

Министерство науки и высшего образования РФ
Национальный исследовательский Нижегородский государственный
университет им. Н.И. Лобачевского

The Digital Scholar: Philosopher's lab

Vol. 3. • no 3.

Цифровой ученый: лаборатория философа

Т. 3. • №3.

Нижний Новгород
Издательство Нижегородского государственного университета
2020

Цифровой ученый: лаборатория философа 2020. Т. 3. №3.

Ежеквартальный научно-теоретический журнал

Главный редактор: И.Т. Касавин (Институт философии РАН, Москва, Россия;
ННГУ им. Н.И. Лобачевского, Нижний Новгород, Россия)

Зам. главного редактора: А.Н. Ткачев (ННГУ им. Н.И. Лобачевского)

Ответственный секретарь: С.В. Шибаршина (ННГУ им. Н.И. Лобачевского)

Редакционная коллегия:

А.М. Бекарев (ННГУ им. Н.И. Лобачевского, Нижний Новгород), Ю.К. Волков (Арзамасский филиал ННГУ им. Н.И. Лобачевского), А. Гельферт (Берлинский технический университет, Берлин, Германия), И.Н. Грифцова (МПГУ, Москва), А.М. Дорожкин (ННГУ им. Н.И. Лобачевского, Нижний Новгород), В.А. Кутырев (ННГУ им. Н.И. Лобачевского, Нижний Новгород), Е.В. Масланов (Институт философии РАН, Москва), М.А. Можейко (БГУКИ, Минск, Республика Беларусь), И.Д. Невважай (СГЮА, Саратов), О.А. Останина (ВятГУ, Киров), К.А. Очеретяный (СПбГУ, Санкт-Петербург), О.В. Париков (Нижегородская академия МВД России, Нижний Новгород), В.С. Пронских (Национальная Ускорительная Лаборатория им. Э. Ферми, Батавия, США; ОИЯИ, Дубна), В.М. Розин (Институт философии РАН, Москва), Н.С. Розов (НГУ, Новосибирск), Руане Луис Паулу (Федеральный университет Сан-Жуан-дел-Рей, Бразилия), А. Рузер (Университет Агдера, Кристиансанн, Норвегия), В.В. Савчук (СПбГУ, Санкт-Петербург), В.А. Фортунатова (ННГУ им. Н.И. Лобачевского, Нижний Новгород), Х. Чанг (Кембриджский университет, Кембридж, Великобритания), Е.Э. Чеботарева (СПбГУ, Санкт-Петербург), Л.В. Шипова (СПбГУ, Санкт-Петербург), А.Ф. Яковлева (МГУ, Москва).

Редакционный совет:

В.А. Бажанов (УлГУ, Ульяновск),
М.Дж. Веллер (Открытый Университет, Милтон-Кинс, Великобритания),
В.А. Лекторский (Институт философии РАН, Москва), В.Н. Порус (НИУ ВШЭ, Москва),
С.У. Фуллер (Университет Уорика, Ковентри, Великобритания),
Чувильдеев В.Н. (ННГУ им. Н.И. Лобачевского, Нижний Новгород).

Журнал включен в Philosophy Documentation Center (PDC),
Ulrichs Periodicals Directory, Российский индекс научного цитирования (РИНЦ).
Цифровой идентификатор объекта (DOI) присваивается через PDC.

Учредитель и издатель: Национальный исследовательский Нижегородский государственный университет им. Н.И.Лобачевского.

Свидетельство о регистрации СМИ: ПИ № ФС 77 — 72454 от 05.03.2018

Периодичность: 4 раза в год. Выходит с 2018 г.

Адрес редакции: 603950, г. Нижний Новгород, просп. Гагарина, 23, корп. 3.
ННГУ им. Н.И. Лобачевского, кафедра философии.

Официальный сайт журнала: <http://www.digital-scholar.unn.ru>

Подписной индекс в Объединенном каталоге "Пресса России" — 39461

The Digital Scholar: Philosopher's lab

2020. Vol. 3. No 3.

Quarterly peer-reviewed journal

Editor-in-chief: Ilya Kasavin (Institute of Philosophy of the Russian Academy of Sciences; National Research Lobachevsky State University of Nizhny Novgorod, Russia)

Editorial Assistants: Andrey Tkachev (National Research Lobachevsky State University of Nizhny Novgorod, Russia); Svetlana Shibarshina (National Research Lobachevsky State University of Nizhny Novgorod, Russia)

Editorial Board:

Adrian Bekarev (National Research Lobachevsky State University of Nizhny Novgorod, Russia), Hasok Chang (University of Cambridge, Cambridge, England, United Kingdom), Elena Chebotareva (Saint Petersburg State University, Saint Petersburg, Russia), Aleksandr Dorozhkin (National Research Lobachevsky State University of Nizhny Novgorod, Russia), Vera Fortunatova (National Research Lobachevsky State University of Nizhny Novgorod, Russia), Axel Gelfert (Technische Universität Berlin, Germany), Irina Griftsova (Moscow State University of Education, Moscow, Russia), Vladimir Kutyrev (National Research Lobachevsky State University of Nizhny Novgorod, Russia), Evgeny Maslanov (RAS Institute of Philosophy, Moscow, Russia), Marina Mozheyko (Belarusian State University of Culture and Arts, Minsk, Belarus), Igor Nevzhay (Saratov State Academy of Law, Saratov, Russia), Konstantin Ocheretyany (Saint Petersburg State University, Saint Petersburg, Russia), Olga Ostanina (Vyatsky State University, Kirov, Russia), Oleg Parilov (Nizhny Novgorod Academy of the Ministry of Internal Affairs of Russia, Russia), Vitaly Pronskikh (Fermi National Accelerator Laboratory, Batavia, USA; Joint Institute for Nuclear Research, Dubna, Russia), Vadim Rozin (RAS Institute of Philosophy, Moscow, Russia), Nikolay Rozov (Novosibirsk State University, Novosibirsk, Russia), Luiz Paulo Rouanet (Universidade Federal de São João Del-Rei, Brasil), Alexander Ruser (University of Agder, Kristiansand, Norway), Valery Savchuk (Saint Petersburg State University, Saint Petersburg, Russia), Lada Shipovalova (Saint Petersburg State University, Saint Petersburg, Russia), Yuri Volkov (Arzamas Branch of Lobachevsky State University of Nizhny Novgorod, Russia), Alexandra Yakovleva (Lomonosov Moscow State University, Moscow, Russia).

Editorial Council:

Valentin Bazhanov (Ulyanovsk State University, Russia), Martin Weller (The Open University, Milton Keynes, England, United Kingdom), Vladislav Lectorsky (Russian Academy of Sciences, Moscow, Russia), Vladimir Porus (National Research University Higher School of Economics, Moscow, Russia), Steve Fuller (University of Warwick, Coventry, England, United Kingdom), Vladimir Chuvildeev (National Research Lobachevsky State University of Nizhny Novgorod, Russia).

Abstracting and indexing: Philosophy Documentation Center (PDC),
Ulrichs Periodicals Directory, Russian Index of Science Citation (RISC).
DOI assignment by PDC.

Publisher: Nizhny Novgorod State University Press.

The Mass Media Registration Certificate: No. FS77-72454 on 05.03.2018

Frequency: 4 times per year. First issue: 2018.

Place of publication: 23 Prospekt Gagarina (Gagarin Avenue), Nizhny Novgorod, 603950, Russia

Official website: <http://www.digital-scholar.unn.ru>

Subscription index in the integrated catalogue «The Press of Russia» is 39461.

СОДЕРЖАНИЕ

РЕДАКЦИОННАЯ СТАТЬЯ

<i>Пронских В.С.</i> Проблемы ядерных технологий и радиационной безопасности.....	6
--	---

ДИСКУССИЯ

<i>Костина А.О.</i> Платформенный урбанизм: новая топология городского пространства.....	25
<i>Шибаршина С.В.</i> От «умных» алгоритмов к цифровому паноптику.....	38
<i>Масланов Е.В.</i> Цифровые платформы, контркультура и контрповедение.....	46
<i>Хусяинов Т.М.</i> Платформенный урбанизм и положение работников.....	53
<i>Костина А.О.</i> Имеет ли цифровая платформа форму паноптикона?.....	60

УНИВЕРСИТЕТ

<i>Демин М.Р.</i> Рандомизированные тесты по философии: опыт использования и статистический анализ.....	65
--	----

КОНЦЕПЦИЯ

<i>Столярова О.Е.</i> Историческая эпистемология и медицина: что они могут дать друг другу?.....	83
---	----

CASE STUDIES

<i>Волков Ю.К.</i> Нормативная природа институтов малой и большой науки.....	101
<i>Соловьева А.С.</i> Закон, философия и религия в архаической Греции: развитие понятий $\theta\epsilon\sigma\mu\acute{o}\varsigma$ и $\nu\acute{o}\mu\circ\varsigma$	118

ПАНОРАМА

<i>Куслий П.С.</i> Кодирование временной информации в естественном языке: обзор новейших исследований	128
--	-----

TABLE OF CONTENTS

EDITORIAL

<i>Vitaly S. Pronskikh. Problems of nuclear technology and radiation safety.....</i>	6
--	---

DISCUSSIONS

<i>Alina O. Kostina. Platform urbanism: new topology of urban space.....</i>	25
<i>Svetlana V. Shibarshina. From smart algorithms towards digital panopticon.....</i>	38
<i>Evgeniy V. Maslanov. Digital platforms, counter-culture and counter-conduct.....</i>	46
<i>Timur M. Khusyainov. Platform urbanism and the situation of employees.....</i>	53
<i>Alina O. Kostina. Are digital platforms panopticon-shaped?.....</i>	60

UNIVERSITY

<i>Maxim R. Demin. Randomized tests in philosophy: implementation and item analysis.....</i>	65
--	----

VIEWPOINTS

<i>Olga E. Stoliarova. Historical epistemology and medicine: what can they give each other?.....</i>	83
--	----

CASE STUDIES

<i>Yuri K. Volkov. The normative nature of institutions of little science and big science.....</i>	101
<i>Aleksandra S. Solovyeva. Law, philosophy and religion in archaic Greece: the development of terms θεσμός and νόμος.....</i>	118

VISTA

<i>Petr S. Kusliy. Encoding temporal information in natural language: an overview of recent works in formal semantics.....</i>	128
--	-----

УНИВЕРСИТЕТ

УДК 167

DOI: 10.5840/dspl20203329

РАНДОМИЗИРОВАННЫЕ ТЕСТЫ ПО ФИЛОСОФИИ: ОПЫТ ИСПОЛЬЗОВАНИЯ И СТАТИСТИЧЕСКИЙ АНАЛИЗ

Демин Максим Ростиславович – кандидат философских наук, доцент. Санкт-Петербургская школа социальных наук и востоковедения, Национальный Исследовательский Университет «Высшая школа экономики». 192171, г. Санкт-Петербург, Седова ул., д. 55 корп. 24;
e-mail: mdemin@hse.ru

В статье описывается опыт применения рандомизированных онлайн-тестов в курсе философии. На основе двух контрольных тестов, выполненных 172 студентами, демонстрируется эффективность методов статистического анализа для измерения успеваемости, а также оценки качества самих тестовых заданий. Рассматриваются четыре критерия: валидность полученных результатов на основе гипотезы нормальности распределения правильных ответов; расчет надежности теста (коэффициент Спирмена-Брауна); индекс трудности и индекс дискриминативности, рассчитываемые для каждого конкретного задания. В ходе постэкзамениционного анализа на примере четырех заданий теста показано, какие вопросы оказались эффективными, а какие требуют дальнейшей доработки. Делается вывод, что тестовые задания не только значимы для оценивания студентов, но и должны рассматриваться как важный элемент обратной связи, а именно – данные, полученные в ходе решения студентами теста, могут быть использованы для последующей корректировки заданий.

Ключевые слова: философия, анализ заданий (item analysis), вопросы с множественным выбором, постэкзамениционный анализ (post-hoc review), рандомизированные тесты

Цитирование: Демин М.Р. Рандомизированные тесты по философии: опыт использования и статистический анализ. 2020. Т. 3. № 3. С. 65-82.
DOI: 10.5840/dspl20203329

RANDOMIZED TESTS IN PHILOSOPHY: IMPLEMENTATION AND ITEM ANALYSIS

Maxim R. Demin – PhD in Philosophy, associate professor. St. Petersburg School of Social Sciences and Humanities, Department of Sociology, National Research University Higher

The article describes an implementation of randomized computer-based tests in a philosophy course. Based on two control tests performed by 172 students, we demonstrate the effectiveness of statistical analysis methods for measuring student performance, as well as evaluating the validity of developed multiple-choice questions. Four

School of Economics.
55 Sedova str., 192171 Saint
Petersburg, Russian Federation;
e-mail: mdemin@hse.ru.

criteria are considered: the validity of the results obtained basing on the hypothesis of the normal distribution of correct answers; calculation of test reliability (Spearman-Brown Coefficient); the index of difficulty and the index of discrimination calculated for each specific task. In the course of post-examination analysis, using the example of four test items, the study shows which questions appear effective and which ones require further refinement. The text argued that multiple-choice questions are both significant for assessing knowledge of philosophical concepts and should be considered an important element of feedback. The data obtained in the course of solving the test by students can be used for subsequent adjustment of the items.

Keywords: philosophy, item analysis, multiple choice questions, post-hoc review, data-driven instruction, randomized tests

Might be cited as: Demin, Maxim. 2020. Randomized tests in philosophy: implementation and item analysis, *The Digital Scholar: Philosopher's Lab*, 3 (3): 65-82. DOI: 10.5840/dspl20203329 (In Russian).

1. Введение

Тесты успешно используются в обучении самым разным дисциплинам от медицины до истории [Кейс, Свэнсон, 1996, web; Давыдова, Михалева, 2013; Mertler, 2014; Secolsky, Denison, 2017]. Они применяются как для измерения когнитивных способностей, так и для оценки усвоения материала конкретной дисциплины. Принимая во внимание международный опыт, в курсах философии тесты используются для оценки критического мышления [Burke et al., 2014]. В то же время, насколько нам известно, существует не так много описанного опыта использования тестов для измерения философских знаний.

Задача данного текста – описать применение тестов в курсе философии и показать, как методы статистического анализа позволяют оценить качество тестовых заданий. Рассматриваемая модель преподавания в течение нескольких лет разрабатывалась и применялась в курсах философии, которые читали М. Р. Демин и С. М. Левин в НИУ ВШЭ-Санкт-Петербург. В рамках этой статьи будут проанализированы результаты только одного курса по философии, прочитанного в качестве обязательной дисциплины для 172-х студентов первого курса образовательной программы «Дизайн» в 2019/2020 академическом году. Курс читался двумя преподавателями, каждый из которых вел по четыре группы (автор программы и ведущий семинаров – М.Р. Демин; ведущий семинаров – Б.В. Фауль).

Нужно оговорить, что мы не учитывали в результатах такие базовые демографические характеристики, как возраст и пол студентов. Относительно возраста состав студентов первого курса программы, на которой читался курс, приблизительно гомогенный: это 17–18-летние молодые люди, как правило, сразу после школы поступившие в вуз. Мы сознательно не учитывали пол студентов, так как задача выяснять восприятие философии в зависимости от половой принадлежности выходит за пределы данного исследования. В целом можно сказать о существенном превалировании женщин среди наших респондентов.

Тестирование и электронное онлайн-сопровождение курса образовательного процесса осуществлялись в бесплатной версии образовательной платформы «Schoology» (Schoology.com). Данная программа позволила собрать данные для последующего анализа в Microsoft Excel.

Разрабатывая тестовые задания и анализируя их эффективность, мы стремились использовать оптимальные решения. В этом отношении личный педагогический опыт призван служить наглядной иллюстрацией того, как можно использовать тесты в курсе философии.

2. Вопросы с множественным выбором в курсе философии

Специалисты в области бакалаврского образования выделяют четыре категории для оценки обучения:

1. Владение информацией (знание): предполагает, что студент должен запоминать определенные факты, имена, определение концепций.

2. Понимание: предполагает, что студент способен объяснить концепцию, распознать ее в частных примерах.

3. Мышление: предполагает, что студент может применить изученное. То есть студент может увидеть проблему и попытаться ее решить, используя знания, полученные в курсе.

4. Обучение: предполагает умение учащегося находить дополнительную информацию [Erickson et al., 2006, p. 50–62].

На мой взгляд, только навыки, входящие в четвертую категорию, сложно измерить при помощи тестирования. Умение работать с дополнительной информацией может быть развито и оценено при помощи других заданий, например, эссе. В то же время тесты могут измерить сформированность первых трех умений и навыков. Это можно сделать, предложив студентам ответить на вопросы с множественным выбором (multiple-choice questions). Существует обширная литература на тему составления тестовых вопросов (см., напр. [Аванесов, 2002; Майоров, 2000; Haladyna et al., 2002]). Вопросы с одним лучшим ответом кажутся наиболее эффективными как для проверки знания, так и для последующей работы с ре-

зультатами теста. Такие вопросы состоят из условия (виньетки, описывающей ситуацию или концепцию), вводного вопроса и нескольких вариантов ответов, среди которых один верный и несколько неверных, отвлекающих ответов. Отвлекающие ответы принято называть дистракторами. В ходе разработки каждого задания мы использовали три или четыре дистрактора.

Создание тестовых заданий предполагает разбор содержания курса с целью точной формулировки перечня фактов, концепций и идей, которые предполагается осваивать на каждом занятии. Программа нашего курса сфокусирована на современных философских проблемах; в качестве материала для обсуждения на семинарских занятиях мы использовали популярную и адаптированную для преподавания литературу [Нагель, 2000; Лоу, 2007; Сэндел, 2013; Дюпре, 2014]. Все тестовые задания были написаны строго по текстам семинаров. Для каждого семинара составлялось около 10 вопросов. Наиболее удачные попадали в банк вопросов, по которым давался тест.

Программа курса, продолжительностью 21 неделю, предполагала следующую формулу оценивания:

- 1) работа на семинарах, 30 % от итоговой оценки;
- 2) тест середины курса на 11-й неделе курса: 20 вопросов на 40 минут (общий банк вопросов = 30), 20% от итоговой оценки;
- 3) экзамен: 40 вопросов на 80 минут (общий банк вопросов = 81), 50% от итоговой оценки.

Ниже мы приведем четыре примера тестовых заданий из тем «Нормативная этика» и «Философия сознания».

1. Катастрофа на подводной лодке

Подводная лодка из-за неисправности оборудования вот-вот нанесет ядерный удар, в результате которого погибнут миллионы людей. Подводную лодку можно уничтожить до нанесения удара, но нет времени на эвакуацию экипажа. Какую из следующих причин утилитарист будет использовать для оправдания уничтожения подводной лодки?

- a. Человек имеет право на жизнь и, в частности, не должен быть убит другим человеком.
- b. Это позволит максимизировать общее количество счастья.
- c. Чем больше людей спасено, тем больше счастья, вероятно, произойдет в будущем.
- d. Экипаж подводной лодки на дежурстве. Поэтому жизнь экипажа может быть принесена в жертву во имя спасения гражданских лиц.
- e. В подлодке нет детей. Права детей важнее, чем права взрослых.

2. Следователь и террорист

Следователь считает, что поймал террориста, и этот человек обладает информацией, располагая которой полиция сможет предотвратить страшный теракт. Можно ли считать этически оправданным применение к подозреваемому пыток? Известно, что Кант, скорее всего, не одобрил бы такой метод получения информации. Выберите один из предложенных доводов, наиболее близкий к позиции Канта.

- a. Под пытками подозреваемый может сообщить неправильную информацию.
- b. Это противоречит гипотетическому императиву.
- c. Нужно относиться к людям так, как Вы бы хотели, чтобы относились к Вам.
- d. Страдания жертв теракта перевешивают страдания одного человека.
- e. Пытать человека – значит сделать его лишь средством для достижения своих целей (спасти невинных людей), а это морально недопустимо.

3. Аргумент против Декарта

Выберите суждение, которое является самым сильным из контраргументов, предложенных против дуализма Декарта:

- a. Сознание животных отличается от вида к виду.
- b. Мы никогда не можем знать ничего определенного.
- c. Нефизическая субстанция по определению не может вызвать изменения в физической субстанции.
- d. Сознание и тело состоят из одного и того же типа субстанции, но имеют разные свойства.
- e. Вода есть H₂O, следовательно, сознание и есть тело.

4. Мэри в черно-белой комнате

В 1982 г. Фрэнк Джексон опубликовал статью, где приводил мысленный эксперимент, получивший название «Мэри в черно-белой комнате». Укажите, что был призван доказать этот эксперимент:

- a. Факты нашего сознания не являются физическими фактами.
- b. Мыслительные эксперименты в философии не доказывают ничего.
- c. Факты нашего сознания могут быть описаны нейтральным научным языком.
- d. Наука успешно может объяснить феномены сознания.
- e. Субстанциональный дуализм Декарта не подтверждается современной наукой.

Что статистический анализ может сказать о качестве этих вопросов? Если хотите проверить свою профессиональную интуи-

цию, ранжируйте эти четыре вопроса от самого эффективного к наименее эффективному вопросу. Правильные ответы и характеристики эффективности вопросов указаны в предпоследнем разделе статьи.

3. Критерии эффективности тестовых заданий

Тестирование не является универсальным инструментом, оно имеет много недостатков и ограничений. Понимание и учет их критически важен для того, чтобы правильно организовать тестирование и оценить его эффективность. Понять, насколько эффективны вопросы теста, можно, статистически анализируя результаты теста. Используя полученные результаты, преподаватель курса и его коллеги могут критически пересмотреть свой педагогический опыт, использовать информацию, полученную от оценивания студентов для того, чтобы улучшить курс [Mertler, 2014].

Существуют стандартные и хорошо прописанные процедуры анализа тестов [Крокер, Алгина, 2010; Давыдова, Михалева, 2013; Lane et al., 2015; Parkes, Zimmaro, 2016]. Мы выбрали четыре ключевых критерия их оценки:

1. Валидность полученных данных. Прежде чем оценить, насколько тестовые задания справляются с их основной задачей – корректно и точно оценивать знания, нужно убедиться, что ответы – результат самостоятельной и индивидуальной работы студента. Валидность теста оценивается на основе гипотезы нормальности распределения правильных ответов и стандартного распределения ошибок. Если распределение не имеет аномалии, то можно говорить, что студенты писали тест самостоятельно.

2. Расчет надежности теста. В этом случае вычисляется согласованность и последовательность вопросов тестов на основании ответов участников. Сильно упрощая идею, лежащую в основе этого параметра, можно отметить следующее: люди, хорошо отвечающие на вопросы одной части теста, должны сравнимо хорошо отвечать и на другие задания и наоборот. Если тенденцию в ответах обнаружить не удастся, то это признак того, что вопросы между собой не связаны. Различные статистические формулы (формула Спирмена-Брауна, формула Кадера-Ричардсона-20) позволяют с разным уровнем приближения вычислить надежность теста. Тест следует признать адекватным, если показатель надежности будет варьироваться от 0,6 до 0,8.

3. Индекс трудности задания. Этот индекс показывает долю студентов, ответивших на задание правильно. Нужно учесть, что если доля студентов, ответивших правильно на вопрос, составляет около 30%, то шансы, что студент правильно ответит на вопрос, близки к случайным. Вопрос с индексом выше 90% следует признать мало информативным. Такие вопросы, скорее всего, стоит

признать плохими. Хорошим считается тот вопрос, на который отвечают от 60 до 90% отвечающих.

4. Индекс дискриминативности задания. Этот индекс отражает способность задания отделять испытуемых с высокой продуктивностью от испытуемых с низкой продуктивностью. Иными словами, он показывает разницу между тем, как на конкретный вопрос отвечали участники, получившие за тест высокой балл, и теми, кто плохо справился с заданием. Коэффициент дискриминации может принимать значения от -1 до +1. Низко дискриминативными будут все вопросы с индексом меньше 0,15; вопросы с индексом выше 0,4 считаются высоко дискриминативными.

Кроме этого, стоит коснуться еще одного аспекта: существенно повлиять на результаты может знание формата тестовых заданий. Поэтому важно ознакомить студентов с форматом тестов заранее.

В нашем случае на семинарах курса были организованы тренировочные задания (то есть тесты, состоящие из 5 заданий, выполнение которых не оценивалось). Тренировочные задания играют очень важную роль, так как они дают студентам и преподавателям важную информацию для прогнозирования результатов будущего контрольного теста.

Таким образом, мы обозначили основные параметры оценки теста, чтобы применить их к нашему опыту.

4. Валидность теста: насколько полученному результату можно доверять?

Вопрос доверия к результатам тестирования особенно важен в том случае, когда преподаватель обеспокоен возможностью списывания ответов на вопросы, которые он так долго придумывал и шифровал. В условиях вынужденного дистанцирования в связи с пандемией коронавируса вопрос о защите теста от списывания встал особенно остро. В этом разделе мы разберем меры, которые были приняты, чтобы минимизировать возможность такого поведения и статистически проверим их эффективность.

Существует несколько стратегий препятствования списыванию тестов, а развитие и доступность технологии компьютерного тестирования делают эти стратегии реализуемыми [Impara, Foster, 2013]. Перечислим основные:

1. Рандомизация ответов на вопрос: варианты ответа перемешиваются.

2. Вариативность вопроса: преподаватели пишут несколько модификаций одного и того же вопроса.

3. Нелинейность теста: в разных вариантах теста вопросы располагаются в разном порядке.

4. Рандомизация вопросов теста: студент получает уникальный набор заданий, случайным образом выбранный из банка заданий.

Возможность сделать все это, и в том числе контролировать выдачу вариативных вопросов или вопросов с приблизительно одинаковой сложностью, определил выбор Schoology как платформы сопровождения курса. Schoology имеет широкий функционал настроек тестов, а также позволяет удобно группировать наборы вопросов по разным темам. Для каждого семинарского теста мы создали по набору вопросов. Мы настроили тест так, что программа могла генерировать случайную последовательность вопросов и не давала возможности возвращаться к предыдущим вопросам. Каждый тест организовывался одновременно для всего потока. В случае с экзаменом, время которого было ограничено 80 минутами, тест был разделен на две части, каждая из которых состояла из 20 вопросов.

Рис. 1. Распределение студентов согласно процентам правильно решенных заданий (тест середины курса, 171 студент).

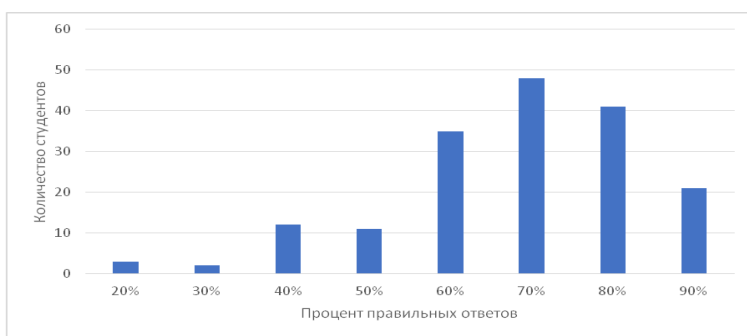


Рис. 2. Распределение студентов согласно процентам правильно решенных заданий (экзамен, вопросы второй части курса, 172 студента).

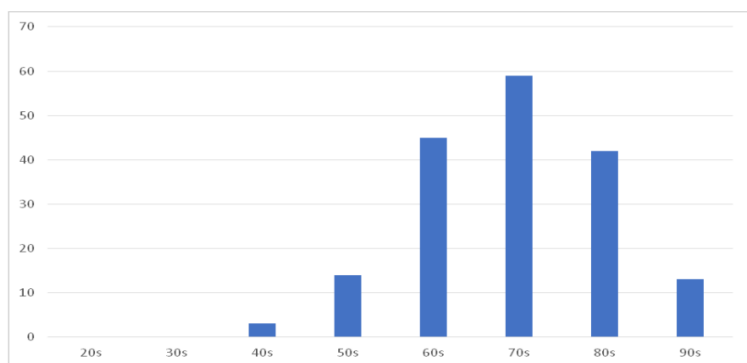


Рис.1 и 2 показывают распределение верных ответов, данных в ходе выполнения теста середины курса и на экзамене. Данные графики показывают нормальное распределение ответов, что говорит о том, что статистически значимых списываний не было. Это доказывает эффективность принятых мер.

Важным аспектом в тестировании является ограничение прохождения теста по времени. Времени не должно быть избыточно много, чтобы не было соблазна консультироваться с другими студентами; но его должно хватить, чтобы осмысленно выполнять задания. После анализа тренировочных заданий было сделано предположение, что 20 вопросов на 40 минут представляют оптимальный баланс между количеством вопросов и предоставленным временем. Эмпирические данные, полученные в результате прохождения студентами теста, позволяют проверить гипотезу.

Рис. 3а. Курс философии для ОП Дизайн. Затраченное время на выполнение теста середины курса и количество правильных ответов студентов четырех групп (76 студентов).

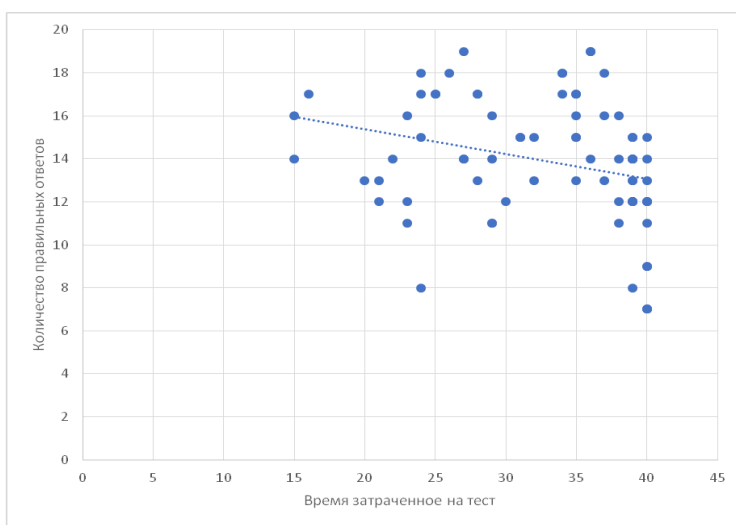
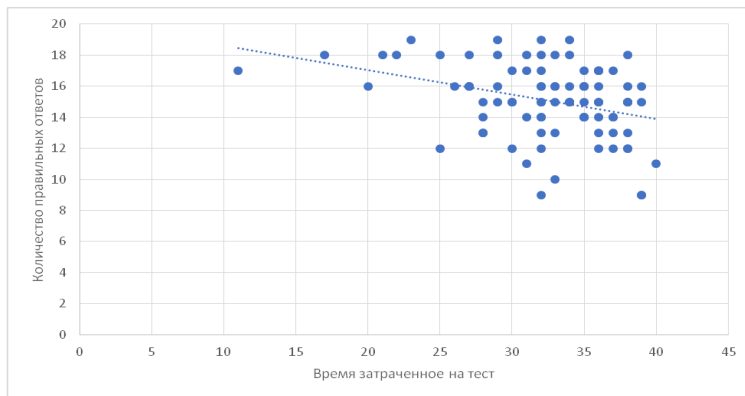


Рис. 3 и 4 показывают время и количество правильных ответов, затраченных на выполнение соответственно теста середины курса и второй части экзамена (напомню, что экзамен был разделен на две части по 40 минут каждая). На обоих рисунках изображена линия тренда, которая показывает, что студенты, набравшие меньшее количество баллов, склонны затрачивать большее количество времени. На экзамене студенты были менее склонны выбирать все время, отведенное на выполнение теста. Это можно объяснить тем, что за пройденный курс студенты стали быстрее решать задания.

Рис. 3. Курс философии для ОП Дизайн. Затраченное время на выполнение второй части экзамена и количество правильных ответов студентов четырех групп (88 студентов).



5. Надежность теста и общая динамика курса

Статистический анализ может помочь в оценке степени взаимосвязанности всех вопросов теста. Формула Кадера-Ричардсона-20 (KR-20) применима на сплошных данных, то есть тест можно анализировать, когда все студенты отвечают на все вопросы. В случае нашего рандомизированного теста это сделать не представляется возможным. Для вопросов теста середины курса полные данные есть только для 18 вопросов, на которые ответили 48 студентов. Рассчитанный по ним показатель Кадера-Ричардсона-20 составил 0,6.

Мы не стали анализировать вопросы экзамена, так как в случае экзамена рандомизация еще больше, поэтому мы рассчитали показатель надежности по формуле Спирмена-Брауна. Этот метод оценки еще называется методом «расщепления» (split-half method). Данная формула позволяет сравнивать медианы разных массивов данных, а именно данные вопросов первой части теста и вопросов второй части теста. Показатели, полученные по этой формуле, считаются менее точными, чем по формуле KR-20. Связано это с приблизительным делением теста пополам. Был рассчитан индекс надежности, исходя из доступных данных, то есть приняв в расчет первые 30 и последние 30 вопросов экзамена. Получился показатель равный 0,658. Этот результат говорит о том, что вопросы теста связаны между собой и тест является относительно надежным.

Рис. 5. Курс философии для ОП Дизайн. Сравнение средних баллов 8 групп по результатам промежуточного теста и экзамена

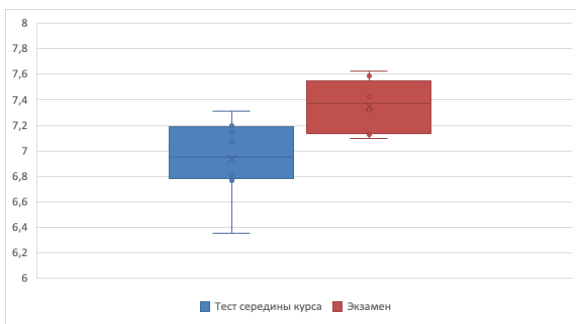
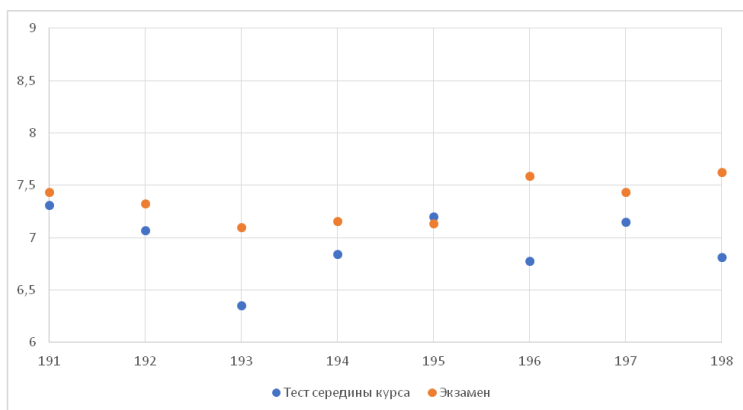


Рис. 6. Курс философии для ОП Дизайн. Сравнение средних баллов по результатам промежуточного теста и экзамена в каждой группе



В дополнение к индексам мы решили проанализировать динамику курса, рассмотрев средние показатели оценки для каждой из 8 групп. Рис. 5 показывает, что разброс среди групп от теста середины курса к экзаменационному тесту уменьшился. Рис. 6 показывает те же данные, но в другом разрезе; на нем мы можем проследить прогресс каждой группы. За исключением одной группы, изначально показывающей хороший результат, все группы свои результаты улучшили. Особенно заметен прогресс в тех группах, которые в начале демонстрировали относительно невысокий результат.

Объективность оценивания может являться вызовом для преподавателя, от которого требуется мотивировать студентов готовиться к тестам, а также поддерживать высокий уровень синхронизации

курса, если его ведут несколько преподавателей. В данном случае можно сказать, что преподавателям удалось мотивировать студентов отстающих групп и приблизить их к лидерам.

6. Трудность и дискриминативность тестовых заданий

Индекс трудности нужен для двух критически важных процедур: 1) для оценки качества вопроса и 2) для регулирования сложности теста.

Регулирование сложности теста, проверяющего знание курса, является сложным процессом, так как зависит не только от качества вопросов, но и от качества подготовки. Также играют роль ожидания студентов обоснованности возложенной на них нагрузки. В этом отношении тест должен иметь определенное распределение частотности вопросов, по индексу трудности попадающих в разные категории (легкие вопросы, вопросы средней сложности, трудные вопросы). Мы не будем приводить частотный разбор вопросов курса по их индексу, так как это будет иметь чисто прикладной интерес только для преподавателей этого курса. Вместо этого мы добавим еще по одной переменной. В одном случае (рис. 7) можно увидеть, как изменятся индексы трудности вопросов, если их задавать повторно. В другом (рис. 8) – распределение индекса трудности в зависимости от преподавателя курса (в нашем случае каждый преподаватель вел по четыре группы).

Рис. 7. Курс философии для ОП Дизайн. Сравнение индекса сложности по вопросам, заданным в тесте середины курса и на экзамене.



Первый график (рис. 7) показывает долю правильных ответов; синяя линия (тест середины курса) почти во всех случаях находится под желтой линией (экзамен), что показывает, что на экзамене студенты ответили лучше, чем в ходе первого тестирования. В случае вопросов 27 и 28 желтая линия отсутствует, так как эти вопросы не задавались на экзамене. Можно говорить о воспроизводимости результатов (по крайней мере, на горизонте нескольких меся-

цев), а также о том, что завершив курс, студенты закрепили знания первых его тем.

Рис. 8. Индекс трудности вопросов второй половины курса по результатам экзамена, данные отображены по преподавателям.

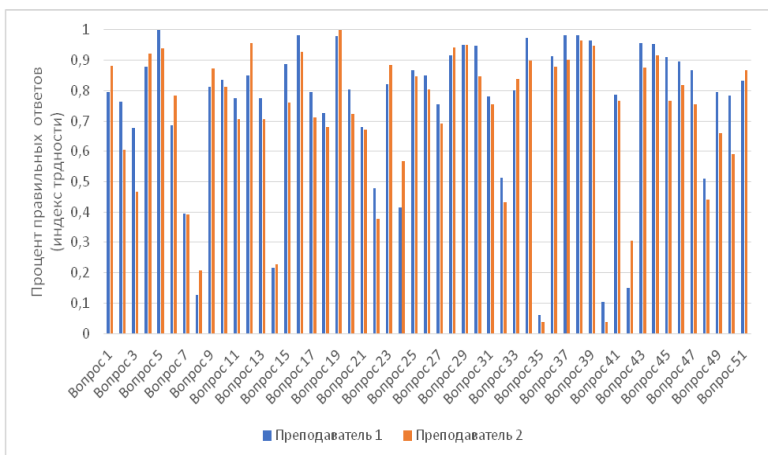
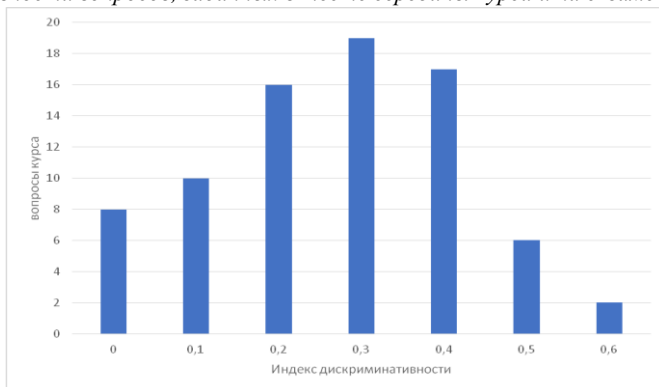


Рис. 8 показывает не только индекс сложности, но и разделение этого индекса по студентам двух преподавателей. На рисунке видно, что есть несколько вопросов, на которые студенты плохо ответили. Их нужно будет пересмотреть. В целом мы наблюдаем очень высокую корреляцию ответов (0,947) между студентами разных преподавателей.

Рис. 9. Курс философии для ОП Дизайн. Распределение индекса дискриминативности вопросов, заданных в тесте середины курса и на экзамене.



Второй критерий, по которому можно оценить эффективность задания, это дискриминативность. Рис. 9 показывает, как, согласно этому индексу, распределились все вопросы курса (данные отра-

жают результаты первого теста). Несмотря на то, что большая часть вопросов имеет приемлемое и даже хорошее (0,4 и выше) значение, из таблицы видно, что как минимум 18 вопросов стоило бы улучшить, так как их индекс ниже 0,2.

Теперь вернемся к рассмотренным выше примерам тестовых заданий. В таблице 1 можно видеть результаты их выполнения студентами. Их анализ позволяет лучше понять, как работают тестовые задания, а учет их результатов может существенно улучшить задания курса. Так, первое и второе задания имеют почти одинаковый индекс трудности, но очень отличаются индексом дискриминации. При этом если сравнить второй и третий вопросы, то второй будет все равно более эффективным, чем третий, хотя третий вопрос оказался более сложным. Четвертый вопрос о Мэри в черно-белой комнате, когда был придуман, субъективно оценивался как хороший, но он оказался очень неинформативным.

Главной стратегией повышения эффективности тестовых вопросов является изменение формулировок дистракторов. Хорошо сформулированный вопрос – вопрос, когда студенты, ошибаясь, выбирают разные дистракторы. Это является признаком того, что дистракторы сформулированы так, что при анализе вариантов отвечающий должен продемонстрировать знание и не попасться в расставленные ловушки. В случае с приведенными заданиями, отвечая на второй и третий вопросы, студенты выбрали разные дистракторы, а отвечая на первый вопрос, ошиблись, выбирая вариант ответа В.

Таблица 1.

Название вопроса	Правильный ответ	Индекс трудности	Индекс дискриминативности
1. Подводная лодка	С.	0,77	0,11
2. Следователь и террорист	Е.	0,76	0,61
3. Аргумент против Декарта	С.	0,71	0,37
4. Мэри в черно-белой комнате	А.	0,92	0,06

Статистический постэкзаменационный анализ показал, что некоторые вопросы теста стоит существенно переработать. Это не самый приятный факт, однако, как показывает исследовательская литература, сформулировать вопросы с низкой эффективностью – нормально; выявление и улучшение таких вопросов явля-

ется частью процесса работы над тестами. Возможная несправедливость в отношении студентов, которым достались слабые вопросы, может быть устранена заранее продуманным методом корректировки оценки [Rudolph et al., 2019].

7. Заключение

Тесты являются только одним из видов заданий. Обсуждение на семинарах, письменные работы, групповая проектная деятельность – все это важная часть образования, позволяющая развить интерес к предмету. Тесты могут измерить полученные умения и навыки, а подготовка к ним – сформировать хорошее знание базовых понятий курса. Тесты могут быть очень гибким инструментом. Создание лишь одного банка заданий и использование его из года в год представляется неверной стратегией. Наоборот, тестовые задания необходимо пересматривать и дополнять новыми заданиями.

Анализ представленных тестов показал, что, несмотря на удачный в целом опыт, многие вопросы теста требуют дальнейшей доработки. Однако такой вывод стоит рассматривать как позитивный результат. Не стоит рассчитывать, что с нуля удастся разработать хорошую систему тестов. Система эффективного тестирования может сложиться как результат апробирования тестовых заданий по мере того, как одна и та же дисциплина будет прочитана неоднократно. Автоматизированное оценивание дает возможность относительно просто анализировать результаты и осуществлять дальнейшие изменения в преподавании на основе данных (data driven curriculum).

Информация о финансировании исследования

Публикация подготовлена в ходе проведения исследования (№ 18-01-0067) в рамках Программы «Научный фонд Национального исследовательского университета «Высшая школа экономики» (НИУ ВШЭ)» в 2018-2019 гг. и в рамках государственной поддержки ведущих университетов Российской Федерации "5-100".

Funding

The article was prepared within the framework of the Academic Fund Program at the National Research University Higher School of Economics (HSE) in 2018- 2019 (grant № 18-01-0067) and by the Russian Academic Excellence Project "5-100".

Информация о конфликте интересов

Автор заявляет об отсутствии конфликта интересов

Declaration of Conflicting Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Список литературы

Аванесов, 2002 – *Аванесов В. С.* Композиция тестовых заданий. Учебная книга для преподавателей вузов, учителей школ, аспирантов и студентов педвузов. 3 изд., доп. М.: Центр тестирования, 2002. 238 с.

Давыдова, Михалева, 2013 – *Давыдова О. В., Михалева Т. Г.* Разработка педагогических тестов по истории. М.: ВАКО, 2013. 189 с.

Дюпре, 2014 – *Дюпре Б.* Философия. 50 идей, о которых нужно знать / пер. с англ. М. Александровой. М.: Phantom press, 2014. 207 с.

Кейс, Свенсон, 1996, web – *Кейс С. М., Свенсон Д. Б.* Создание письменных тестовых вопросов по базисным и клиническим дисциплинам / пер. с англ. Э. Боровик. Национальный Совет Медицинских Экзаменаторов. Филадельфия, 1996. URL: <https://www.testcentr.org.ua/books/sozdaniye-pismennykh-testovykh-voprosov.pdf> (дата обращения: 04.09.2020)

Крокер, Алгина, 2010 – *Крокер Л., Алгина Д.* Введение в классическую и современную теорию тестов: Учебник / пер. с англ. под общ. ред. В.И. Звонникова, М.Б. Чельшковой. М.: Логос, 2010. 668 с.

Лоу, 2007 – *Лоу С.* Философский Тренинг / пер. с англ. А.Л. Никифорова. М.: Аст, 2007. 350 с.

Майоров, 2000 – *Майоров А. Н.* Теория и практика создания тестов для системы образования. М.: Народное образование, 2000. 351 с.

Нагель, 2000 – *Нагель Т.* Что все это значит? Очень краткое введение в философию / пер. с англ. А. Толстова. М.: Идея-Пресс, 2000. 83 с.

Сэндел, 2013 – *Сэндел М.* Справедливость. Как поступать правильно? / пер. с англ. А. Калинина. М.: Манн, Иванов и Фербер, 2013. 338 с.

Burke et al., 2014 – *Burke B. L., Sears S. R., Kraus S., Roberts-Cady S.* Critical analysis: a comparison of critical thinking changes in psychology and philosophy classes // *Teaching of Psychology*. 2014. Vol. 41. No 1. P. 28–36. doi: 10.1177/0098628313514175

Erickson et al., 2006 – *Erickson B. L., Peters C. B., Strommer D. W., Erickson B. L.* Teaching First-year College Students / rev. and expanded ed. San Francisco, CA: Jossey-Bass, 2006. 284 p.

Haladyna et al., 2002 – *Haladyna T. M., Downing S. M., Rodriguez M. C.* A review of multiple-choice item-writing guidelines for class-

room assessment // *Applied Measurement in Education*. 2002. Vol. 15. No 3. P. 309–333. doi:10.1207/S15324818AME1503_5

Impara, Foster, 2013 – *Impara J. C., Foster D.* Item and test development strategies to minimize test fraud // *Handbook of Test Development*. L.: Routledge, 2013. P. 91–114.

Lane et al., 2015 – *Lane S., Raymond M. R., Haladyna T. M.* (eds.). *Handbook of Test Development* / 2nd ed. L.: Routledge, 2015. 659 p.

Mertler, 2014 – *Mertler C. A.* The data-driven classroom: How do I use student data to improve my instruction? Alexandria, VA: ASCD, 2014. 42 p.

Parkes, Zimmaro, 2016 – *Parkes J., Zimmaro D.* Learning and Assessing with Multiple-choice Questions in College Classrooms. L.: Routledge, 2016. 138 p.

Rudolph et al., 2019 – *Rudolph M. J., Daugherty K. K., Ray M. E., Shuford V. P., Lebovitz L., DiVall M. V.* Best practices related to examination item construction and post-hoc review // *American Journal of Pharmaceutical Education*. 2019. Vol. 83. No 7. P. 1492–1503. doi: 10.5688/ajpe7204

Secolsky, Denison, 2017 – *Secolsky C., Denison D. B.* (eds.) *Handbook on Measurement, Assessment, and Evaluation in Higher Education* / 2nd ed. L.: Routledge, 2017. 680 p.

References

Avanesov, V.S. Kompozitaiia testovykh zadaniy [Composition of Test Items]: An tutorial for university teachers, school teachers, graduate students and students of pedagogical universities. 3rd ed., suppl. Moscow: Tsentr testirovaniia Publ., 2002. (In Russian)

Burke, B.L., Sears, S.R., Kraus, S., Roberts-Cady, S. “Critical analysis: a comparison of critical thinking changes in psychology and philosophy classes”, *Teaching of Psychology*, 2014, vol. 41, no. 1, pp. 28–36. doi: 10.1177/0098628313514175

Case, S.M., Swanson, D.B. Constructing Written Test Questions for the Basic and Clinical Sciences / transl. from English by E. Borovik. Philadelphia: National Board of Medical Examiners, 1996. Available at: <https://www.testcentr.org.ua/books/sozdaniye-pismennykh-testovykh-voprosov.pdf> (accessed on September 4, 2020). (In Russian)

Crocker, L., Algina, J. Introduction to Classical and Modern Test Theory / transl. from English under ed. by V.I. Zvonnikov, M.B. Chelyshkova. Moscow: Logos Publ., 2010. (In Russian)

Davydova, O.V., Mikhaleva, T.G. Razrabotka pedagogicheskikh testov po istorii [Development of Pedagogical Tests in History]. Moscow: VAKO Publ., 2013. (In Russian)

Dupre, B. 50 Philosophy Ideas You Really Need to Know / transl. from English by M. Alexandrova. Moscow: Phantom Press, 2014. (In Russian)

Erickson, B.L., Peters, C.B., Strommer, D.W., Erickson, B.L. Teaching First-year College Students / rev. and expanded ed. San Francisco, CA: Jossey-Bass, 2006.

Haladyna, T.M., Downing, S.M., Rodriguez, M.C. “A review of multiple-

choice item-writing guidelines for classroom assessment”, *Applied Measurement in Education*, 2002, vol. 15, no 3, pp. 309–333. doi:10.1207/S15324818AME1503_5

Impara, J.C., Foster, D. Item and test development strategies to minimize test fraud, in: *Handbook of Test Development*. London: Routledge, 2013, pp. 91–114.

Lane, S., Raymond, M.R., Haladyna, T.M. (eds.). *Handbook of Test Development* / 2nd ed. London: Routledge, 2015.

Law, L. *The Philosophy Gym* / transl. from English by A.L. Nikiforov. Moscow: AST Publ., 2007. (In Russian)

Maierov, A. N. *Teoriia i praktika sozdaniia testov dlia sistemy obrazovaniia* [Theory and Practice of Creating Tests for the Education System]. Moscow: Narodnoe obrazovanie Publ., 2000.

Mertler, C.A. *The data-driven classroom: How do I use student data to improve my instruction?* Alexandria, VA: ASCD, 2014.

Nagel, T. *What Does It All Mean?: A Very Short Introduction to Philosophy* / trans. from English by A. Tolstov. Moscow: Idea-Press, 2000. (In Russian)

Parkes, J., Zimmaro, D. *Learning and Assessing with Multiple-choice Questions in College Classrooms*. London: Routledge, 2016.

Rudolph, M.J., Daugherty, K.K., Ray, M.E., Shuford, V.P., Lebovitz, L., DiVall, M.V. “Best practices related to examination item construction and post-hoc review”, *American Journal of Pharmaceutical Education*, 2019, vol. 83, no 7, pp. 1492–1503. doi: 10.5688/ajpe7204

Sandel, M. *Justice: What's the Right Thing to Do?* / transl. from English by A. Kalinin. Moscow: Mann, Ivanov and Ferber Publ., 2013. (In Russian)

Secolsky, C., Denison, D.B. (eds.) *Handbook on Measurement, Assessment, and Evaluation in Higher Education* / 2nd ed. London: Routledge., 2017.

Поступила в редакцию 07.09.2020

The Digital Scholar: Philosopher's lab

Цифровой ученый: лаборатория философа

2020. Т. 3. №3.

Учредитель и издатель: ФГАОУ ВО «Национальный исследовательский
Нижегородский государственный университет им. Н.И. Лобачевского».

Главный редактор: И.Т. Касавин
Зам. главного редактора: А.Н. Ткачев
Ответственный секретарь: С.В. Шибаршина

Дизайн обложки: Н.Д. Асташова
Компьютерная верстка: Т.М. Хусяинов

Формат 60х100 1/16.
Гарнитура «Times», «Calibri»
Уч.-изд. л. 9. Усл. печ. л. 8,85. Тираж 500 экз. Заказ № .
Подписано в печать с оригинал-макета: г.

Отпечатано в типографии Нижегородского государственного университета
им. Н.И. Лобачевского 603000, г. Нижний Новгород, ул. Б. Покровская, 37

Зарегистрирован Федеральной службой по надзору в сфере связи,
информационных технологий и массовых коммуникаций (Роскомнадзор)
Свидетельство ПИ № ФС 77 — 72454 от 05.03.2018

Официальный сайт журнала «The Digital Scholar: Philosopher's Lab /
Цифровой ученый: лаборатория философа»
<http://www.digital-scholar.unn.ru>