for entrance exam at secondary schools and universities. The test represents an exam with the same conditions for all tested persons and with quantitative character of their results. The didactic test is a special kind of test for evaluation of results in educational process. There exist a lot of different definitions of didactic test concept. Some authors, e. g. Kalhous, Obst (2003), define didactic test as "set of tasks identical for certain groups of pupils". The tasks are chosen, ordered, submitted and evaluated to recognize the results of education and to know knowledge and skills of pupils in certain subject matter and time period. The didactic test has usually a written form and in comparison with a verbal exam it has some clear advantages (objectivity, the same submission and condition for all pupils, less consumed time). The didactic tests fulfill several functions in educational process - diagnostic, control, orientation, prognostic and classification.

To be effective tool for measurement of the results of education the didactic test has to have certain properties – validity, reliability, sensibility, objectivity and practicalness. According to Chráska (2006) the most important properties are validity (content, criterion-related, and construct) and reliability because they contain the other properties as their parts. The didactic test is created from particular test questions. Kinds of the questions are distinguished according to pupil answer to open-ended and closed-ended ones.

The open-ended questions are questions when a pupil creates a free answer herself or himself. Correction of that kind of questions is mostly time consuming and it can be also

PROBLEMS
OF EDUCATION
IN THE 21st CENTURY
Volume 28, 2011

subjective. To prevent subjectivity one can add certain number of points for correct and complete answer and take off certain number of points for missing or incomplete answer. Further it is possible to correct the test anonymously. These problems are divided to problems with wide answer or with brief one.

In the case of closed-ended questions a pupil has to choose an answer from given variants. These questions are the most frequented form in the world and they are also quite commonly used at Czech schools. Evaluation of that questions is less time consuming and subjective. The closed-ended questions are divided to dichotomic, with choice of answers, assignment and ordering.

Van den Heuvel-Panhuizen, M. & Becker, J. (2003) compare and contrast the didactic and psychometric models of assessment design in mathematics education. Among others, they describe a particular mathematical problems there and compare two approaches how to solve them – as problem with multiple solutions (open-ended problem) and as problem with multiple choices (closed-ended problem). Discussion concerning that topic does not have a unique conslusion because appropriate assessment problems depend very much on how one thinks about the aims of mathematics education: What mathematical ideas are important for pupils, how is mathematics learned, and how is mathematics taught? Evidently, new approach to mathematics education in many parts of the world calls for a new approach to educational assessment in mathematics (see e.g., Niss (1993); Clarke (1996); Hashimoto & Becker (1999)). A consequence of this thinking is that as mathematics content and teaching is reformed, assessment also needs to be reformed. This is not a new idea, about twenty years ago, Romberg, Zarinnia and Collis (1990) predicted a new future world-view of assessment in mathematics.

Problem of Research

The problem solved in the research was dependence of pupils achievement in mathematical didactic test at 9th class of basic school on the kind of questions (open-ended or closed-ended). The problem was consequence of discussion of teachers of mathematics who prepare pupils for entrance exam to upper secondary school. The entrance tests contain mostly the closed-ended questions. On the other hand, teachers of mathematics at elementary schools prefer tests with open-ended questions or verbal examination.

The aim of the research was to find out whether the kind of test questions has effect on pupils achievement. In case of positive answer the teachers preparing their pupils for mathematical entrance exam should change the form of didactic tests and use the closed-ended questions in the tests more.

Research Focus

Since very few teachers of mathematics use the tests with closed-ended questions pupils mostly have bigger experience with the didactic tests with open-ended questions. Therefore our objective hypothesis will be formulated as follows: „Pupils at basic school achieve better results in didactic tests with open-ended questions than with closed-ended ones."

Methodology of Research

General Background of Research

There were set the following two hypotheses for the research:
H_0: There does not exist statistically significant difference between average number of points

achieved by pupils in test 1 and test 2.

H_A: There exists statistically significant difference between average number of points achieved by pupils in test 1 and test 2.

Pedagogical experiment was chosen as research instrument for verification of the hypotheses. Pythagorean Theorem was chosen as the topic suitable for using the experiment. The experiment was realized at 9th class of basic school.

Sample of Research

The research was realized within student practices where the sample of experiment was represented by 85 pupils (49 girls and 36 boys) of 9th class of two elementary schools in Olomouc.

Instrument and Procedures

Using standardized didactic tests were obtained data for the first variant of tests (with open-ended questions, see Table 1.) and for the second variant of tests (with closed-ended questions, see Table 2.). Each test contains 10 questions intent on memorizing and understanding of knowledge and ability of their application in standard and problem situation. Both versions of the tests were pointed and numbers of points were translated to values 1-5. The standardization of the tests was realized before the experiment according to Chráska (1999). Student′s t-test, Mann-Whitney´s U-test and single-factor analysis of dispersion were used to validate the hypotheses because of the type of data. The significance level α was chosen 0,05 for all statistical tests.

Results of Research

The test of variant 1 was done by 43 pupils (25 girls and 18 boys) from Elementary school Holečkova Olomouc (numbers 1 – 21 in Table 1) and from Elementary school Hněvotín Olomouc (letters A – V in Table 1).

Table 1. Test of variant 1 (open-ended questions).

	Number of points			Number of points	
1	42	1764	A	40	1600
2	41	1681	B	40	1600
3	40	1600	C	37	1369
4	38	1444	D	36	1296
5	34	1156	E	36	1296
6	32	1024	F	35	1225
7	32	1024	G	33	1089
8	30	900	H	32	1024
9	29	841	I	32	1024
10	26	676	J	30	900
11	23	529	K	30	900
12	17	289	L	25	625
13	17	289	M	16	256
14	16	256	N	16	256

PROBLEMS
OF EDUCATION
IN THE 21st CENTURY
Volume 28, 2011

15	15	225	O	13	169
16	15	225	P	12	144
17	15	225	Q	12	144
18	12	144	R	11	121
19	12	144	S	10	100
20	11	121	T	10	100
21	2	4	U	5	25
			V	2	4
				∑1012	∑29828

The test of variant 2 was done by 42 pupils (24 girls and 18 boys) from Elementary school Holečkova Olomouc (numbers 1 – 20 in Table 2) and from Elementary school Hněvotín Olomouc (letters A – V in Table 2).

Table 2. Test of variant 2 (closed-ended questions).

	Number of points			**Number of points**	
1	46	2116	A	42	1764
2	46	2116	B	41	1681
3	41	1681	C	41	1681
4	41	1681	F	36	1296
5	38	1444	E	35	1225
9	37	1369	D	32	1024
8	35	1225	H	30	900
6	35	1225	J	25	625
10	35	1225	M	23	529
7	30	900	L	23	529
11	29	841	K	22	484
12	29	841	N	21	441
13	29	841	I	21	441
15	26	676	G	20	400
14	22	484	Q	19	361
19	19	361	U	18	324
16	17	289	P	18	324
18	13	169	O	7	49
20	9	81	R	4	16
17	4	16	S	0	0
			T	0	0
			V	0	0
				∑1059	∑33675

Student's t-test

Using the data from the tables one can compute value of t-test criterion $t = 0{,}85$ (for details see e.g. Chráska (2007)). Since the table value of the criterion $t_{0,05}(83)$ for significance

level $\alpha = 0{,}05$ and for number of degrees of freedom $f = 43 + 42 - 2 = 83$ is equal to 1,990 > 0,85 the hypothesis H_0 is accepted. It means that for this significance level there does not exist statistically significant difference between average number of points achieved by pupils in tests of version 1 and version 2.

Mann-Whitney's U – test

To verify the hypotheses H_0 and H_A we order the data in the Table 1 and Table 2 from the least value to the greatest one (see Table 3). We assign order to each point value for both versions and for pupils with the same results we assign the average order.

Table 3. Ranked data of both test variants.

Variant 1					Variant 2			
Points	Order	Points	Order		Points	Order	Points	Order
2	4,5	26	44,5		0	2	29	47,5
2	4,5	29	47,5		0	2	29	47,5
5	8	30	52		0	2	29	47,5
10	11,5	30	52		4	6,5	30	52
10	11,5	30	52		4	6,5	30	52
11	13,5	32	57		7	9	32	57
11	13,5	32	57		9	10	35	64
12	16,5	32	57		13	19,5	35	64
12	16,5	32	57		17	28	35	64
12	16,5	33	60		18	30,5	35	64
12	16,5	34	61		18	30,5	36	68
13	19,5	35	64		19	32,5	37	70,5
15	22	36	68		19	32,5	38	72,5
15	22	36	68		20	34	41	79
15	22	37	70,5		21	35,5	41	79
16	25	38	72,5		21	35,5	41	79
16	25	40	75		22	37,5	41	79
16	25	40	75		22	37,5	42	82,5
17	28	40	75		23	40	46	84,5
17	28	41	79		23	40	46	84,5
23	40	42	82,5		25	42,5		
25	42,5				26	44,5		
$n_1 = 43$		∑1758,5			$n_2 = 42$		∑1896,5	

Using the data from the Table 3 one can compute value of standardized normal U-test criterion $u = 0.7955$. Since the table value of the criterion $u_{0,05} = 1.96 > 0.7955$ the hypothesis H_0 is accepted. It means that for the significance level $\alpha = 0.05$ there does not exist statistically significant difference between average number of points achieved by pupils in tests of version 1 and version 2.

PROBLEMS
OF EDUCATION
IN THE 21st CENTURY
Volume 28, 2011

Single-factor Analysis of Dispersion

The results of both test variants are contained in Table 4 which is completed with another data necessary for the computation.

Table 4. Data for single-factor analysis of dispersion.

	Number of points	Number of pupils	Total number of points	x^2	Arithmetic average
variant 1	42 41 40 40 40 38 37 36 36 35 34	43	1012	29828	23,53
	33 32 32 32 32 30 30 30 29 26 25				
	23 17 17 16 16 16 15 15 15 13 12				
	12 12 12 11 11 10 10 5 2 2				
variant 2	46 46 42 41 41 41 41 38 37 36 35	42	1059	33675	25,21
	35 35 35 32 30 30 29 29 29 26 25				
	23 23 22 22 21 21 20 19 19 18 18				
	17 13 9 7 4 4 0 0 0	∑85	∑2071	∑63503	

There were set the following two hypotheses:

H_{02}: There does not exist statistically significant difference between dispersion within the groups and between them.

H_{A2}: The dispersion between the groups is bigger than within them.

Total sum of squares = $63503 - \frac{2071^2}{85} = 13\ 044$

Sum of squares between the groups = $\frac{1012}{43} + \frac{1059}{42} - \frac{2071}{85} = 24.38$

Sum of squares within groups = 13044 – 24.38 = 13019.62

Table 5. Results of single-factor analysis of dispersion.

Source of dispersion	Sum of squares	Degrees of freedom	Dispersion	*F*
Between the groups	24.38	1	24.38	0.1554
Within groups	13019.62	83	156.86	
Total	13044	84		

The obtained test criterion $F = 0.1554$ we compare with the critical value $F_{0,05}(1,83) = 3,97$. Since the table value of the criterion 3.97 > 0.1554 the hypothesis H_{02} is accepted. It means that for the significance level $\alpha = 0.05$ there does not exist statistically significant difference between dispersion within the groups and between them.

Discussion

The problem of open-ended and closed-ended questions is very important for surveys in many different branches. For example, in branch of agricultural economy Kealy, M.J. and Turner, R.W. (1993) tested whether open-ended and closed-ended contingent valuation mechanisms lead to significantly different results. Their test was based on joint estimation of willingness to pay responses to open- and closed-ended questions asked of the same sample of individuals. In a public good example, individuals do respond differently depending on question format. Possible explanations include different incentives for strategic behavior and respondents' lack of familiarity with the open-ended question type. No differences in willingness to pay were found in a private good example. In hospitality management Lockyer, T. (2005) investigated the factors that influence the selection of hotel accommodation by guests. Much of the previous research into this topic used a variety of closed and open question surveys. As another example can be used survey in social gerontology. Krause, N. (2002) describes strategy that can be used to improve the quality of closed-ended survey items that assess a wide range of topics in social gerontology.

The paper of Reaves, D.W., Kramer R. A. and Holmes T.P. (1998) contains results of survey in environmental economy. A three-way treatment design is used to compare contingent valuation response formats. Respondents are asked to value an endangered species (the red-cockaded woodpecker) and the restoration of its habitat following a natural disaster. For three question formats (open-ended, payment card, and double-bounded dichotomous choice), differences in survey response rates, item non-response rates, and protest bids are examined. Bootstrap techniques are used to compare means across formats and to explore differences in willingness to pay (WTP) distribution functions. Convergent validity is found in a comparison of mean WTP values, although some differences are apparent in the cumulative distribution functions. Differences across formats are also identified in item non-response rates and proportion of protest bids. Overall, the payment card format exhibits desirable properties relative to the other two formats.

Veisten, K. (2007) investigated four convenience samples comprising customers of two IKEA stores, one in England the other in Norway. The survey was done for the purpose of investigating willingness to pay (WTP) for an environmental attribute through certification and eco-labelling. Two survey-based valuation methods were applied in each store: conjoint analysis (CA) and contingent valuation (CV). In the sample of English IKEA customers responding to CA questions, extra median WTP for the eco-labelled alternative was 16% of the price of the existing unlabelled alternative. In the sample responding to CV questions, median estimate of the price premium was 7.5%. In the samples of Norwegian IKEA customers, the CA median was 2%, while the CV median was 6%. Only in the English cases did the relation between CA and CV estimates turn out as expected.

Of course, an assessment in mathematics education is very specific. The pupils have to organize the data, develop or select a model, and select a manner of notation by which they can deal with the data, make use of their knowledge of daily-life measures, reason in order to find ways for combining information, and select or develop adequate and efficient solution strategies. The assessment should make these processes explicity clear. This is particularly true for classroom assessment. Classroom teachers need detailed information on their students' mathematical thinking in order to tailor lessons appropriate for them. These requirements cannot be easily met, however, by only using results from standardized tests. By utilizing only the psychometric model for assessment design, we are depriving teachers of problems that are crucial for informing them about their students' mathematization abilities. Obviously, developing of the assessment in mathematics education will be always very important didactic topic (see e.g. Romberg, Zarinnia and Collis (1990)).

PROBLEMS
OF EDUCATION
IN THE 21st CENTURY
Volume 28, 2011

Conclusions

There were used three statistical tests of significance (Student′s t-test, U-test and single-factor analysis of dispersion) to verify weather there exists statistically significant difference between results in mathematical didactic tests with open-ended questions and with closed-ended questions at elementary school. In all three cases we accepted the zero hypothesis, i. e. there does not exist the statistically significant difference.

The objective hypothesis which supposed that the tests with open-ended questions are more suitable for pupils of 9th classes at elementary school than the tests with closed-ended questions was not validated.

Particularly, it means that the kind of test does not have effect on pupils results in mathematics. Important consequence of the results for teachers is that it does not matter which kind of test questions they use to exam their pupils. The most relevant fact for education is good understanding of a subject matter. Therefore a teacher should take notice of understandable interpretation and proper exercise. Another thing important for teachers is knowledge of didactic tests theory and ability to create their own quality tests.

Acknowledgements

This work was supported by the project CZ.1.07/2.2.00/15.0310 "Professional preparation of teachers of scientist branches for use in competitive environment".

References

Clarke, D. (1996). Assessment. In *International Handbook of Mathematics Education.* Dordrecht, The Netherlands: Kluwer Academic Publishers, p. 327-370.

Hashimoto, Y. & Becker, J. P. (1999). The Open Approach to Teaching Mathematics – Creating a Culture of Mathematics in the Classroom. In *Developing Mathematically Promising Students.* Reston, VA: National Council of Teachers of Mathematics, p. 101- 119.

Chráska, M.(2007). *The Methods of Pedagogical Research.* Praha: Grada.

Chráska, M.(1999). *The Didactic Tests.* Brno: Paido.

Ivis, F.J. & Bondy S.J. & Adlaf E.M. (1997). The Effect of Question Structure on Self-Reports of Heavy Drinking: Closed-Ended versus Open-Ended Questions. Journal of Studies on Alcohol, Vol. 58.

Kalhous, Z. , Obst, O. (2003). *Didaktika sekundární školy.* Olomouc: Palacký University.

Kealy, M.J. & Turner, R.W. (1993). A Test of the Equality of Closed-Ended and Open-Ended Contingent Valuations. *American Journal of Agricultural Economics*, Vol. 75, No. 2, p. 321-331.

Kratochvílová, L. (2007). *Didaktické testy v práci učitele matematiky na druhém stupni základní školy.* Diploma theses. Olomouc: Palacký University.

Krause, N. (2002). A Comprehensive Strategy for Developing Closed-Ended Survey Items for Use in Studies of Older Adults. *J. Gerontol. B. Psychol. Sci. Soc. Sci.* Vol. 57, No. 5, p. 263-274.

Lockyer, T. (2005). Understanding the Dynamics of the Hotel Accommodation Purchase Decision. *International Journal of Contemporary Hospitality Management,* Vol. 17, Iss: 6, p. 481- 492.

Niss, M. (1993). *Cases of Assessment in Mathematics Education.* Dordrecht, The Netherlands: Kluwer

Academic Publishers.

Reaves, D.W. & Kramer R. A. & Holmes T.P. (1998). Does Question Format Matter? Valuing an Endangered Species. *Environmental and Resource Economics,* Vol. 14, No. 3, p. 365-383.

Romberg, T. A. & Zarinnia, E. A., & Collis, K. F. (1990). A New World View of Assessment in Mathematics. In *Assessing Higher Order Thinking in Mathematics.* Washington, D. C.: AAAS, p. 21–38.

Van den Heuvel-Panhuizen, M. & Becker, J. (2003). Towards a Didactic Model for Assessment Design in Mathematics Education. In *Second International Handbook of Mathematics Education*, Dordrecht: Kluwer Academic Publishers. p. 689–716.

Veisten, K. (2007). Willingness to Pay for Eco-labelled Wood Furniture: Choice-based Conjoint Analysis versus Open-ended Contingent Valuation. *Journal of Forest Economics,* Vol. 13, No. 1, p. 29-48.

Advised by Vincentas Lamanauskas, University of Siauliai, Lithuania

Bronislava Štěpánková	PaedDr., Ph.D., Assistant Professor, Palacký University, Pedagogical faculty, Žižkovo nám. 5, 771 40 Olomouc, Czech Republic. E-mail: bronislava.stepankova@upol.cz Website: http://upol.cz
Petr Emanovský	Doc., RNDr., Ph.D., Associate Professor, Palacký University, Faculty of Science, tř. 17. listopadu 12, 771 46 Olomouc, Czech Republic. E-mail: emanovsky@inf.upol.cz Website: http://upol.cz