Students´ Knowledge, or Random Choice in ESP?

Ivana Šimonová

Abstract—As widely accepted, didactic multiple-choice tests are referred as a tool providing feedback easily and quickly. Despite the final test scores are corrected by a special formula and number of high plausibility distractors is taken into consideration, the results may be influenced by the random choice. The survey was held in three academic years at the Faculty of Informatics and Management, University of Hradec Kralove, Czech Republic, where the multiple-choice test scores were compared to the open-answer ones. The research sample included 567 respondents. The collected data were processed by the NCSS2007 statistic software by the method of frequency and multiple regression analysis and presented in the form of figures and tables. The results proved statistically significant differences in test scores in academic years 2 and 3, and were discussed from the point of the credit system and conditions for teaching/learning English in the Czech education system.

Keywords—ESP, higher education, multiple-choice test, open-answer test.

I. INTRODUCTION

THE importance of education substantially increased within the period of last 20 years and the implementation of latest information and communication technologies provided serious impact on the process of instruction. The changes were crucial, instructors were to adjust their time- and experience-tested teaching strategies to new conditions – they provided learners with more autonomous and learner-centered opportunities for learning, re-defined some of the concepts and methodologies of teaching/learning. This has been done in terms of enriching classroom activities, re-organizing course structures, changing assessment tools for evaluation learners´ performance and knowledge. The assessment is considered an inseparable part of the process of instruction. Under the conditions of i-society and e-society it is often supported by the information and communication technologies (ICT). Above all, there is always concern about the format of the assessment tool [1]. Following didactic principles applied on assessment procedures, the format of the test is pre-defined by the educational content and the main assessment objective.

Didactic tests can be used in various phases of the process of instruction. The process of testing often runs in the ICT-supported environment and electronic tests comprise fast, precise, modern and efficient form of collecting feedback from students to teachers. To be considered a highly formalized instrument assessing learners´ knowledge, the electronic test plays a unique role in the process of instruction and has become an assessment tool frequently used on all levels of education. Various types of questions and tasks can (and

Ivana Simonova is with the University of Hradec Kralove, Czech Republic (phone: 00420 49 333 2313; fax: 00420 49 5513 251; e-mail: ivana.simonova@uhk.cz).

should be) be applied, e.g. multiple choice, matching, filling the gaps, true/false, yes/no, short/long answer and many others [2].

II. PROJECT DESIGN

A. Research Problem

The problem of testing is often considered from the view of general pedagogy and didactics. Another problem appears if the aspect of students´ individual learning styles and preferences is taken into consideration. Most educators have been well aware of the factor of learning styles and have adjusted their process of instruction (*teaching*) to learners´ preferences, e.g. by implementing approaches and strategies preferred by single types of learners and/or excluding those not appreciated. But, what about another part of the process of instruction – the field of *testing* learners´ knowledge and skills? Are the "testing styles", i.e. preferences of various formats of testing, taken into account? Unfortunately, despite the assessment is recognized a crucial part of the process of instruction, teachers often tend to use tests (either standardized or non-standardized ones) of the same types for *all* learners, i.e. learners´ individual preferences in testing are not reflected at all.

Leither states teachers are pushed to make assessment more systematic, transparent, objective, so that to provide all students with the *same* conditions [3]. But – this ´fair´ treatment is the cause of the ´unfair´ conditions from the point of individual preferences in styles of learning and testing. That was why she started experimenting with giving students choices on their exams when offering the option of taking exams in the multiple-choice or open-answer format. She ran the pedagogical experiment in two phases: first, she detected students´ learning preferences by Solomon-Felder Learning Style Index; second, she correlated the data to six exam formats (multiple choice, essay, short answer, combination, true/false, and other). Briefly summarized, the group where preferences of testing style were reflected reached significantly higher test scores (the difference was 5.51 %, p < 0.05 level).

As her research sample included students of political science (88.6 %) and non-political science (11.4 %) [3], in the near future we are going to follow this pedagogical experiment focusing on sample group of bachelor and master technical and engineering IT students, particularly in an IT subject, Management and foreign language – English for specific purposes (i.e. IT English). Before we started, the question whether/to what extent the results of various test formats are comparable should be researched. Despite various researches have been done, e.g. [4]–[6], we focused on students of Informatics and relating study programmes (approx. 2,000 of

them study at the Faculty of Informatics and Management, University of Hradec Kralove).

Šimonová and Poulová detected learning preferences of students enrolled in bachelor study programme of Applied Informatics and master study programme of Information Management at the same faculty [7]. The strongest learning preferences of these (i.e. IT) students were detected as follows: sequential processors – 70%, technical – 37%, precise – 23% and confluent processors – 9%. They applied the Christine A. Johnston´s Learning Combination Inventory (LCI) [8] which (instead of others) includes the question: ´If you could choose, what would you do to show your teacher what you have learned?´ [8]. But, this promising question is not reflected in characteristics of single types of processors and preferred types of knowledge assessment.

While a multiple-choice test format (both in electronic and printed version) provides an efficient tool for assessment because it is easily (automatically) graded, instructors sometimes prefer an open-answer format questions because learners´ responses facilitate a more accurate understanding of students' thought processes and allow displaying the extent of their understanding of the subject matter. Thus, an equitable compromise between two assessment tools (formats) appears [9]. Lin and Singh presented a study investigating the relationship between students' performance on open-answer and multiple-choice questions [9]. As expected, their results verified that the correlation coefficient between open-answer response and dichotomous multiple-choice performance was higher for the question with a single stronger distractor choice compared to the question with several distractor choices.

B. Research Objective

The main objective of this research was to verify whether students reach higher level of knowledge (i.e. higher test scores) in the multiple-choice or open-answer test format.

C. Research Sample

Totally, the research sample included 567 students of the Faculty of Informatics and Management, University of Hradec Kralove (136 students in year 1 (2011/12), 159 students in year 2 (2012/13), 272 students in year 3 (2013/14), who enrolled in the first year of the bachelor study programme of Applied Informatics and master study programme of Information Management. The testing was held at the beginning of the second term.

D. Process of Research and Methodology

The process of testing ran three times, in three academic years: 2011/12, 2012/13 and 2013/14. Two types of test were used: the open-answer format was applied (Test 1, T1) and multiple-choice test format was applied providing four distractors per task (Test 2, T2).

The process of testing was run in two phases, each of them applying a different test format: first, the open-answer format was applied, i.e. respondents translated Czech sentences into English; second, the multiple-choice test format was applied with four distractors per task.

Each test contained 11 tasks valued one point for correct answer, i.e. students could reach the maximum score of 11 point. The selection of grammar items resulted from the expert analysis of 16 academicians from the Applied Linguistics Department, Faculty of Informatics and Management and Faculty of Education, University of Hradec Kralove; Slovak University of Agriculture, Nitra; Department of Language and Intercultural studies, Faculty of Education, Constantine the Philosopher University, Nitra and University of Economics, Bratislava. Following grammar topics were included in the test: task/question Q1 Present Continuous, Q2 Present Simple, Q3 Past Simple, Q4 Present Perfect Simple and Past Simple, Q5 Present Perfect Continuous, Q6 Gerund, Q7 Indirect Question, Q8 Expressing the future – going to, Q9 Sequence of Tenses, Q10 Conditional Clause, Q11 Wish Clause.

The tasks (sentences) were based on the same lexical material so that the result, i.e. knowledge of English tenses, was minimally influenced by lexical and other factors. In each sentence *Mr. Parker* did an activity, i.e. *to wash the car*. The verb *wash* and noun *car* belong to basic vocabulary acting under regular rules without any exceptions which might influence the result. This vocabulary is used in each sentence describing various situations by different tenses, e. g. *Mr. Parker is washing his car., Have you washed your car, Mr. Parker?* The visual similarity of the sentences emphasized differences in the application and translation of single tenses, which could help discover learner´s potential misconceptions in this field [10]. This intention was supported by the order of single grammar items, which has been used in most grammar books, i.e. from easier to more complicated ones, using the comparison of similar forms of various tenses (Present Simple/Continuous, Present Perfect Simple/Past Simple, Gerunds, Indirect Question, Sequence of Tenses, Conditionals etc.).

Each test was set independently, i.e. the multiple-choice test was assigned just after the translation had been finished and test sheets collected from the students.

The following hypothesis was tested:

H_1: Students will reach higher test scores in the multiple-choice test (T2) in comparison to the open-answer format (T1).

III. RESEARCH RESULTS

The data were processed by the method of frequency and multiple regression analysis by NCSS 2007 statistic software [11]. The descriptive statistics data are displayed in Table I.

Spearman´s coefficient reached significant values in all years (0.1676 in year 1; 0.4891 in year 2; 0.1233 in year 3; while critical value was 0.113 for year 1, 2 and 0.088 for year 3). These results imply that the respondents´ answers were based on good knowledge (very good knowledge in year 2, where the Spearman´s coefficient value was high), not on the random choice. T-value was similar in year 1 (1.967; crit. 1.978) and 3 (2.042; crit. 1.969) but it reached high values in year 2 (7.027; crit. 1.975).

TABLE I
DESCRIPTIVE STATISTIC RESULTS

	Year 1		Year 2		Year 3	
	OA	MCh	OA	MCh	OA	MCh
N	136	136	159	159	272	272
Mean	5.10	5.44	5.52	5.77	5.36	5.44
SD	2.4473	1.5142	2.4472	1.8857	2.4695	1.5114
Min	1	2	1	1	1	2
Max	11	8	10	10	11	8
ND	Yes	Yes	Yes	Yes	Yes	Yes
T-value	**1.967** (crit. 1.978)		**7.027** (crit. 1.975)		**2.042** (crit. 1.969)	
Cor.	0.1676 (crit. 0.113)		0.4891 (crit. 0.113)		0.1233 crit. 0.088)	
H₁	rejected		accepted		accepted	

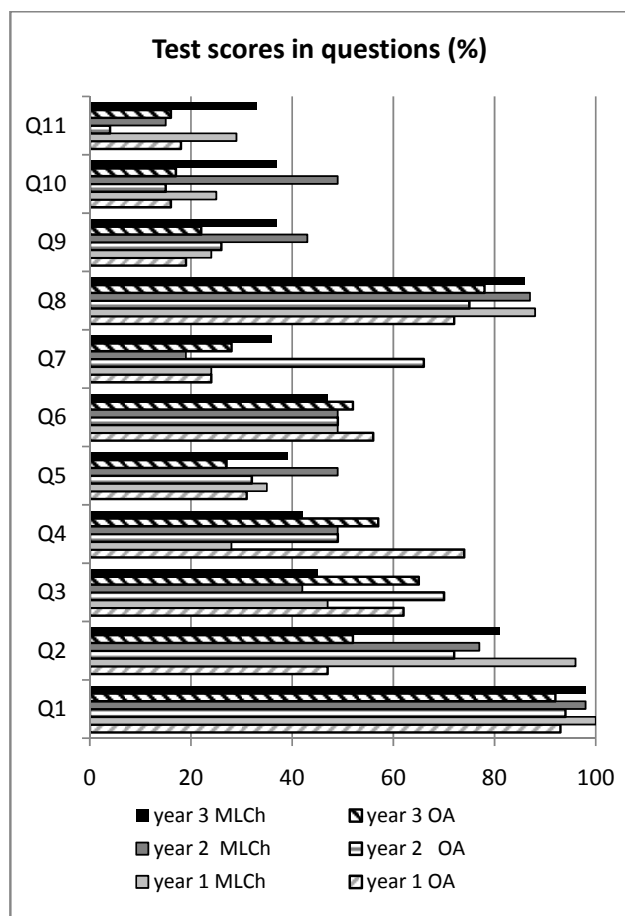ND: normal distribution; Cor.: Spearman correlation coefficient

Fig. 1 Test scores per question and year
Grammar topics - tasks: Q1 Present Continuous, Q2 Present Simple, Q3 Past Simple, Q4 Present Perfect Simple and Past Simple, Q5 Present Perfect Continuous, Q6 Gerund, Q7 Indirect Question, Q8 Future – going to, Q9 Sequence of Tenses, Q10 Conditional Clause, Q11 Wish Clause.

Thus the hypothesis (Students will reach higher test scores in the multiple-choice test (T2) in comparison to the open-answer format (T1) was verified.

Further on, test scores in single questions were considered in detail, i.e. test score in question 1 (Q1), year 1, was compared to years 2 and 3etc. The results are presented in Fig. 1: the open-answer test scores are displayed in a structured pattern, while the MLCH test scores are in full colours of the appropriate shadows (light grey for year 1, dark grey for year 2, black colour for year 3).

IV. DISCUSSIONS AND CONCLUSION

The comparison of test scores per academic year having been made, a deeper insight in single tasks would be helpful for the English teachers.

The highest test scores in all tests were reached in task Q1 (Present Continuous) and task Q8 (Expressing the future - going to), followed by task Q2 (Present Simple) where lower test scores in open-answer format are caused by omitting –s in 3.sg. and by task Q3 (Past Simple). Surprisingly, the high score also appeared in task Q4 (combining Present Perfect and Past tenses, 67 %). A reason might be this task follows the one where Present Perfect was applied. Generally, the use of this tense is often difficult for numerous students, which is the reason why special attention is paid to this tense in the first semester, and the test result might be positively influenced by this feature (mainly in year 2). Higher mean test scores in multiple-choice tests show that numerous students are able to select the correct answer from distractors (i.e. result based on good knowledge) but despite this fewer of them are not so good at (actively) using the tense correctly in the Czech-English translation.

Nearly identical test scores were reached in Q1 (Present Continuous) and Q8 (Expressing the future – going to). The reason might be these tenses are easy and most English textbooks in the Czech Republic start with them. As a consequence, learners fix them well and then, they make mistakes in situations, when present simple or Present Perfect is required using the –ing form in inappropriate context. In task Q6 (Gerund) the test scores are very close but lower than in Q1 and Q8. The result (49 – 56 % of correct answers) is surprising because this grammar item is very difficult to be used correctly by Czech learners, as the grammar rule differs substantially.

The lowest test scores appeared in task Q11 (wish clause), task Q9 (sequence of tenses) and task Q7 (indirect question). The reason might be these grammar items are more difficult in general and either do not appear in Czech language grammar, or they are applied according to different rules.

Despite the tested hypothesis confirmed our expectation (students reach higher test score in the multiple-choice test and the difference in comparison to the open-answer test scores is statistically significant in year 2 and 3), this result will lead us to further research activities in the field of "testing styles", i.e. preferred ways of assessment, as mentioned above. Generally, there exist at least two main factors relating to the foreign language teaching/learning within the Czech education system.

First, what the general level of foreign language knowledge is. What we consider to be very important in this problem is the fact that the level of knowledge of the first-year IT

students is low (lower than it has been ever before) and they did not meet the B2 level required from secondary school graduates. Differences in partial test scores and their seeming illogicality result from and reflect the current state of Czech upper secondary school system. While the grammar school graduates meet all requirements, the secondary professional school graduates (mainly those who did not have English as a subject of school leaving examination) have crucial problems. Generally, it is not exactly their guilty or fault, but partly the blame should be laid on the education system. The situation requires more from both the teachers and learners, including those who reach the adequate level. Working in classes of different level of knowledge is demanding and stressful for both the teachers, who are constantly looking for adequate didactic strategies, and for the learners, whose level of knowledge differs significantly from the requirements. Taking the university credit system into account, the weak students have enough time (18 months) to study independently (individually) and reach the required level of knowledge. Providing the motivation (both inner and outer) and didactic support can lead to succeeding in this process. The organization of study, which reflects the situation from the institutional point of view, and providing additional courses organized by the Institute of Further Education at the Faculty of Informatics and Management support those students who are really interested and willing to make efforts to meet the demands.

Second, the process of implementation of information and communication technologies on all levels of education has been running quickly in the Czech Republic and institutions providing technical (IT) education are the leaders in this process. The originally high expectations of radically increased efficiency of the process of instruction (i.e. higher increase in knowledge resulting from the ICT-supported process of learning) were not proved in numerous studies, e.g. [12]. This resulted in emphasizing other aspects which more clearly prove positive and evident evidence of the ICT contribution to the process of instruction, mainly in the affective domain of cognitive and learning processes, mainly in motivation to learning, in deepening the relation between education and learners´ success on the labour market etc. Additionally, the efficient organization of education is also included as a factor of positive contribution on all three levels: (1) in the macro-structure (study programmes), (2) in the mezo-structure (complementarity of teaching and learning in the group of learners) and in the micro-structure (learning processes of an individual). Such partial positive outcomes can be also detected among IT students. Then, the above proposed future research focus might be another contribution of ICT implementation to the process of bachelor and master education of IT specialists, not only at the Faculty of Informatics and Management, University of Hradec Kralove. In literature the multiple-choice format is generally favoured for testing receptive skills, mainly listening. Cheng emphasizes their preference to the open-answer format saying test scores are higher in multiple-choice answers due to guessing, memory constraints and the ability to predict to

some extent what is coming; the same situation is with testing grammar [13]. Linn and Miller state similar problem is the ambiguity of open answers which is impossible with multiple-choice format [14]. Above all, this format allows learners with good knowledge but poor writing abilities to reach higher test scores [15]. Opposite to this opinion, Currie and Chiramanee emphasize the possibility of guessing the correct option (25 % in four distractors) [16]. The choice can be done either by random ´picking´ one of them, or eliminating incorrect answers without knowing the answer, as Lee recommends [17]. This approach is prevented by using strongly plausible distractors, as mentioned above. DiBattista, Sinnige-Egger and Fortuna recommend the ´none of the above´ option should not be used in multiple-choice tests as it makes the test more difficult without providing a correct answer (19 % of learners wrote incorrect answer, 26 % did not show any answer) [18]. Currie and Chiramanee compared the difference between the responses in multiple-choice format to open-answer test (they call it constructed-response format) in a grammar test (as we did) [16]. Having the research sample of 152 learners, they took the translation first, followed by the multiple choice test where incorrect answers were used as distractors. Their study revealed the increase in multiple-choice test scores where only 26 % of responses were identical with the translation test format. The assumed that guessing the answer could be a factor affecting the multiple-choice test score. This result is comparable to our findings in years two and three. Above all, Burgerova and Cimermanova detected similar results when researching selected massive open online courses (MOOC) courses [19].

References

[1] W. A. Mehrens and I. L. Lehmann, *Using standardized tests in education.* OUP: Longman, 4 sub edition, 1986.

[2] P. Horovčák and B. Stehlíková, "Electronic testing and factors influencing its results," in *Acta Montanistica Slovaca,* 2007, vol. 12, no. 2, pp. 102-113.

[3] A. Leither, "Do student learning styles translate to different "testing styles?," in *Journal of political science education*, 2011, vol. 7, no. 4, pp. 416-433.

[4] D. Andrich, I. Marais and S. Hurphy, S. "Using the theorem by Andersen and the dichotomous Rasch model to access the presence of random guessing in multiple choice items," in *Journal of educational and behavioral statistics*, 2012, vol. 37, no. 3, pp. 417-442.

[5] M. N. Desrochers and J. M. Shelnutt, "Effect of answer format and review method on college students´ learning," in *Computers & Education*, 2012, vol. 59,no. 3, pp. 946-951.

[6] M. Waalkens, V. Aleven and N. Taatgen, "Does supporting multiple student strategies lead to greater learning and motivation? Investigating a source of complexity in the architecture of intelligent tutoring systems," in *Computers&Education*, 2013, vol. 60, no. 1, pp. 159-171.

[7] I. Šimonová and P. Poulová, *Learning style reflection within tertiary e-education.* Hradec Kralove: WAMAK, 2012, ch. 6.

[8] C. A. Johnston, *Unlocking the will to learn.* Thousand Oaks: Corwin Press, Inc., 1996, p. 121

[9] S. Lin and Ch. Singh, "Can multiple-choice questions simulate free-response questions?," in *Physics Education Research Conference*, Omaha, Nebraska: August 3-4, 2011, 1413, pp. 47-50, [Online]. Retrieved January 04, http://www.compadre.org/portal/items/detail.cfm

[10] T. M. Haladyna, *Developing and validating multiple-choice test items*. Lawrence Erlbaum Associates, 1994.

[11] E. Hatch and A. Lazaraton, *The research manual. Design and statistics for applied linguistics.* USA: Heinle&Heinle Publishers, 1991.

[12] D. H. Hymes, *"Toward linguistic competence,"* [Online], Retrieved November 18, 2013, http://www.aila.info/download/publications/review/AILA02.pdf.

[13] H. Cheng, "A comparison of multiple-choice and open-ended response formats for the assessment of listening performance," in *Foreign language annuals*, 2004, vol. 37, no. 4, pp. 544-555.

[14] M. Linn and M. Miller, *Measurement and assessment in teaching*. 9th edition. New Jersey: Pearson education, 2005.

[15] P. McCoubrie, "Improving the fairness of multiple-choice questions: a literary review," in *Medical teacher*, 2004, vol. 26, no. 8, pp. 709-712.

[16] M. Currie and T. Chiramanee, "The effect of multiple-choice item format on the measurement of knowledge of language structure," in Language *testing*, vol. 27, pp. 471-491.

[17] J. Lee, "Second language reading topic familiarity and test score: test taking strategies for multiple-choice comprehension questions," Ph.D thesis, 2011, in I. Ibadurrahman, *"The use of multiple-choice questions in language testing,"* Retrieved April 11, 2014, http://www.slideshare.net/IhsanIbadurrahman/pros-and-cons-of-multiple-choice-question-in-language-testing.

[18] D. DiBattista, J. A. Sinnige-Egger and G. Fortuna, "The "none of the above" option in multiple-choice testing: an experimental study," in *Journal of experimental education*, vol. 82, no. 2, pp. 168-183.

[19] J. Burgerová and I. Cimermanová, "Creating a sense of presence in online learningenvironment," *DIVAI – Proceedings of 10th International scientific conference on distance learning in applied informatics*. Wolters Kluwer, 2014, pp. 275-284.