Психологическая наука и образование
2021. Т. 26. № 1. С. 76—91
DOI: https://doi.org/10.17759/pse.2021260105
ISSN: 1814-2052
ISSN: 2311-7273 (online)

Psychological Science and Education
2021. Vol. 26, no. 1, pp. 76—91
DOI: https://doi.org/10.17759/pse.2021260105
ISSN: 1814-2052
ISSN: 2311-7273 (online)

Academic Outcomes of Students in University Digital Environment at Different Levels of Higher Education: Who Is More Successful?

Marina G. Sorokova
Moscow State University of Psychology & Education, Moscow, Russia,
ORCID: http://orcid.org/0000-0002-1000-6487, e-mail: sorokovamg@mgppu.ru

The problem of empirical assessment of various aspects of learning in the digital educational environment seems to be of particular relevance. At the same time, it is emphasized that there is a lack of comparative studies of students' educational outcomes at different levels of higher education who have completed e-courses. The study was conducted at the Moscow State University of Psychology and Education (MSUPE), the sample size is N = 424 students. Subject of the study is immediate and long-term academic achievements of students in two levels of higher education who completed e-courses. Purpose of the study is to evaluate the differences in the academic achievements and knowledge retention of graduate and undergraduate students. Key findings of students' academic outcomes comparative analysis are the following: (1) No differences were found between graduate and undergraduate students in the pretest, final test and overall e-course grade indicators. (2) The same tendency was revealed in students of both groups: pretest scores are low, posttest scores significantly and strongly increase, and then after 1.5 - 4 months they significantly decrease, while remaining significantly higher than the input indicators. The knowledge retention scores are very scattered in comparison with the direct ones. (3) The gain score effect size and the improvement index are significant for the final test only without adjusting for clustering, i.e. ignoring the fact that the sample consists of several student groups. A median graduate student would have a higher score than a median undergraduate student. Cluster-level effect size is not statistically significant. Cluster-level effect size for overall e-course grade indicators with difference-in-differences adjustment is also not reliable. (4) The knowledge retention scores in both students' categories do not differ. The gain score effect sizes for knowledge retention, taking into account both the final test and the pretest, are not significant. (5) The psychometric characteristics of the academic achievement test in the field of empirical data quantitative analysis can be considered satisfactory.

Keywords: blended learning, flipped classroom model, e-course, mass open online course (MOOC), digital technologies in education, university digital environment, higher education, academic outcomes, knowledge retention, effect size, improvement index.

Funding. The reported study was funded by the Moscow State University of Psychology and Education (MSUPE) in the framework of the research project "Digital Technologies in Higher Education: Development of Technology for Individualizing Education Using E-Courses".

For citation: Sorokova M.G. Academic Outcomes of Students in University Digital Environment at Different Levels of Higher Education: Who Is More Successful? *Psikhologicheskaya nauka i obrazovanie = Psychological Science and Education*, 2021. Vol. 26, no. 1, pp. 76—91. DOI: https://doi.org/10.17759/pse.2021260105 (In Russ.).

Предметные результаты студентов в цифровой среде университета на разных уровнях высшего образования: так кто же более успешен?

Сорокова М.Г.
ФГБОУ ВО «Московский государственный психолого-педагогический университет» (ФГБОУ ВО МГППУ),
г. Москва, Российская Федерация
ORCID: https://orcid.org/0000-0002-1000-6487, e-mail: sorokovamg@mgppu.ru

Отмечается, что проблема эмпирической оценки различных аспектов обучения в цифровом образовательном пространстве приобретает особую актуальность. В то же время подчеркивается, что существует дефицит сравнительных исследований образовательных результатов студентов на разных уровнях высшего образования, завершивших электронные курсы. Представлены результаты исследования, проведенного на выборке объема N=424 студентов (обучающиеся в Московском государственном психолого-педагогическом университете). Предметом работы были непосредственные и отдаленные образовательные результаты студентов двух уровней высшего образования, завершивших электронные курсы. Целью проведенного исследования было оценить различия в образовательных результатах студентов магистратуры и программ второго высшего образования, с одной стороны, и студентов бакалавриата и специалитета программ первого высшего образования — с другой. Сравнительный анализ результатов показал следующее: 1) различий между двумя категориями студентов в результатах входного теста, итогового теста и общей оценке за электронный курс не выявлено; 2) обнаружена одна и та же тенденция у студентов обеих категорий: на входе результаты по тесту достижений низкие, на выходе они достоверно и сильно возрастают, а затем через 1,5—4 месяца достоверно снижаются, оставаясь при этом достоверно выше входных результатов. При этом отдаленные результаты сильно рассеяны по сравнению с непосредственными; 3) размер эффекта и индекс улучшения достоверны для итогового теста с учетом результатов входного теста без поправки на кластеризацию, т.е. игнорируя тот факт, что выборка состоит из нескольких студенческих групп. Медианный студент программ магистратуры и второго высшего образования имел бы более высокий результат, чем медианный студент в группе первого высшего. Размер эффекта с коррекцией на кластеризацию статистически не значим. Размер эффекта для общей оценки за электронный курс с учетом входного теста также не достоверен; 4) отдаленные результаты обеих категорий студентов не различаются. Размеры эффекта по отдаленным результатам с учетом как итогового теста, так и с учетом входного теста недостоверны; 5) психометрические характеристики теста достижений в области количественного анализа эмпирических данных можно считать удовлетворительными.

Ключевые слова: смешанное обучение, модель «перевернутый класс», электронный курс, массовый открытый онлайн-курс (МООК), цифровые технологии в образовании, цифровая среда университета, образовательные результаты, отдаленные результаты, размер эффекта, индекс улучшения.

Финансирование. Работа выполнена при финансовой поддержке ФГБОУ ВО «Московский государственный психолого-педагогический университет» в рамках научно-исследовательского проекта «Цифровые технологии в высшем образовании: разработка технологии индивидуализации обучения средствами электронных учебных курсов».

Для цитаты: *Сорокова М.Г.* Предметные результаты студентов в цифровой среде университета на разных уровнях высшего образования: так кто же более успешен? // Психологическая наука и образование. 2021. Том 26. № 1. С. 76—91. DOI: https://doi.org/10.17759/pse.2021260105

Introduction

Currently, the digitalization of education has become global. Thanks to modern digital technologies, universities around the world interact in a network form, develop their own e-courses and use online courses of other universities, increasing the accessibility and quality of education.

Analyzing current trends in education, B. Williamson [22] stresses the influence of “platform capitalism” or “surveillance capitalism” represented by online educational program management platforms, digital learning platforms and intensively using data analytics to realize their strategic business priorities on the landscape of higher education. According to him, these forms of digital capitalism are merging with existing political demands for universities to become more data-driven, competitive, and market-focused. P. Prinsloo [17] in a review of Williamson’s book cites the author’ point of view on the role of datafication and digitization in education: they point to how “software and digital data are becoming integral to the ways in which educational institutions are managed, how educators’ practices are performed, how educational policies are made, how teaching and learning are experienced, and how educational research is conducted” [17, p. 183].

In Russia, online education is one of the priority areas of public policy. The state program of the Russian Federation “Development of education” for 2018 — 2025 includes the implementation of the federal project “Digital educational environment”. The project aims to “create conditions for the introduction by 2024 of a modern and secure digital educational environment that ensures the formation of self-development and self-education value among students of educational institutions of all types and levels by updating the information and communication infrastructure, training personnel, creating federal digital platform” [1].

Universities that intend to be powerful educational clusters must prepare interactive courses with elements of distance learning [7]. The modern educational paradigm involves the creation of smart universities in order to provide each student with the opportunity to build an individual profile of competencies with which he will enter the labor market in the digital economy and will be in demand there [2]. And even the external conditions associated with the force majeure circumstances of the spread of pandemics of viral infections force universities to completely switch to distance learning formats as soon as possible. Thus, the problem of the empirical assessment of various aspects of learning in the digital educational space is of particular relevance.

Previous studies findings

In international studies, various aspects of the digitalization of education are actively explored. One of the most important issues is to assess the impact of blended learning and “flipped classroom” model on students’ educational outcomes. A.A. Margolis [3] characterized various models of blended learning and presented the analytic review of international studies of its effectiveness in general education compared to traditional full-time and distance learning. The results are controversial: a number of studies confirm its advantages, others do not, the research design is also criticized. Note that blended learning is a combination of full-time study with digital and online formats.

Currently, one of the most popular approaches in blended learning is the “flipped classroom” (FC), which involves a combination of extracurricular self-study of students using lectures video recordings and a variety of online tutorials and teaching materials with face-to-face sessions aimed at updating the independently studied content and developing the desired competencies using interactive methods [3]. A number of empirical studies are devoted to assessing its’ effectiveness in higher education in various aspects when studying various courses, for example, English [9], physiotherapy [19], management [18].

In the paper [10], self-efficacy, autonomy, and academic workload, as well as knowledge retention of students studying a course of hematology and oncology using the FC, were compared with traditional full-time study. There were no differences in the workload and autonomy indicators, but students’ self-efficacy in the FCs was significantly higher, and the time to prepare for the exam was significantly less. 10 months after the course, long-term outcomes and self-efficacy showed no differences.

The study [16] emphasizes the importance of categorizing assessment tasks in accordance with Bloom’s taxonomy for evaluating the FC-effect. It was shown that the academic achievements of students studying anatomy in the FC are significantly better than students in the traditional full-time approach when performing tasks requiring a higher level of knowledge, i.e. analysis, and there are no differences in the results of tasks requiring a lower level cognition, i.e. knowledge or its application.

The structure of e-courses includes different assessment types, and from this point of view, the study findings [13] of the motives of students using online formative assessments (OFAs) when preparing for summative assessments are important. Formative assessments intend to provide feedback on student performance in order to improve and accelerate learning. Revealed motives for using OFAs are the collection of information, obtaining feedback or direct communication. The main reasons for abandoning the use of OFAs were lack of time and having completed the questions before, but they can also be associated with the students themselves, teachers or fashion.

Analytical review of the advantages and disadvantages of the FC [8] shows that the most frequently reported advantage of this model is improved student performance. Most of the problems of this model are associated with extracurricular activities, for example, inadequate preparation of students for full-time sessions.

Very interesting study of the effectiveness of the FC-approach based on massive open online courses (MOOCs) [21] shows that students in the FC based on MOOCs on average showed better academic achievements in the inorganic chemistry course than in the traditional class. In addition, most FC-students received favorable experience in terms of interaction with group students, the availability of learning materials and the results of active learning. The authors discuss the possibilities of using common MOOCs in traditional university curriculums.

In our pilot empirical study [6] the potential of the e-course “Mathematical Methods in Psychology” as a digital educational resource of blended learning using the FC is examined, reliable relationship between the students’ positive assessment of their educational achievements and their positive attitude to the new format is confirmed . The students’ academic outcomes after completing the e-course significantly improved. The FC-students final assessments are on average significantly higher than in the group of traditional full-time learning. A comparative analysis of perceived learning experiences in the e-course of graduate and under graduate students was carried out by us in [5].

Of particular interest are the results of a meta-analysis conducted by the authors of articles [11], [12] and [14], since they analyzed not only scientific papers, but also electronic databases provided by their authors or available in data repositories.

In a meta-analysis [11], aimed at testing the FC-effectiveness compared to the traditional approach in medical education, this model is evaluated as a promising learning approach to increase student motivation and interest. However, the influence of the FC on changes in knowledge and skills was less convincing, since the effect sizes ranged from d = -0.27 to 1.21 with a median of 0.08. The different effect size direction and its' modulus testified to the lack of convincing evidence of the FC-effectiveness in relation to academic achievements beyond traditional teaching methods. The need for further studies of the long-term impact of this model with respect to knowledge retention and the transfer of knowledge into professional practice and patient care is emphasized. Another meta-analysis of comparative studies conducted in [14] showed a general statistically significant effect in favor of the FC compared to the traditional approach for teaching professions in the field of medicine and healthcare. The emphasis was on research, where students were given video lectures before face-to-face sessions. The effect sizes were also calculated, and possible moderators and systematic publication errors were analyzed.

A meta-analysis [12] of studies on a sample of students studying both medicine and health care sciences, as well as sciences from other areas, showed that the outcomes of FC-students were significantly better than traditional full-time approach students, in exam scores (after intervention between groups and as changes in indicators before and after the intervention) and in course grades, but not in objective structured clinical examination scores. Analysis of the subgroups showed that the advantage of FC was not observed in RCTs, non-USA countries, as well as in earlier years of publication (2013 and 2014). Cumulative analysis and meta-regression suggested a tendency to gradually improve the results by year.

We did not find comparative studies of the educational outcomes of students at different levels of higher education who completed e-courses with FC-approach.

Methods

The quasi-experimental study was carried out at the Moscow State University of Psychology and Education (MSUPE) in the fall semester of 2019 and in the spring semester of 2020 in the framework of the research project “Digital Technologies in Higher Education: Development of Technology for Individualizing Education Using E-Courses”, website project https://dthe.mgppu.ru. The sample consists of N1 = 234 students attending master's programs and the second higher education programs (IG1, graduate students) as well as N2 = 190 students attending undergraduate programs and program majors (IG2, undergraduate students) covering program tracks in psychology and education. Total sample size is N = 424 students relating to 23 academic groups of 6 faculties of the university. All students completed e-courses developed by us: the e-course “Statistical and Mathematical Methods in Psychological and Educational Researches” for masters' level, and the e-course “Mathematical Methods in Psychology” for undergraduate level. Both e-courses are hosted on the LMS Moodle platform and are available at https://e-learning.mgppu.ru.

We collected contextual data about the sample using the feedback survey, accessible after the completion of the e-course. 28 students did not fill out this survey, therefore, we present the testees' socio-demographic characteristics according to the sample of N = 396 students, of which N1 = 213 students belong to IG1 and N2 = 183 students belong to IG2. There are no gender differences between the groups ($p = 0.613$): there are 21.1% of men and 78.9% of women in IG1, and 18.6% of men and 81.4% of women in IG2. Both groups significantly differ in age (Chi-square test, $p < 0.001$). The IG1 group is mainly adults: 16.0% are students aged 20—24, 12.2% are 25—29 years old, 29.1% are 30—34 years old and 42.7% are 35 years old and older, while in IG2 youth predominates — 16.9% under the age of 20 years, 81.4% — 20—24 years old, and only 1.6% are 25 years old and older. Both groups also significantly differ in the nature of employment (Chi-square, $p < 0.001$). Compared to IG2, in IG1, work is related to the program track they attend in 51.2% vs 5.5% students, not connected — in 32.9% vs 41.5%, and 16.0% vs 53.0% currently do not employ at all.

Both e-courses are aimed at developing competencies and basic skills for quantitative analysis of empirical data in research and scientific and practical activities in SPSS and consist of 3 identical modules regarding basic methods of mathematical statistics, and the master's course also has the 4th additional module "Multidimensional Statistical Methods " for advanced students. We compared perceived experiences and academic outcomes of students who completed 3 mandatory modules. Educational outcomes were evaluated using 5 tests inside the e-course, i.e. pretest, 3 learning tests inside the modules, final test, and an individual case-task including 6 cases. Cases in different case-tasks varied in data sets. Students performed case-tasks in SPSS, we evaluated and commented on them. Students who completed the e-course filled out an anonymous feedback survey. After 1.5—4 months, students participated at full-time testing again at the Department for Monitoring the Quality of Professional Education (DMQPE) at the MSUPE to evaluate knowledge retention. The pretest, the final test and the test at DMQPE are the same.

Both e-courses were studied in blended learning FC-format, which implies a transition from teacher-centered approach to student-centered learning management approach using the LMS Moodle platform. The videolectures were offered to students for independent pre-class preparation along with presentations, videos demonstrating case solving in SPSS, data files and output files, hyperlinks to textbooks in electronic library and articles in scientific journals of the PsyJournals portal https: //psyjournals.ru/, illustrating the application of the studied methods in real researches. At the face-to-face sessions, students, individually using presentations as a guideline, updated information: answered instructor questions, participated in group discussions of the most difficult issues, but, most importantly, they solved in SPSS authentic cases concerning researches in psychology and education, learned to choose methods, analyze data and interpret the results. We supported the interaction and mutual assistance of students.

Subject of the study: immediate and long-term academic achievements of students in two levels of higher education who completed e-courses.

Purpose of the study: to evaluate the differences in the academic achievements and knowledge retention of graduate and undergraduate students.

Tasks: 1. compare the outcomes of the pretest, the final test and the overall e-course grade for students of both categories and identify their similarities and differences; 2. assess the differences between the students of both categories in knowledge retention; 3. evaluate effect sizes and improvement indices for the measured parameters; 4. check the psychometric characteristics of the achievement test.

Research question: how do the educational outcomes on the e-courses in quantitative analysis of empirical data correlate among graduate and undergraduate students of psychology and education areas? Which student category is more successful?

The research database is publicly available in the Mendeley Data repository [20] and is available for download. Data analysis was performed in SPSS V23 using the methods of descriptive statistics, Mann-Whitney test, Wilcoxon test, binomial test, Spearman's rank correlation coefficient [4]. The effect sizes and improvement indices were calculated using the WWC Version 4.1 Procedures Handbook methodology presented on the IES What Works Clearinghouse portal [15].

Results

First, we compared the educational outcomes of graduate (IG1, N1 = 234) and undergraduate students (IG2, N2 = 190) on 3 parameters, i.e. according to the pretest, final test and the overall e-course assessment. The comparison results by the Mann-Whitney test are shown in Table 1.

Table 1 shows that neither the pretest nor the final test scores revealed differences between the students of both groups. Graduate students showed significantly better outcomes in the overall e-course assessment ($p < 0.05$), however, in absolute terms, the means difference is small (M1 = 84.12 vs M2 = 82.93). The standard deviations (SD1 vs SD2) for each of the 3 parameters in both groups are approximately the same, therefore, the scattering of test scores near group means also does not differ. Note that in IG1, eight students did not perform the pretest, and four other students completed the e-course without passing the final test, so the sample sizes of IG1 students in these parameters are 226, 230 and 234 students, respectively, while the sample size of IG2 did not change and amounted to 190 students.

To evaluate the impact of the intervention, i.e. training in e-course, in both groups on the academic outcomes, we also compared the pretest and final tests indicators with each other according to the Wilcoxon test (Table 2).

As can be seen from Table. 2, the differences between the final test and pretest indicators are highly significant ($p < 0.001$) in both groups, and the indicators of the final test are on average significantly higher by 50.95 percentage points in IG1 and by 48.66 in IG2. The standard deviation in IG1 did not change, while in IG2 it increased only slightly, which indicates approximately the same scattering of test scores around group means and a comparable uniformity of the

Table 1

Educational outcomes comparison of 2 students groups on the pretest, the final test and the overall e-course assessment according to the Mann-Whitney test (N = 424)

Parameter	IG1 graduate students		IG2 undergraduate students		Mann-Whitney U	p — value
	M1	SD1	M2	SD2		
Pretest	33,90	10,83	34,78	9,82	20270,5	0,326
	N1 = 226		N2 = 190			
Final test	84,85	10,24	83,44	11,18	20342	0,233
	N1 = 230		N2 = 190			
Overall e-course assessment	84,12	9,94	82,93	8,36	19592,5	0,036*
	N1 = 234		N2 = 190			

***** differences are statistically significant at $p < 0,05$

results at the input and output. The sample size of IG1 was reduced to N1 = 222 (see Table 1) due to those 12 students who did not pass at least one of these 2 tests.

However, the most interesting for us was to compare the knowledge retention after 1.5—4 months after the e-course completion. If students performed the pretest and final test inside the e-course independently without external control, then the long-term results were tested at the Department for Monitoring the Quality of Professional Education (DMQPE) in person in the presence of an employee of this department, so the outcomes of this test can be considered an independent assessment. Testing at DMQPE was not carried out by all students who completed the e-course, therefore, when analyzing the knowledge retention results, they dropped out, and the sample sizes were reduced to N1 = 149 (IG1) and N2 = 139 (IG2). Since baseline equivalence could be broken due to dropouts, we again compared both groups according to 3 parameters (Table 3), as well as indicators of the pretest, the final test and the test at DMQPE in each group separately (Table 4).

Table 3 reflects the absence of differences between students of IG1 and IG2 for all 3 studied parameters.

Table 4 shows the same trend in both groups: the pretest' indicators are low, at the output they significantly increase on average by ca 50 percentage points, and the knowledge retention indicators are on average significantly lower by ca 30 percentage points than the final test scores, but still significantly higher than the pretest indicators. For all 3 comparisons, the differences were significant at $p < 0.001$. Note that the knowledge retention indicators are strongly scattered around group means: the standard deviation for the test at DMQPE is almost twice as high as for the indicators of the pretest and final test.

Effect sizes and Improvement indices

In international studies, it is commonly used effect size indices, improvement coefficients and their p-values to evaluate the

TTable 2

Educational outcomes comparison for the pretest and the final test with each other in each of the 2 groups of students according to the Wilcoxon test (N = 412)

Group	Test	Mean M	Standard deviation SD	Wilcoxon Z	p — value
IG1 graduate students (N1 = 222)	Pretest	33,90	10,83	-12,918	0,000***
	Final test	84,85	10,24		
IG2 undergraduate students (N2 = 190)	Pretest	34,78	9,82	-11,952	0,000***
	Final test	83,44	11,18		

*** differences are statistically significant at p < 0,001

Table 3

Educational outcomes comparison of 2 groups students on the pretest, the final test and the test at DMQPE according to Mann-Whitney (N = 288)

Parameter	IG1 graduate students		IG2 undergraduate students		Mann-Whitney U	p — value
	M1	SD1	M2	SD2		
Pretest	33,00	9,47	34,12	9,81	10049	0,664
Final test	85,50	9,85	83,12	11,18	9002	0,082
Test at DMQPE	57,16	18,95	55,84	17,85	9945,5	0,561

Table 4

Comparison of immediate educational outcomes and knowledge retention in each of 2 groups of students by Wilcoxon test (N = 288)

Group	Test	Mean M	Standard deviation SD	Min	Max	p — value
IG1 graduate students (N1 = 149)	Pretest	33,00	9,47	1	56	0,000***
	Final test	85,50	9,85	56	100	
	Test at DMQPE	57,16	18,95	16	97	
IG2 undergraduate students (N2 = 139)	Pretest	34,12	9,81	1	66	0,000***
	Final test	83,12	11,18	53	100	
	Test at DMQPE	55,84	17,85	22	97	

*** differences are statistically significant at p < 0,001

intervention effectiveness. Compared to other statistical criteria, these indices have the following advantages. Firstly, they are expressed in standard units, which allows you to compare the intervention effects for different tests. Secondly, they allow you to calculate effect size both without taking into account the input slice in the 2 compared groups, and taking into account the input data (''difference-in-differences adjustment''), i.e., gain score effect size. Thirdly, the methodology allows you to calculate the effect of the intervention, taking into account the quasi-experimental design (cluster-level effect size), i.e., given that the sample is composed of several subsamples — student groups, data within which may be more related. Finally, translating effect sizes into coefficients or improvement indices allows them to be clearly interpreted. We calculated the effect sizes and improvement indices according to the WWC Version 4.1 Procedures Handbook methodology presented on the IES What Works Clearinghouse portal [15].

The effect size was calculated according to g Hedges' formula, where IG1 (graduated students) was considered as a conditional "intervention group", and iG2 (undergraduate students) as a "comparison group". The results of calculating the effect sizes and improvement indices for the data of the pretest, the final test (posttest) and the overall e-course assessment for IG1 as compared to IG2 for the whole sample are presented in Table 5.

As Table 5 shows, the effect sizes for IG1 in comparison with EG2 were first calculated according to the pretest (g = — 0.085), the final test (g = 0.132) and the overall e-course assessment (g = 0.128). They reveal that at the baseline the graduate students had on average slightly lower scores, than undergraduate students, but after the e-course completion their scores were already higher. The effect is not significant even unadjusted for clustering (p = 0.164 and p = 0.178 for output indicators).

Next, the gain score effect size and its significance level for the final test was calculated taking into account the pretest (g1 = 0.214, p = 0.030) without adjustment for clustering, i.e., ignoring the fact that the combined sample consists of several subsamples — student groups. Since the effect is significant, for a better interpretation, we calculated the improvement index (8.47%). The improvement index is the difference in the percentile rank between the middle member of the "intervention group" and the middle member of the "comparison group" in the distribution of the "comparison group". In our case, the improvement index means that the median graduate student would be 8.47 percentage points to the right of the median undergraduate student (which, by definition, has the 50th percentile), i.e. he would

Table 5

Effect sizes and improvement indices for the pretest, the final test and the overall e-course assessment data for IG1 as compared to IG2 for the whole sample (N = 424)

Group, Index	Statistics	Test			Notes
		Pretest	Final test	Overall e-course assessment	
IG1 graduate students	Mean M1	33,90	84,85	84,12	IG1 is taken as an "intervention group"
	Std. Deviation SD1	10,83	10,24	9,94	
	N1	226	230	234	
IG2 undergraduate students	Mean M2	34,78	83,44	82,93	IG2 is taken as a "comparison group"
	Std. Deviation SD2	9,81	11,18	8,36	
	N2	190	190	190	
Effect size	g Hedges	- 0,085	0,132	0,128	Calculated by Hedges' g and SE(g) formulas not taking pretest into account
Effect size standard error	SE(g)	0,098	0,098	0,097	
t -statistics	t		1,391	1,349	Unadjusted for clustering
p-value	p		0,164	0,178	
Effect size	g1	0,214			The final test scores in 2 groups were compared taking into account the pretest. The gain score effect size was calculated.
Effect size standard error	SE(g1)	0,129			
Improvement index 1	U3 — 50%	58,47% — 50% = 8,47%			
t -statistics	t	2,172			Unadjusted for clustering
p-value	p	0,030*			
t-statistics corrected for clustering	t_a	1,023			Clustering correction was calculated, since the effect size without adjusting for clustering is significant
Clustering-corrected statistical p value	p	0,307			
Effect size	g2			0,143	The overall e-course assessment in 2 groups were compared taking into account the pretest. The gain score effect size was calculated.
Effect size standard error	SE(g2)			0,125	
Improvement index 2	U3 — 50%			55,68% — 50% = 5,68%	
t -statistics	t			1,456	Unadjusted for clustering
p-value	p			0,146	

have a rank of 58.47 in this group, reflecting a higher result. Since the effect is significant, according to the WWC methodology, cluster-level effect size and its p-value (p = 0,307) was calculated, and it turned out to be statistically insignificant. The effect size for the overall e-course assessment, taking into account the pretest, was also not reliable (g2 = 0.143, p = 0.146).

The Table 6 shows the effect sizes and improvement indices for the pretest, final test and knowledge retention data for reduced sample, i.e. with drop-out of those students who did not executed testing at DMQPE.

Table 6

Effect sizes and improvement indices for the pretest, final test and knowledge retention data in IG1 as compared to IG2 for reduced sample (N = 288)

Group, Index	Statistics	Test: Pretest	Test: Final test	Test: Test at DMQPE	Notes
IG1 graduate students	Mean M1	33,00	85,50	57,16	IG1 is taken as an “intervention group”
	Std. Deviation SD1	9,47	9,85	18,95	
	N1	149	149	149	
IG2 undergraduate students	Mean M2	34,12	83,12	55,84	IG2 is taken as a “comparison group”
	Std. Deviation SD2	9,81	11,18	17,85	
	N2	139	139	139	
Effect size	g Hedges	0,332			The final test scores in 2 groups were compared taking into account the pretest. The gain score effect size was calculated.
Effect size standard error	SE(g)	0,157			
Improvement index 1	U3 — 50%	63,00% — 50% = 13,00%			
t -statistics	t	2,796			Unadjusted for clustering
p-value	p	0,006 **			
t-statistics corrected for clustering	t_a	1,444			Clustering correction was calculated, since the effect size without adjusting for clustering is significant
Clustering-corrected statistical p value	p	0,150			
Effect size	g2		- 0,057		The test at DMQPE scores in 2 groups were compared taking into account the final test. The gain score effect size was calculated.
Effect size standard error	SE(g2)		0,157		
Improvement index 2	U3 — 50%		47,72% — 50% = -2,28%		
t -statistics	t		- 0,483		Unadjusted for clustering
p-value	p		0,629		
Effect size	g3			0,132	The test at DMQPE scores in 2 groups were compared taking into account the pretest. The gain score effect size was calculated.
Effect size standard error	SE(g3)			0,140	
Improvement index 3	U3 — 50%			55,25% — 50% = 5,25%	
t -statistics	t			1,118	Unadjusted for clustering
p-value	p			0,264	

As Table 6 shows, graduate students show significantly better outcomes compared with undergraduate students in the final test taking into account the pretest (g1 = 0.332, p = 0.006). The improvement index is 13%. It means that the rank of the average graduate student would correspond to the 63rd percentile in the undergraduate students group, or, equivalently, the median student of masters' level would have a rank of 13 percentage points higher than the median student of undergraduate level, which, by definition, corresponds to the 50th percentile. However, clustering adjustment gives $p = 0.150$, and the effect becomes unreliable. The effect size for master's level group compared to the undergraduate level group is not significant for the test at DMQPE taking into account the final test scores ($g2 = -0.057$, $p = 0.629$), as well as the pretest ($g3 = 0.132$, $p = 0.264$). Thus, the knowledge retention in both categories of students does not differ.

The psychometric characteristics of the achievement test developed by us were evaluated using database of the test at DMQPE and calculated using the software of the HT-Line laboratory. The test consists of 32 items with 4 optional answers, one of which is correct. According to the binomial criterion for a series of 32 trials with a probability of success of $P = 0.25$, if a student scored 40.6 or more percentage points, then with a probability of 95% we can assume that this is not a random choice of answers, but if 46.8 or more, then the probability increases to 99%.

Discrimination coefficients of 31 test items exceed 3, which is favorable. The empirically calculated difficulty of the test items gives 3% of easy, 60% of medium and 37% of difficult ones. To check the construct validity, the Spearman rank correlation coefficient was used. Knowledge retention outcomes significantly correlate with the test scores on the subject "Research and Forecasting in Education" ($\rho = 0.273$, $p < 0.01$) and the overall e-course assessment ($\rho = 0.507$, $p < 0.01$) in masters' level students, and with English test scores ($\rho = 0.283$, $p < 0.01$) and also with overall e-course assessment ($\rho = 0.346$, $p < 0.01$) in undergraduate students. In all cases, a weak or medium direct relationship was obtained: the best indicators for one test were associated with the best indicators for others. This is probably due to the fact that successful students tend to show high academic achievements in various subjects, regardless of their specificity.

Discussion

The findings of comparing the academic achievements of graduate and undergraduate students are of interest for several reasons.

Firstly, these are representatives of 2 different generations. It is believed that the "millennium generation" is better adapted to learning in a digital environment, and one would expect that 3rd year undergraduate students would show higher educational outcomes, but the experiment did not reveal differences: effect size in all cases was insignificant. Adults and more mature people engaged in professional activities and having family responsibilities, after completing the e-course, show the same high academic achievments as student youth. Perhaps this is due to their more conscious attitude to training and higher motivation to develop competencies in the chosen profile. It is also possible that they are helped by self-organization and learning skills acquired earlier at undergraduate level. A certain role could also be played by a sense of responsibility or anxiety of some students of a more mature age. In addition, the professional activity and everyday life of most people is directly related to the Internet, so the difficulties of mastering digital competencies in mature people can be exaggerated, and this is just a stereotype.

Secondly, we observed the same tendency among students of both categories: pretest scores are low, posttest scores significantly and strongly increase, and then after 1.5—4 months they significantly decrease, while remaining significantly higher than the input indicators. At the same time, the knowledge retention scores are very scattered in comparison with the direct ones: the standard deviation almost doubles. How can this be interpreted? There can be several reasons, and one of them is the use of dishonest strategies when testing and completing case studies. In article [5] we already stressed that more than 70%

of graduate students and more than 85% of undergraduate students believe that there will inevitably be students using dishonest strategies in testing, although this does not mean that they use such strategies themselves. Note that the final test was part of the e-course and, in principle, allowed for the use of dishonest strategies, while long-term outcomes were tested in person with external control, so the assessment can be considered independent, and the results can be taken sufficiently objective and reliable.

Another reason for decreased long-term results compared with the direct ones could be the lack of regularity in the study of the e-course and the habit of storming by some students, as well as their lack of motivation. Another most important reason is the insufficient development of competencies and practical skills in solving cases in SPSS: because of this, information only briefly remains in the memory, but the competencies necessary for independent research and scientific and practical activities do not develop. In addition, the skills of quantitative data analysis should be supported, first of all, in the processing of their own empirical research data. All this determines the direction of the instructor's further activities in motivating students, improving e-courses structure using interactive components and individualizing learning methods.

Conclusions

No differences were found between graduate and undergraduate students in the pretest, final test and overall e-course grade indicators. At the baseline the graduate students had on average slightly lower scores, than undergraduate students, but after the e-course completion their scores were already higher. The effect size is not significant even unadjusted for clustering

The same tendency was revealed in students of both groups: pretest scores are low, posttest scores significantly and strongly increase, and then after 1.5—4 months they significantly decrease, while remaining significantly higher than the input indicators. The knowledge retention scores are very scattered in comparison with the direct ones, the standard deviation is almost doubled.

The gain score effect size and the improvement index are significant for the final test only without adjusting for clustering, i.e. ignoring the fact that the sample consists of several student groups. A median graduate student would have a higher score than a median undergraduate student. Cluster-level effect size is not statistically significant. Cluster-level effect size for overall e-course grade indicators with difference-in-differences adjustment is also not reliable.

After dropping out students who did not participate at knowledge retention testing, reliable effect size and an improvement index for the final test were obtained taking into account the pretest only without correction for clustering. Graduate students show significantly better outcomes compared to the undergraduate ones. However, cluster-level effect size becomes unreliable. The effect size according to knowledge retention indicators, taking into account both the final test and the pretest, is also not significant. Thus, the long-term results of both categories of students do not differ.

The psychometric characteristics of the achievement test in the field of empirical data quantitative analysis can be considered satisfactory. In both groups, a weak and medium direct relationship was found between the achievment test scores, the overall e-course assessment and two tests in non-mathematical subjects: the best indicators for one test are associated with the best indicators for the other tests.

Further studies suggest the improvement of e-courses in terms of motivating students, using active and interactive components, as well as individualizing teaching methods. It would be very interesting to compare the various aspects of blended and online higher education.

References

1. Gosudarstvennaya programma Rossiiskoi Federatsii «Razvitie obrazovaniya» [The state program of the Russian Federation "Development of education"]. Bank dokumentov. Ministerstvo prosveshcheniya Rossiiskoi Federatsii [Bank of documents. Ministry of Education of the Russian Federation]. Available at: https://docs.edu.gov.ru/document/3a928e13b4d292f8f71513a2c02086a3/ (Accessed: 05.06.2020). (In Russ.).

2. Dmitrievskaya N.A., Goremykina G.I. Modelirovanie sistemy upravleniya po rezul'tatam deyatel'nosti smart-universiteta v usloviyakh tsifrovizatsii ekonomiki i obshchestva [Modeling a management system based on the results of a smart university in the context of digitalization of the economy and society]. Materialy mezhdunarodnoi konferentsii "*E-Learning Stakeholders and Researchers Summit*" (g. Moskva, 5—6 dekabrya 2018 g.) [Materials of the international conference "*E-Learning Stakeholders and Researchers Summit*"]. Moscow: Publ. Izd. dom Vysshei shkoly ekonomiki, 2018, pp. 39—46. (In Russ.).

3. Margolis A.A. Chto smeshivaet smeshannoe obuchenie? [What Kind of Blending Makes Blended Learning?]. *Psikhologicheskaya nauka i obrazovanie = Psychological Science and Education*, 2018. Vol. 23, no. 3, pp. 5—19. DOI:10.17759/pse.2018230301 (In Russ., abstr. in Engl.).

4. Sorokova M.G. Matematicheskie metody v psikhologo-pedagogicheskikh issledovaniyakh: Uchebnoe posobie [Mathematical methods in psycho-educational researches]. Moscow: Publ. Neolit, 2020. 216 p. DOI:10.17759/psychlib/978-5-6043562-0-3 (In Russ., in Engl.).

5. Sorokova M.G. Digital Educational Environment in University: Who is More Comfortable Studying in It? *Psikhologicheskaya nauka i obrazovanie = Psychological Science and Education*, 2020. Vol. 25, no. 2, pp. 44—58. DOI:10.17759/pse.2020250204 (In Russ., in Engl.).

6. Sorokova M.G. Elektronnyi kurs kak tsifrovoi obrazovatel'nyi resurs smeshannogo obucheniya v usloviyakh vysshego obrazovaniya [E-Course as Blended Learning Digital Educational Resource in University]. *Psikhologicheskaya nauka i obrazovanie = Psychological Science and Education*, 2020. Vol. 25, no. 1, pp. 36—50. DOI:10.17759/pse.2020250104 (In Russ., in Engl.).

7. Chkhutiashvili L.V. Gosudarstvennaya politika v sfere onlain-obrazovaniya [State policy in the field of online-education]. Materialy mezhdunarodnoi konferentsii "*E-Learning Stakeholders and Researchers Summit*" (g. Moskva, 5—6 dekabrya 2018 g.) [Materials of the international conference "*E-Learning Stakeholders and Researchers Summit*"]. Moscow: Publ. Izd. dom Vysshei shkoly ekonomiki, 2018, pp. 18—21. (In Russ.).

8. Akcayır G., Akcayır M. The flipped classroom: A review of its advantages and challenges. *Computers & Education,* 2018. Vol. 126, pp. 334—345. DOI:10.1016/j.compedu.2018.07.021

9. Arıf S., Omar İ. Effectiveness of Flipped Classroom in Teaching Basic English Courses. *Yükseköğretim Dergisi*, 2019. Vol. 9, no. 3, pp. 279—289. DOI:10.2399/yod.19.003

10. Bouwmeestera R.A.M., de Kleijnb R.A.M., van den Bergc I.E.T., ten Cated O.Th.J., van Rijena H.V.M., Westervelde H.E. Flipping the medical classroom: Effect on workload, interactivity, motivation and retention of knowledge. *Computers & Education*, 2019. Vol. 139, pp. 118—128. DOI:10.1016/j.compedu.2019.05.002

11. Chen F., Lui A.M., Martinelli S.M. A systematic review of the effectiveness of flipped classrooms in medical education. *Medical Education*, 2017. Vol. 51, no. 6, pp. 585—597. DOI:10.1111/medu.13272

12. Chen K.S., Monrouxe L., Lu Y.H., Jenq C.C., Chang Y.J., Chang Y.C. et al. Academic outcomes of flipped classroom learning: A meta-analysis. *Medical Education*, 2018. Vol. 52, no. 9, pp. 910—924. DOI:10.1111/medu.13616

13. De Kleijn R.A.M., Bouwmeester R.A.M., Ritzen M.M.J., Ramaekers S.P.J., Van Rijen H.V.M. Students' motives for using online formative assessments when preparing for summative assessments. *Medical Teacher*, 2013. Vol. 35, no. 12, pp. 1—7. DOI:10.3109/0142159X.2013.826794

14. Hew K.F., Lo C.K. Flipped classroom improves student learning in health professions education: a meta-analysis. *BMC Med Educ*, 2018. Vol. 18, 38. DOI:10.1186/s12909-018-1144-z

15. IES What Works Clearinghouse. Handbooks and other resources. Available at: https://ies.ed.gov/ncee/wwc/Handbooks (Accessed 05.06.2020).

16. Morton D.A., Colbert-Getz J.M. Measuring the impact of the flipped anatomy classroom: The importance of categorizing an assessment by Bloom's taxonomy. *Anatomical Sciences Education*, 2017. Vol. 10, no. 2, pp. 170—175. DOI:10.1002/ase.1635

17. Prinsloo P. Big data in education. The digital future of learning, policy and practice. *International Studies in Sociology of Education*, 2020. Vol. 29, no. 1—2, pp. 183—186. DOI:10.1080/09620214.2019.1690546

18. Rajaram K. Flipped classrooms: Scaffolding support system with real-time learning interventions. *Asian Journal of the Scholarship of Teaching and Learning*, 2019. Vol. 9, no. 1, pp. 30—58. Available at: http://www.nus.edu.sg/cdtl/engagement/publications/ajsotl-home/archive-of-past-issues/v9n1/flipped-classrooms-providing-a-scaffolding-support-system-with-real-time-learning-interventions (Accessed 02.04.2020).

19. Røe Y., Rowe M., Ødegaard N.B. et al. Learning with technology in physiotherapy education: design, implementation and evaluation of a flipped classroom

teaching approach. *BMC Med Educ,* 2019. Vol. 19, 291. DOI:10.1186/s12909-019-1728-2
20. Sorokova M. (2020) "Educational outcomes of graduate and undergraduate students who completed e-courses in mathematical methods in psychological and educational researches", *Mendeley Data*, v1 http://dx.doi.org/10.17632/hvfkdpfwnr.1
21. Wang K., Zhu C. MOOC-based flipped learning in higher education: students' participation, experience and learning performance. *Int J Educ Technol High Educ*, 2019. Vol. 16, 33. DOI:10.1186/s41239-019-0163-0
22. Williamson B. Making markets through digital platforms: Pearson, edu-business, and the (e)valuation of higher education. *Critical Studies in Education*, 2020. DOI:10.1080/17508487.2020.1737556

Литература

1. Государственная программа Российской Федерации «Развитие образования» [Электронный ресурс] // Банк документов. Министерство просвещения Российской Федерации URL: https://docs.edu.gov.ru/document/3a928e13b4d292f8f71513a2c02086a3/ (дата обращения: 05.06.2020).
2. *Дмитриевская Н.А., Горемыкина Г.И.* Моделирование системы управления по результатам деятельности смарт-университета в условиях цифровизации экономики и общества // Материалы международной конференции «E-Learning Stakeholders and Researchers Summit» (г. Москва, 5—6 декабря 2018 г.). Москва: Изд. дом Высшей школы экономики, 2018. С. 39—46.
3. *Марголис А.А.* Что смешивает смешанное обучение? // Психологическая наука и образование. 2018. Том 23. № 3. С. 5—19. DOI:10.17759/pse.2018230301
4. 4. *Сорокова М.Г.* Математические методы в психолого-педагогических исследованиях: Учебное пособие. М.: Неолит, 2020. 216 с. DOI:10.17759/psychlib/978-5-6043562-0-3
5. *Сорокова М.Г.* Цифровая образовательная среда университета: кому более комфортно в ней учиться? // Психологическая наука и образование. 2020. Том 25. № 2. С. 44—58. DOI:10.17759/pse.2020250204
6. *Сорокова М.Г.* Электронный курс как цифровой образовательный ресурс смешанного обучения в условиях высшего образования // Психологическая наука и образование. 2020. Том 25. № 1. С. 36—50. DOI:10.17759/pse.2020250104
7. *Чхутиашвили Л.В.* Государственная политика в сфере онлайн-образования // Материалы международной конференции «E-Learning Stakeholders and Researchers Summit» (г. Москва, 5—6 декабря 2018 г.). Москва: Изд. дом Высшей школы экономики, 2018. С. 18—21.
8. *Akcayır G., Akcayır M.* The flipped classroom: A review of its advantages and challenges // Computers & Education. 2018. Vol. 126. P. 334—345. DOI:10.1016/j.compedu.2018.07.021
9. *Arıf S., Omar İ.* Effectiveness of Flipped Classroom in Teaching Basic English Courses // Yükseköğretim Dergisi. 2019. Vol. 9. № 3. P. 279—289. DOI:10.2399/yod.19.003
10. *Bouwmeestera R.A.M., de Kleijnb R.A.M., van den Bergc I.E.T., ten Cated O.Th.J., van Rijena H.V.M., Westervelde H.E.* Flipping the medical classroom: Effect on workload, interactivity, motivation and retention of knowledge // Computers & Education. 2019. Vol. 139. P. 118—128. DOI:10.1016/j.compedu.2019.05.002
11. *Chen F., Lui A.M., Martinelli S.M.* A systematic review of the effectiveness of flipped classrooms in medical education // Medical Education. 2017. Vol. 51. № 6. P. 585—597. DOI:10.1111/medu.13272
12. *Chen K.S., Monrouxe L., Lu Y.H., Jenq C.C., Chang Y.J., Chang Y.C. et al.* Academic outcomes of flipped classroom learning: A meta-analysis // Medical Education. 2018. Vol. 52. № 9. P. 910—924. DOI:10.1111/medu.13616
13. *De Kleijn R.A.M., Bouwmeester R.A.M., Ritzen M.M.J., Ramaekers S.P.J., Van Rijen H.V.M.* Students' motives for using online formative assessments when preparing for summative assessments // Medical Teacher. 2013. Vol. 35. № 12. P. 1—7. DOI:10.3109/0142159X.2013.826794
14. *Hew K.F., Lo C.K.* Flipped classroom improves student learning in health professions education: a meta-analysis // BMC Med Educ. 2018. Vol. 18, 38. DOI:10.1186/s12909-018-1144-z
15. IES What Works Clearinghouse. Handbooks and other resources. Available at: https://ies.ed.gov/ncee/wwc/Handbooks (Accessed 05.06.2020).
16. *Morton D.A., Colbert-Getz J.M.* Measuring the impact of the flipped anatomy classroom: The importance of categorizing an assessment by Bloom's taxonomy // Anatomical Sciences Education. 2017. Vol. 10. № 2. P. 170—175. DOI:10.1002/ase.1635
17. *Prinsloo P.* Big data in education. The digital future of learning, policy and practice // International Studies in Sociology of Education. 2020. Vol. 29. № 1—2. P. 183—186. DOI:10.1080/09620214.2019.1690546
18. *Rajaram K.* Flipped classrooms: Scaffolding support system with real-time learning interventions // Asian Journal of the Scholarship of Teaching and Learning. 2019. Vol. 9. № 1. P. 30—58. Available at: http://www.nus.edu.sg/cdtl/engagement/publications/ajsotl-home/archive-of-past-issues/v9n1/flipped-classrooms-providing-a-scaffolding-support-system-with-real-time-learning-interventions (Accessed 05.06.2020).

19. *Røe Y., Rowe M., Ødegaard N.B. et al.* Learning with technology in physiotherapy education: design, implementation and evaluation of a flipped classroom teaching approach // BMC Med Educ. 2019. Vol. 19, 291. DOI:10.1186/s12909-019-1728-2

20. *Sorokova M.* (2020) “Educational outcomes of graduate and undergraduate students who completed e-courses in mathematical methods in psychological and educational researches”, Mendeley Data, v1 http://dx.doi.org/10.17632/hvfkdpfwnr.1

21. *Wang K., Zhu C.* MOOC-based flipped learning in higher education: students’ participation, experience and learning performance // Int J Educ Technol High Educ. 2019. Vol. 16, 33. DOI:10.1186/s41239-019-0163-0

22. *Williamson B.* Making markets through digital platforms: Pearson, edu-business, and the (e)valuation of higher education // Critical Studies in Education. 2020. DOI:10.1080/17508487.2020.1737556

Information about the authors

Marina G. Sorokova, Doctor of Education, PhD in Physics and Mathematics, Head of Scientific and Practical Center for Comprehensive Support of Psychological Research «PsyDATA», Professor, Chair of Applied Mathematics, Faculty of Information Technology, Moscow State University of Psychology & Education, Moscow, Russia, ORCID: http://orcid.org/0000-0002-1000-6487, e-mail: sorokovamg@mgppu.ru

Информация об авторах

Сорокова Марина Геннадьевна, доктор педагогических наук, кандидат физико-математических наук, руководитель Научно-практического центра по комплексному сопровождению психологических исследований PsyDATA, профессор кафедры Прикладной математики факультета Информационных технологий, Московский государственный психолого-педагогический университет (ФГБОУ ВО МГППУ), г. Москва, Российская Федерация, ORCID: http://orcid.org/0000-0002-1000-6487, e-mail: sorokovamg@mgppu.ru

Получена 09.06.2020
Принята в печать 11.02.2021

Received 09.06.2020
Accepted 11.02.2021