International Journal of Engineering and Management Sciences (IJEMS) Vol. 8. (2023). No. 4.
DOI: 10.21791/IJEMS.2023.030

Testing the Sametest-effect in a BSc-level Business Communication Course Examination

András István KUN[1], JÁNOS FARKAS[2], CSILLA JUHÁSZ[3]

[1]University of Debrecen, Faculty of Economics and Business, Department of Human Resource Management, kun.andras.istvan@econ.unideb.hu
[2]University of Debrecen, Faculty of Economics and Business, Institute of Business Communication and Professional Language Studies, farkas.janos@econ.unideb.hu
[3]University of Debrecen, Faculty of Economics and Business, Department of Organization and Communication, juhasz.csilla@econ.unideb.hu

Abstract. Using secondary data, we empirically examine two biasing effects that may arise in the written evaluation of large groups of students. Suppose the students take the examination in consecutive groups, and we wish to avoid the distortion caused by tests of different difficulty. In that case, we can decide to use the same examination questions. However, the danger of the "same test effect" arises, according to which the group writing later can perform better if it receives information from the examinees in the previous round. Using the same examination tests cannot be recommended if that effect is significant. Another related potential phenomenon is the "revealed sameness effect". Accordingly, if the examinees are aware of the repetition of the questions, it significantly increases the scores of the following group. We tested these phenomena using the data of a three-round written examination. A previously published analysis of a larger sample found that the "same test effect" can be expected if the students decide in which round they take the examination. Since it was possible to freely register for the examination rounds for the assessment analyzed in this study, we assume that the "same test effect" will be significant. Based on the literature, we also expected that the "revealed sameness effect" would occur in the third round. The performed linear regression analysis (N=77) only found some weak evidence for the 'revealed sameness effect' but not for the 'same test effect'.

Keywords: Same Test Effect, Higher Education, Business Communication, Educational Assessment, Assessment Bias

Introduction

Evaluating performance is a defining task from early childhood. It is no different later, neither in secondary nor in higher education, but it is also an everyday task during work and talent management (Mándy et al., 2017; Gergely et al., 2022). In higher education, a common problem with the organization of written examinations for large groups is that it is impossible to conduct assessments simultaneously due to a lack of space, staff, or other reasons. In such cases, the obvious question is how to assemble the successive groups (candidates can freely apply for individual rounds or should we divide them according to some rule) and whether we should change the tasks and questions of the examination (Kun et al., 2021).

Overlooking management aspects such as increased time and labour and material costs caused by the compilation of different assessment instruments (assessment instruments, hereafter 'tests'), then the answer can be given by comparing the potential distortions caused by the same and various tests. It is

DOI: 10.21791/IJEMS.2023.030

not only the validity (do all tests measure the same thing) and reliability of the test (does the result of the evaluation depend not only on knowledge but also on getting into the round of tests) but also its fairness (justice, objectivity) make it essential (Csíkos and B. Németh, 2002; Rasooli, Zandi and DeLuca, 2018). In other words, the decision about the identity or difference of the tests affects both the objective and perceived subjective quality of education received by the students.

In our study, among the three main categories of test-related biases (van de Vijver and Tanzer, 2004) – construct bias (e.g. definitions and factors associated with the construct), method bias (e.g. sample, instrument or administration), and item bias (e.g. translation, explicit item content and cultural context) – this study focuses on the difference between method bias and the other two biases. Distortion falling into the construct bias category (potentially different composition and difficulty of various tests) can be avoided by trivially using the same assessment tools. Moreover, it can be handled non-trivially by thoroughly checking, comparing and fine-tuning the different tests, reducing the possibility of bias resulting from differences. However, it cannot make it zero (Csíkos and B. Németh, 2002). Of course, the latter design of the tests has a high working time cost and is usually not feasible. Is it enough to think about whether there is enough time available and who will be the test subjects? Electronic solutions can help: tests compiled from random items individually by the software can result in nearly equal chances, without systematic differences, if the tests are long enough and the test bank is large enough. However, it is not always possible to apply them. In the case of different tests, other errors (composition, typing and grammatical errors, etc.) may affect only one or the other test and thus lead to unequal conditions (item bias category).

If the students write the same test in several groups and there is a systematic difference between these groups, i.e. distortion, which is caused by something necessarily independent of the test, the test's predictive validity deteriorates (Jensen, 1980). Such biases belong to the 'procedural and situational bias' type (Jensen, 1980) within the method bias category.

If we use the same tests in successive rounds, all students face the same questions, but distortions from other sources may occur. The most important is the risk of information leaking from the first to the second round. If this happens, even to a small extent, students in the second group may gain an unfair advantage. "The increase in the test scores of the second turn originating from the identical nature of the tests will be referred to as 'same test effect'" (Kun, 2015, p. 1160).

Of course, the resulting distortion increases if the students write the tests in more than two rounds since the groups following the second have more time to utilize the possible information. On the other hand, they also realize that the questions and tasks are repeated. The latter is the "revealed sameness effect" (Kun, 2015).

According to previously cited study, aggregating 13 examination sessions in 4 subjects (N=1221), the 'same test effect' did not occur under security measures that can be considered as usual (i.e. under conditions that are neither extremely strict nor excessively lax). However, the 'revealed sameness effect' occurred in a smaller number of examinations written in three rounds (N=235). In other words, using identical tests can be a safe tool with sufficient security measures. Still, if information flows between students, it causes distortion and gives an unfair advantage (cheating) to those taking the examinations in later rounds. Based on the previous study, see above, the 'same test effect' intensifies if the students

DOI: 10.21791/IJEMS.2023.030

can choose the test round themselves. In the subsample of such students (N=201), the effect was significant, while it was not if the students were assigned to the rounds based on the list of students. We set up the hypotheses based on these conclusions, although in the case of the first one with the opposite sign to the previous results.

1. Hypotheses

According to our first hypothesis, the 'same test effect' exists, so:

H1: Compared to those who took the examination in the first round, ceteris paribus, i.e. depending solely on belonging to the round, better results are obtained in the subsequent rounds.

According to our second statement, the 'revealed sameness effect' will also be supported, i.e.:

H2: Candidates in the third round, who could learn about using the same tests from the information obtained from the candidates of the previous two rounds, achieve a significantly higher score than those who sat for the test in the previous rounds.

2. Data and methods

We relied on secondary data for the analysis: we used existing examination results after anonymization. The disadvantage is that we only had minimal background variables, and it was impossible to obtain them afterwards. Table 1 presents the available characteristics (variables) used for the subsequent analysis.

Variable	**Description**
SCORE	Student's actual score
TURN2	1 if the student took the exam in the second round, 0 otherwise
TURN3	1 if the student took the exam in the third round, 0 otherwise
GROUP_B	1 if the student wrote the "A" group test, 0 otherwise.

Table 1. Defining the variables

Note: Since those who wrote the test in group 'A' served as the basis, the dummy variable for Group 'A' is omitted.

Our data comes from the end-of-semester assessment of the 2021 fall semester of the Business communication undergraduate course. The examination was conducted in three rounds, and the students signed up for one of the three consecutive rounds. Based on the literature, we expect this to strengthen the 'same test effect'. Among the examinees were one second-year student, two third-year students, and one fourth-year student. Due to their low number, they were omitted from the analysis. All students in the sample attempted to fulfil the course for the first time. Thus, the valid element number is 77, of which 37 are written in 'A' and '40' in group 'B' (in all three rounds groups A-B, were the same). We did not have information about the gender and age of the students. Twenty-nine students took the examination in the first round, twenty-eight in the second, and twenty in the third. The examination consisted exclusively of a multiple-choice test, and one round lasted 30 minutes.

DOI: 10.21791/IJEMS.2023.030

A multivariate linear regression model was used for data analysis. The dependent variable of the model was the score achieved. We did not apply standardization because the only known and relevant group-forming variable (GROUP_B) was included in the model as a control variable.

First, we analyzed a complete model in which all the known variables were included (Model 1), and then based on this, we removed the non-significant variables. Thus, according to all variables, we created a narrowed model, significant at least at the 10% level (Model 2).

3. Results

The linear regression analysis presented in Table 2 shows no significant connection between turns and the achieved score at the 5% level. The only significant variable is the constant in both models. However, in Model 2, the *p*-value of TURN3 is slightly lower than 0,1. Thus, we have a weak clue that those writing the test in the third turn could have expected a little higher score than their fellows in the previous two turns. It should be noted that even in Model 2, the adjusted R^2 value is close to zero. The models are also presented in equations (1) and (2).

Variable	**Model 1**			**Model 2**		
	Coefficient	**Std. coeff.**	***t***	**Coefficient**	**Std. coeff.**	***t***
CONSTANT	12.8164		14.3595***	12.7807		23.4981***
GROUP_B	0.7089	0.0858	0.7508			
TURN2	-0.8069	-0.0940	-0.7362			
TURN3	1.3537	0.1438	1.1255	1.7793	0.1890	1.6672*
N	77			77		
F	1.2674			2.7797		
adj. R^2	0.0104			0.0228		

Table 2. The results of linear regression analysis.

*Notes: * p < 0.1, *** p < 0.01. all variables are entered into Model 1, but only the variables with p <0,1 into Model 2.*

The first model expressed in the form of equation is,

$$SCORE = 12.8164 + 0.7089 \times GROUP\_B - 0.8069 \times TURN2 + 1.3537 \times TURN3 + \varepsilon, \quad (1)$$

while, the second, reduced linear regression model is

$$SCORE = 12.7807 + 1.7793 \times TURN3 + \varepsilon, \quad (2)$$

where ε is the error term and the variables are defined according to Table 1.

DOI: 10.21791/IJEMS.2023.030

4. Conclusions

Our small sample of students showed no significant difference between the first and the second and between the first and the third turn. However, there is very weak support for the' revealed sameness effect' because TURN3 is positive and significant at the 10% level. In other words, even if examinees in the second round gain no advantage, those writing in the third round might do. And this could be a consequence of some information about the test as in the previous two groups and its recurrence.

Our result, therefore, confirmed the experience of previous research (Kun, 2015) since we did not find evidence for the existence of the 'same test effect', under natural examination conditions, in the case of a new subject and a new group of students.

For the sake of correct interpretation, we must also note that our research was conducted on a sample with a small number of items. Therefore we also included the result currently significant at 10% since it is especially indicative because it would be the same as the experience of the previous, to which similar research referred. In addition to the number of elements, a severe limitation is the narrow range of available background variables. In the referenced antecedent research, gender, the number of examination admissions, the grade, and the major significantly influenced the result. Of these, in the present study, our population was homogeneous according to the number of examination admissions, grades and major, but we had no information on the students' gender.

Based on the fact that the repeated test did not confirm the existence of the 'same test effect' but found evidence for the 'revealed sameness effect', by analyzing a total of 14 test occasions of 5 subjects together with the previous one, we can already draw a few conclusions. The most straightforward point is that writing the same test can be used to avoid distortions caused by different tests if basic safety measures are used (not much time passes between rounds, students are not allowed to use phones, etc.), but only if the students are not aware of it (so, for example, we do not use it in every examination), or if we use it in no more than two rounds. The third round showed an upward bias in the score achieved in both studies. We emphasize that we did not take into account the economic benefits, i.e., we do not make our proposal because of the time saved for the examination preparers. Of course, a certain amount of time is saved, but from the point of view of educational ethics (Barizsné and Filep, 2022), it could not justify the potential bias resulting from the same tests. We recommend not discarding the careful use of the same tests because we have not found evidence that it causes bias. Still, it can reduce other biases (mainly those resulting from the different challenges caused by various tests).

References

[1] Barizsné, H.E and Filep, R. (2022) ' Mennyire vagyunk etikusak? és mások?' *In: Dajnoki, K. and Boros, J. (ed.) Tanulmányok Berde Csaba tiszteletére: [Tanulmánykötet]* Debreceni Egyetem Gazdaságtudományi Kar VESZI, Debrecen, 328 p. pp. 77-83. , 7 p.

[2] Csíkos, C. and B. Németh, M. (2002) 'A TESZTEKKEL MÉRHETŐ TUDÁS', in Csapó, B. (ed.) Az iskolai tudás. Budapest: Osiris Kiadó, pp. 91–122.

DOI: 10.21791/IJEMS.2023.030

[3] Gergely, É., Szántó, G. and Szondi, R. (2022) 'A tehetségmenedzsment nyomában' In: Dajnoki, K. and Boros, J. (ed.) Tanulmányok Berde Csaba tiszteletére: [Tanulmánykötet] Debreceni Egyetem Gazdaságtudományi Kar VESZI, Debrecen, 328 p. pp. 133-144. , 12 p.

[4] Jensen, A. R. (1980) 'Précis of Bias in Mental Testing', Behavioral and Brain Sciences, 3(3), pp. 325–333. doi: 10.1017/S0140525X00005161.

[5] Kun, A. I. (2015) 'Does the repetitive use of the same test in consecutive examination sessions facilitate cheating?', Assessment & Evaluation in Higher Education, 40(8), pp. 1159–1181. doi: 10.1080/02602938.2014.977845.

[6] Kun, A.I., Kotsis, Á. and Boros, J. (2021) 'A dunning-kruger-hatás a felsőoktatási szakképzésben : online feleletválasztós tesztek vizsgálata', In: Molnár, Gy. and Tóth, E. (ed) A neveléstudomány válaszai a jövő kihívásaira : XXI. Országos Neveléstudományi Konferencia Szeged, 2021. november 18-20. : program, előadás összefoglalók Szeged, Magyarország : MTA Pedagógiai Tudományos Bizottság, SZTE BTK Neveléstudományi Intézet 690 p. p. 535

[7] Mándy, Zs., Kun, A., Boros, J. and Dajnoki, K. (2017) PhD-hallgatók elégedettsége és a képzésre való jelentkezés motivációs tényezőinek összefüggései, In: Molnár, Gy. and Tóth, E. (ed) A neveléstudomány válaszai a jövő kihívásaira : XXI. Országos Neveléstudományi Konferencia Szeged, 2021. november 18-20. : program, előadás összefoglalók Szeged, Magyarország : MTA Pedagógiai Tudományos Bizottság, SZTE BTK Neveléstudományi Intézet 690 p. p. 535

[8] Rasooli, A., Zandi, H. and DeLuca, C. (2018) 'Re-conceptualizing classroom assessment fairness: A systematic meta-ethnography of assessment literature and beyond', *Studies in Educational Evaluation*, 56, pp. 164–181. doi: 10.1016/J.STUEDUC.2017.12.008.

[9] Van de Vijver, F. and Tanzer, N. K. (2004) 'Bias and equivalence in cross-cultural assessment: an overview', *European Review of Applied Psychology*, 54(2), pp. 119–135. doi: 10.1016/J.ERAP.2003.12.004.