

Современные модели и методы оценивания качества учебного процесса

Modern models and methods of assessment the quality of the educational process

doi 10.26310/2071-3010.2022.284.5.006



П. В. Конюховский,

д. э. н., профессор, Российский государственный университет им. А. И. Герцена, г. Санкт-Петербург
✉ kon_pv@mail.ru

P. V. Konyukhovskiy,

professor, department of sectoral economics and finance, Herzen state pedagogical university, St. Petersburg

В статье рассматриваются вопросы создания, развития и применения методик оценивания результатов и качества учебного процесса, которые были бы адекватны новым целям и задачам образовательной деятельности на этапе глобальной цифровой трансформации экономики и общества.

Основное внимание уделяется проблемам согласования абсолютного и относительного подходов к оцениванию. Для решения этой задачи предлагается использовать p -value статистического критерия хи-квадрат. Результаты исследования ориентированы на внедрение и имплементацию в рамках программного обеспечения учебного процесса.

The article deals with the creation, development and application of methods for assessment the results and quality of the educational process, which would be adequate to the new goals and objectives of educational activities at the stage of global digital transformation of the economy and society.

The main attention is paid to the problems of harmonization of absolute and relative approaches to assessment. To solve this problem, it is proposed to use the p -value of the chi-square statistical test. The results of the study are focused on implementation and implementation within the software of the educational process (learning management systems).

Ключевые слова: качество образовательного процесса, оценивание результатов, методики оценивания, статистические модели оценивания.

Keywords: quality of the educational process, assessment of results, assessment methods, statistical assessment models.

Введение

Радикальное изменение роли и места сферы образования, обусловленные глобальной цифровой трансформацией экономики и общества, привели к возрастанию актуальности проблемы адекватного, объективного и достоверного оценивания эффективности образовательной деятельности. Это, в свою очередь, усиливает требования к качеству и адекватности моделей, методик и методологий контрольно-тестовых процедур.

Критика недостатков, объективно присущих традиционной «балльной» системы, стала в некотором роде «общим местом» научных и научно-публицистических работ по данной тематике.

Фундаментальные недостатки балльных методик в первую очередь определяются тем, что это инструменты «доцифровой эпохи», изначально не ориентированные на использование преимуществ современных информационных технологий. Таких как интеллектуальный, предиктивный анализ данных, Big Data, data mining. Более того, даже традиционные («классические») методы математической статистики и эконометрики реализуются в программном обеспечении учебной деятельности (LMS) в очень ограниченных масштабах.

В то же время результаты развития информационных технологий, достигнутые в последние десятилетия, позволяют говорить о начале нового этапа эволюции процедур оценивания результатов учебной деятельности.

Особое значение задачи совершенствования систем оценивания имеют в контексте проблематики инновационных решений в сфере образования. В последние годы наблюдается активный рост внимания к данной тематике и, как следствие, увеличение числа научных публикаций.

Среди них, в частности, заслуживают безусловного внимания исследования организационно-педагогических предпосылок внедрения инноваций в образовательный процесс [15]. В фокусе рассмотрения данной работы находятся вопросы соотношения онлайн- и оффлайн-обучения, анализ объективных и субъективных факторов, определяющих инновационность развития образовательной сферы.

В статье [5] исследуется специфика понятия инноваций применительно к сфере образования, выявляются связи между инновациями в образовании, с одной стороны, и инновациями в предпринимательстве, с другой.

Современные проблемы процедур оценивания результатов образовательного (учебного) процесса

Перманентная цифровая трансформация экономики достаточно быстро охватила сферу образования. Это породило дополнительные требования к результативности и эффективности работы образовательных учреждений. Как в области высшего, так и среднего, специального, начального и дополнительного образования.

Настоящая статья является логическим продолжением ранее опубликованных исследований [9-12]. В них были подняты проблемы соотношения абсолютной и пропорциональной систем оценивания учащихся (в первую очередь, студентов высших учебных заведений). Рассматривались как достоинства, так и недостатки, присущие обеим системам. Это позволило сформулировать методы и подходы к решению задач согласования выводов, получаемых на основе той или иной системы.

Актуальность проблемы не вызывает особых сомнений и возражений. Первоочередным критерием качества работы учебных заведений являются знания и компетенции выпускников, которые были получены ими в ходе образовательного процесса. В качестве принципиального достоинства компетентностного подхода выделяется его ориентация не на сумму знаний, получаемых в процессе обучения, а на развитие способностей успешно действовать в ситуациях, типичных для будущей профессиональной деятельности [25].

К сожалению, показатели, характеризующие компетенции, в подавляющем большинстве случаев не могут быть измерены и оценены непосредственно, с помощью прямых, транспарентных методик. Это обстоятельство порождает обширное семейство косвенных методов и подходов, которые вносят неизбежные искажения и перекосы. Речь идет о таких характеристиках, как, например, обеспеченность площадями, библиотечными фондами либо иными материальными ресурсами. Довольно неоднозначными являются и критериальные показатели, основанные на последующем трудоустройстве выпускников вузов. Нельзя не признать, что карьерные перспективы бывших студентов являются следствием совокупного воздействия целого комплекса трудно наблюдаемых и зачастую нерегистрируемых факторов. Это могут быть региональные различия в доступе к образовательным и информационным ресурсам, флуктуации и асимметрия рынка труда, и, наконец, социальное неравенство.

Дополнительную сложность вносят факторы, вызванные процессами цифровой трансформации экономики и общества. В предшествующих работах [9-12] были рассмотрены критически опасные аспекты цифровизации, связанные с социальной дифференциацией, порождаемой неравенством доступа к информационным ресурсам и подразумевающей разделение общества на управляющее меньшинство, которое «ведет базы данных», и управляемое большинство, на которое «базы данных ведутся». Причем это расщепление носит объективный, определяемый текущим уровнем экономического развития общества, характер. Иными словами, принципиальная возможность управления большинством меньшинством за счет достижений современных информационных технологий будет объективно, неотвратно порождать возникновение такого меньшинства («информационной элиты»).

В этих условиях может существенно возрасти роль университетов и образовательных учреждений как инструментов формирования новой информационной

элиты, обеспечения ее рациональной ротации и мобильности. Было бы неверным утверждать, что ранее учебным заведениям указанная роль была не свойственна. Однако в условиях «трансформационных» социальных инноваций ее значение может радикально усиливаться. Соответственно, возрастает актуальность и востребованность методов и методик оценивания учащихся, на основании которых должны выделяться «достойные переходы на следующий уровень», что, в конечном счете, является необходимым условием формирования качественного, эффективного «управляющего меньшинства», способного реализовывать сбалансированную стратегическую линию, ориентированную на фундаментальные социальные цели, а не на локальные эгоистические интересы.

Еще одним важным социальным эффектом, порожденным цифровой трансформацией, стало изменение способов восприятия информации широкими слоями населения. Невозможно отрицать, что новые поколения людей, воспитываемые в среде, насыщенной (и в определенном смысле — перенасыщенной) мобильными устройствами, вырабатывает качественно иные методы и навыки работы с информационными сигналами и вызовами окружающего мира.

Например, теряются компетенции, предполагающие наличие хорошей памяти, объема знаний, которыми человек может воспользоваться без обращения к внешним источникам информации. У значительного числа учащихся наблюдается снижение способностей самостоятельно формулировать мысли и суждения. Наоборот, возрастает роль компетенций, позволяющих быстро найти нужные сведения во внешней информационной среде, в первую очередь, в Интернете. Более того, в последние годы даже эти компетенции вытесняются их имитациями и так называемыми «симулякрами». Реальные навыки по поиску нужной информации сменяются иллюзиями класса «если мне действительно понадобится, то я всегда смогу найти». Что, вообще говоря, совсем не обязательно соответствует реальности. Подобные представления подпитываются перманентным пребыванием в «информационном море», что и порождает необоснованные ощущения причастности к окружающим знаниям.

Указанные изменения информационного фона общества носят массовый и объективный характер, что, безусловно, предъявляет новые требования к организации и методикам образовательных процессов. Неправомерными и неконструктивными являются надежды на возврат к «прекрасному прошлому», равно как и сетования на то, что «студент пошел не тот». Все реалистичные пути решения текущих проблем, так или иначе, завязаны на смену образовательных технологий.

При этом речь идет не о какой-то новой «чудо-технологии», которая заменит ранее существовавшие, а о переходе к процессу перманентных инновационных изменений образовательных методик, то есть такому порядку, когда их непрерывное последовательное замещение воспринимается как нечто естественное и необходимое. Ускорение процессов общественного развития, смены старых норм новыми правилами не-

предложно диктует адекватные требования в части гибкости и адаптируемости в образовательной сфере.

В ситуации увеличения гибкости и, в определенном смысле, нестабильности учебно-методических схем и технологий особую актуальность приобретает проблематика релевантности методик анализа их эффективности и последствий применения. В течение последних лет в профессиональном дискурсе значительное внимание уделялось таким понятиям, как индивидуальные образовательные траектории, устаревание знаний, адаптируемость к новым условиям. Крайне популярным стал термин софт скиллз (soft skills — мягкие, гибкие навыки). Под ними понимают широкий спектр умений, кросс-функциональных навыков, которые оказываются востребованными безотносительно к конкретной профессии. В то же время, для того, чтобы благие интенции не превращались в банальные декларации, необходимы релевантные методики оценивания.

Дискуссия о системах оценивания и оценках к настоящему моменту имеет достаточно представительную историю. Среди знаковых выступлений на эту тему можно отметить выступление видного российского журналиста и общественного деятеля Ивана Засурского [7], опубликованное под броским заголовком «система троечников должна быть снесена». В нем сформулирован достаточно яркий тезис о том, что «люди, которые могли бы умножать знания, работают на фальсификацию знаний». Отмечается, что фальсификация образовательного и научного статуса в современных условиях становятся особой системой обращения капитала.

Достаточно яркие и острые тезисы сформулированы в выступлении спецпредставителя президента РФ по вопросам цифрового и технологического развития Дмитрия Пескова [16]. Он охарактеризовал практику выдачи дипломов «на всю жизнь» как абсурдную. Основной аргумент выглядит достаточно убедительно — невозможно гарантировать, что полученные знания с течением времени сохранятся и будут актуальны. В качестве ведущего требования к современной системе образования названа высокая скорость разработки учебных программ, удовлетворяющих потребностям рынка труда, нацеленность на подготовку команд профессионалов, а не отдельных специалистов. Отдельное внимание уделено необходимости отказа от стандартов, построенных в «аналоговом мире».

Европейская образовательная система имеет богатый опыт решения проблем, связанных с конвертацией баллов при переходе студентов из одного университета в другой. До 2009 г. согласно Европейской системе перевода и накопления баллов (European Credit Transfer and Accumulation System, ECTS) абсолютные оценки выставлялись в зависимости от места студента в рейтинге, см. рис. 1, см. [26].

Данная система подвергалась серьезной и острой критике. В первую очередь по причинам игнорирования специфики и уровня различных университетов, а также содержательных аспектов образовательных программ. После 2009 г. осуществляется процедура непосредственной конвертации оценок между двумя

университетами на основе принципа соответствия процентилей, то есть на основе оценки, полученной в одном университете, выставляется оценка, ближайшая к процентиллю оценок другого университета.

Собственно, системы оценок, используемые на настоящий момент, в учебных заведениях различных странах является отдельным, важным и чрезвычайно интересным объектом для изучения. Если обратиться к открытым общедоступным источникам [18, 30], то несложно осознать широчайшее разнообразие, которое существует на текущий момент в области методов, методик и алгоритмов оценивания. Даже для таких исторически близких как экономически, так и духовно стран, каковыми являются Россия и Белоруссия, системы оценивания различаются. В то время, как в Российской Федерации по-прежнему доминирует «классическая» пятибалльная система, Белоруссия перешла на 10-балльную шкалу. Дополнительным аргументом в пользу тезиса о сложности и многообразии оценочных систем служит описание ECTS [26], особенно с учетом ее различных версий.

Достаточно существенные различия можно обнаружить и на уровне отдельных российских вузов. Например, и в Санкт-Петербургском государственном университете (СПбГУ), и в Российской академии народного хозяйства и государственной службы (РАНХиГС) используется «буквенная» система оценивания, но при этом в СПбГУ оценка «В» интерпретируется как «очень хорошо», а в РАНХиГС как «отлично», которое «чуть менее отлично, чем отлично — А». Другие университеты, например, МГИМО (Московский государственный институт международных отношений), имеют собственные подходы [31]. Подсчет GPA, [27] (Grade Point Average) — усредненного балла по аттестату или диплому за весь период обучения, величина которого является важнейшим условием поступления в зарубежный университет, как правило, оказывается очень непростой и нетривиальной задачей [27].

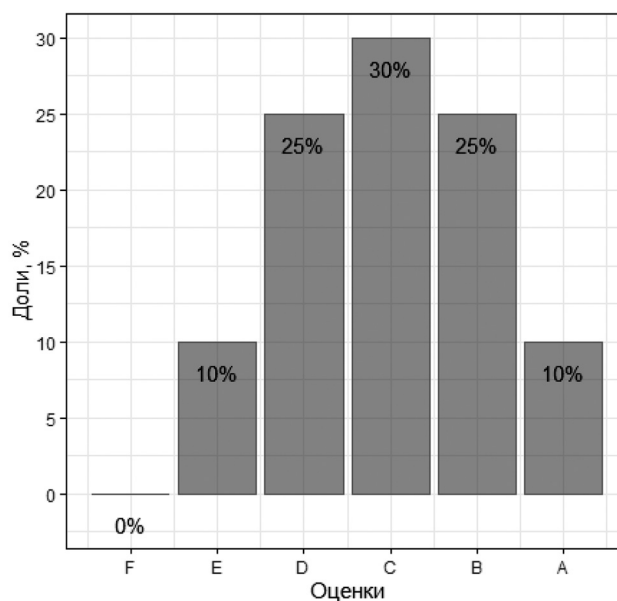


Рис. 1. Метод конвертирования оценок в рамках ECTS, до 2009 г.

Предшествующие исследования

К настоящему моменту существует обширное множество научных исследований, посвященных проблематике оценивания результатов образовательного процесса.

В работе сотрудников Московского института электроники и математики НИУ ВШЭ [6] рассматриваются вопросы повышения эффективности учебных мероприятий за счет усиления обратной связи со студентами. Акцент делается на меры регулярного контроля, в том числе не только на семинарских и лабораторных занятиях, но и на лекциях. По мнению авторов, процедуры сбора и перманентного статистического анализа данных и организованного на их основе мониторинга позволяют выявить широкий спектр закономерностей, определяющих успеваемость студентов, их мотивацию и, в конечном счете, повысить качество освоения учебного материала.

Особого внимания заслуживают работы, посвященные применению «продвинутых» математических моделей и методов. Статья [19] посвящена анализу проблем перехода от непараметрической статистики успеваемости отдельного студента к параметрической статистике описания успеваемости большой выборки студентов. Авторами предложена предварительная классификация выборок студентов по признаку отсутствия функциональной зависимости дисперсии от средней успеваемости. Также в этой работе сформулированы конкретные рекомендации для условий перехода от непараметрической статистики к параметрической на примере анализа средней успеваемости студентов.

Работа [13] посвящена применению математического аппарата марковских цепей с дискретным временем перехода для моделирования динамики показателей успеваемости студентов. По результатам проведенного анализа авторы приходят к заключению о нестационарности марковской цепи, что интерпретируется как отсутствие предопределенности в учебных достижениях студентов. Как следствие, надежность предсказаний будущих результатов вызывает вполне обоснованные сомнения.

Достаточно представительным является класс работ, в которых на основе методов эконометрического анализа исследуются взаимосвязи между результатами ЕГЭ абитуриентов и их последующей успеваемостью в вузах.

Одним из первых исследований в данном направлении стала статья [17]. В ней рассматриваются проблемы изучения зависимости среднего балла студентов первого курса от результатов их вступительных испытаний. В качестве эмпирической базы была использована статистика по факультету экономики ВШЭ. Автором предложена модель бинарного выбора для вероятности выбытия как функции результатов ЕГЭ. В работе делаются заключения о позитивных, в целом, свойствах модели. Также отмечается, что наибольшее влияние на вероятность отчисления оказывает ЕГЭ по математике.

В работе [3] по данным «нестоличных» российских университетов на основе регрессионного анализа

влияния баллов ЕГЭ на успеваемость выявлены дополнительные факторы: в частности, средний балл аттестата, место жительства студента, временной лаг между окончанием школы и началом обучения в университете.

Одной из активно дискутируемых тем является так называемое «оценивание по кривой» (grading on curve, marking on curve). Под этим термином понимается система методов, применяемых преподавателями для корректировки «распределений» оценок. По умолчанию речь идет о приближении распределения оценок, выставляемых в рамках той или иной группы (популяции) студентов к нормальному. С одной стороны, оценивание по кривой является мерой, противодействующей инфляции оценок, предотвращающей ситуации уравнивания успехов, при которых более половины класса может получить отличную оценку. С другой стороны, данный тип оценок конфликтует с принципом оценивания «по критериям» и, в соответствии с достижением абсолютных результатов, может порождать токсичную атмосферу среди учащихся, стимулировать нездоровую конкуренцию. Данные проблемы, в частности, поднимаются в работе [29].

Важным и актуальным направлением является сравнительный анализ образовательных систем различных стран. Результаты социально-экономического развития, которые демонстрирует Китай в течение последних десятилетий, безусловно, привлекают пристальное внимание к опыту организации образовательной сферы в данной стране. Среди работ, посвященных данной проблематике, можно отметить статью [21]. В ней достаточно подробно раскрывается специфика организации системы высшего образования в КНР и прослеживаются механизмы ее влияния на экономические успехи страны.

Статья [1] затрагивает вопросы качества образования с позиций субъективных показателей. В работе поднимаются вопросы о значимости отношения студентов к своему образованию как к важнейшему фактору повышения результативности учебного процесса. Что, в конечном счете, оказывается главным условием эффективного использования научно-педагогического потенциала вуза. Проблемы рейтинговой системы оценивания успешности студентов в Китае, ее достоинства и недостатки рассматриваются в работе [14].

В научном направлении, относящемся к проблематике оценивания образовательного процесса, можно выделить два принципиальных субнаправления. А именно, работы, посвященные методам анализа и интерпретации результатов. Их примеры приведены выше. С другой стороны — работы, посвященные анализу оценочных средств (контрольно-тестовых материалов, заданий и т. п.)

Среди исследований этого субнаправления можно выделить статью [20]. В ней предлагается модель оценивания контрольно-тестовых материалов, опирающаяся на анализ вероятностных распределений успеваемости студентов. Инструментальной основой модели является математический аппарат теории марковских процессов. На основе сконструированной модели получены и проанализированы вероятности переходов между студенческими группами успеваемости

и устойчивые распределения по группам успеваемости, характеризующие контрольно-измерительные материалы. Авторами особо подчеркивается актуальность исследования для онлайн-курсов, которые становятся непременным атрибутом образовательного процесса в условиях цифровой трансформации.

Говоря о современных или относительно современных методах исследования, нельзя упускать из внимания те методы, которые эффективно и конструктивно проявили себя в прошлом. При этом их потенциал далеко не исчерпан. В этом ряду можно отметить методы принятия решений в условиях риска и неопределенности. Они, в частности, получили плодотворное развитие в работах [22-24]. В них предложен фундаментальный теоретический аппарат принятия решений в условиях неполной, неточной, нечисловой информации (так называемой ННН-информации). На его основе могут быть сформированы алгоритмы построения интегрированных сводных показателей, что представляет несомненный интерес с точки зрения задач оценивания результатов образовательного процесса.

Методология исследования

Традиционные методики и модели оценивания образовательного процесса, так или иначе, могут быть разделены на два принципиальных класса:

- абсолютное оценивание — оценивание на основе сопоставления результатов контрольно-тестовой процедуры с некоторой абсолютной шкалой;
- относительное оценивание — оценивание на основе разбиения совокупности (группы) проверяемых на подгруппы (кластеры) в соответствии с заданными пропорциями.

Не вызывает сомнений то, что в зависимости от применяемого принципа результаты оценочных процедур могут ощутимо различаться. Это, в свою очередь, порождает задачу их согласования, как максимум, и осмысления причин расхождения, как минимум.

В этом плане перспективным представляется подход, основанных на использовании статистических критериев согласия. На начальных этапах исследования в роли вполне приемлемого инструмента может выступить критерий χ^2 Пирсона. В последующем могут быть рассмотрены и другие критерии согласия, например, критерий Колмогорова-Смирнова.

Напомним, что «исходная» сфера применения критерия χ^2 связана с проверкой статистической гипотезы о принадлежности эмпирической выборки теоретической генеральной совокупности. Последняя, как правило, идентифицируется с некоторой случайной величиной с известной функцией распределения.

В рамках решаемой нами задачи в роли теоретического распределения выступают эталонные пропорции распределения проверяемых учащихся по группам (карманам, кластерам), соотносимым с балльными оценками.

Преобразованное эталонное распределение представлено в табл. 1. Максимально возможный результат за выполнение контрольно-тестового задания полага-

ется равным единице. Интервал допустимых значений результаты разбивается на равные отрезки длиной в 0,1 ($0 \leq 0,1 \leq 0,2 \dots \leq 0,9 \leq 1$). Каждому из отрезков ставится в соответствие доля от полной совокупности тестируемых учащихся, чей результат должен оказаться в пределах данного отрезка. Например, ожидается, что при «идеальной» («эталонной») тестовой процедуре результат от 0,9 до 1 (от 90 до 100%) покажут 5% учащихся, результат от 0,8 до 0,9 — 10% и т. д.

Как несложно догадаться, конкретные числа не оказывают существенного воздействия на работоспособность рассматриваемой методики.

Эталонное распределение можно рассматривать в качестве генеральной совокупности. Результаты конкретных контрольно-тестовых процедур при этом получают интерпретацию эмпирических выборок. Как следствие, вполне корректной и осмысленной представляется задача проверки гипотезы об их принадлежности генеральной совокупности, т. е. получения ответа на вопрос «можно ли их считать частными реализациями случайной величины, ассоциируемой с генеральной совокупностью?».

В ходе проверки гипотезы с помощью критерия χ^2 находится соответствующее p -value (уровень значимости — максимальное значение ошибки первого рода для значения эмпирической статистики $\chi^2_{\text{эмп}}$).

При «высоких» значениях p -value гипотеза о принадлежности эмпирической выборки генеральной совокупности не отвергается. При малых p -value — отвергается, т. е. появляются веские основания для выводов о существенных расхождениях между «эталонном» (ожидаемыми результатами) и результатами фактическими.

Очевидным «простым» выходом из подобных ситуаций является переработка методического контента тестовых процедур (так называемого фонда оценочных средств).

В то же время нельзя игнорировать направления совершенствования процедур оценивания, предполагающего корректировку «эталонных распределений».

С учетом логики предшествующего исследования может быть предложена процедура гибкого (динамического) мониторинга текущих результатов на основе их сопоставления с эталонными распределениями, характеризующимися относительно несложными (транспарентными) параметрами. На эту роль, в частности, могут претендовать:

- асимметричное треугольное определение;
- бета-распределения, в том числе их частный случай — распределение PERT.

Асимметричное треугольное распределение на интервале [0; 1] однозначно определяются модой, см. рис. 2.

Процедура предполагает выбор такого значения моды, которое обеспечивает минимальное значение эм-

Таблица 1
Примеры «эталонных» распределений оценок

$\leq 0,1$	$\leq 0,2$	$\leq 0,3$	$\leq 0,4$	$\leq 0,5$	$\leq 0,6$	$\leq 0,7$	$\leq 0,8$	$\leq 0,9$	≤ 1
1,25%	1,25%	2,5%	5%	5%	30%	20%	20%	10%	5%

Источник: разработано автором

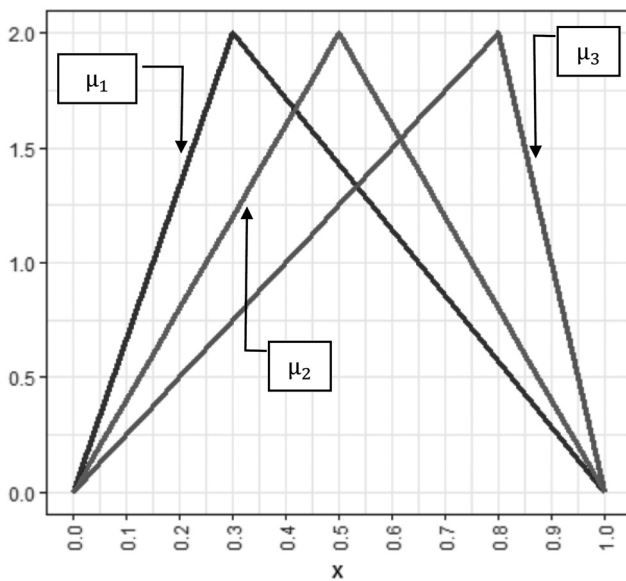


Рис. 2. Графики плотности эталонных треугольных распределений на интервале [0; 1] с различными модами ($\mu_1=0,3; \mu_2=0,5; \mu_3=0,8$)

пирической статистики $\chi^2_{эмп}$ (максимальное p -value), рассчитываемых при проверке гипотезы о принадлежности эмпирической выборки (частных результатов) генеральной совокупности.

Значение моды, на котором достигается наилучший результат, определяет распределение, которое в последующем может рассматриваться в качестве кандидата в «эталонные».

Информационная база исследования

В качестве информационной базы исследования были использованы данные по:

- 1) финальному экзаменационному тесту по курсу «Общая теория статистики», Институт экономики и управления, РГПУ им. А. И. Герцена, 2020-2022 гг., всего 126 наблюдений (информация по структуре текста представлена в табл. 2);

- 2) финальному экзаменационному тесту по курсу «Социально-экономическая статистика», Институт экономики и управления, РГПУ им. А. И. Герцена, 2020-2022 гг., всего 103 наблюдения (информация по структуре текста представлена в табл. 3);
- 3) финальному экзаменационному тесту по курсу «Эконометрика», Институт экономики и управления, РГПУ им. А. И. Герцена, 2021-2022 гг., всего 45 наблюдений (информация по структуре текста представлена в приложении, табл. А-1);
- 4) финальному экзаменационному тесту по курсу «Исследование операций», Северо-Западный институт управления РАНХиГС, 2022 г., всего 69 наблюдений (информация по структуре текста представлена в приложении, табл. А-2);
- 5) финальному экзаменационному тесту по курсу «Теория игр», Северо-Западный институт управления РАНХиГС, 2022 г., всего 49 наблюдений (информация по структуре текста представлена в приложении, табл. А-3);
- 6) финальному экзаменационному тесту по курсу «Моделирование конфликтных ситуаций в социально-экономических процессах», Северо-Западный институт управления РАНХиГС, 2022 г., всего 42 наблюдения (информация по структуре текста представлена в приложении, табл. А-4).

Первичная обработка данных предполагает их нормировку (масштабирование) к базовому интервалу [0; 1]. Принципиальная схема преобразования (для случая максимально возможной оценки за тест, равной 40 баллам) представлена на рис. 3.

Процедура применения критерия χ^2 , вообще говоря, не требует предварительных трансформационных (масштабирующих) преобразования. Однако в данном случае ее целесообразность обосновывается последующими возможностями сравнительного анализа результатов тестовых процедур с различными значениями максимальных баллов.

Таблица 2

Структура теста по курсу «Общая теория статистики», Институт экономики и управления, РГПУ им. А.И. Герцена, 2020-2022 гг.

Общая теория статистики ($\Sigma_i q_i=40$)									
Q-01	Q-02	Q-03	Q-04	Q-05	Q-06	Q-07	Q-08	Q-09	Q-10
1	1	1	1	2	2	2	2	2	2
Q-11	Q-12	Q-13	Q-14	Q-15	Q-16	Q-17	Q-18	Q-19	Q-20
2	2	2	2	2	2	3	5	1	1

Источник: разработано автором

Таблица 3

Структура теста по курсу «Социально-экономическая статистика», Институт экономики и управления, РГПУ им. А.И. Герцена, 2020-2022 гг.

Социально-экономическая статистика ($\Sigma_i q_i=40$)								
Q_01	Q_02	Q_03	Q_04	Q_05	Q_06	Q_07	Q_08	Q_09
2	3	1	2	2	2	2	2	2
Q_10	Q_11	Q_12	Q_13	Q_14	Q_15	Q_16	Q_17	Q_18
2	2	3	3	1	2	3	3	3

Источник: разработано автором

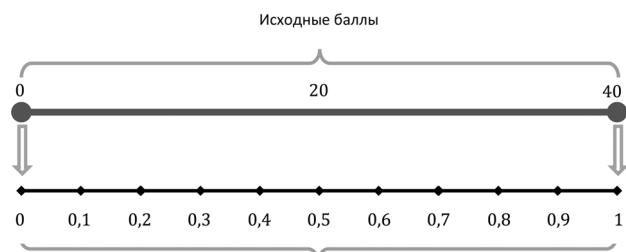


Рис. 3. Схема трансформации исходных баллов в аналитические

Эмпирические исследования, апробация

Рис. 4 визуализирует процедуру сравнительного анализа фактических результатов, показанных студентами при выполнении финального теста по курсу «Общая теория статистики», с эталоном, заданным в табл. 1. Разбиение полного интервала значений на 10 карманов (кластеров) с точки зрения конкретики и практики реальных тестовых процедур представляется вполне рациональным и обоснованным.

Диаграмма (рис. 4) наглядно отражает радикальные различия между фактом и эталоном ($\chi^2_{эмп} \approx 458,99$, $p\text{-value} \rightarrow 0$). В подобной ситуации очевидным представляется решение о радикальной корректировке учебно-методического обеспечения соответствующих оценочных средств.

Результаты аналогичной процедуры для данных по результатам контрольно-тестовой процедуры по курсу «Социально-экономическая статистика» представлены на рис. 5.

В данном случае гипотеза о принадлежности выборки фактических результатов эталонной генеральной совокупности также должна быть отвергнута с $p\text{-value} = 0,0002$ ($\chi^2_{эмп} \approx 31,97$). Однако на уровне визуального анализа мы можем констатировать «качественно» иной уровень близости между фактом и эталоном по сравнению с предшествующим примером.

Результаты аналогичного сравнительного анализа по другим наборам данных представлены в приложении, см. рис. А-1–А-4.

Существенные расхождения между фактическими и эталонными выборками являются вполне

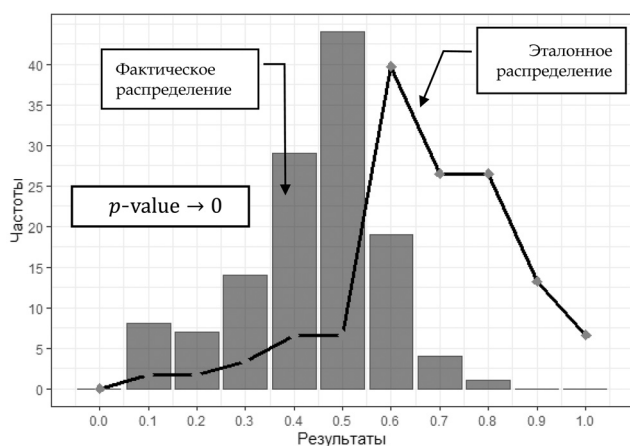


Рис. 4. Эмпирическое и эталонное распределение, результаты теста по курсу «Общая теория статистики», ИЭиУ, РГПУ, 2020-2022 гг.

естественными и предсказуемыми. Скорее, их близость свидетельствовала бы о случайном совпадении либо подгонке.

Объективная природа контрольно-тестовых процедур в образовательных процедурах неизбежно порождает «уходы и смещения» от изначального целевого состояния. Поэтому крайне важны методики, позволяющие оперативно выявлять факты отклонения и, что особенно важно, строить их количественные характеристики.

Методы корректировки оценочных систем

Как уже упоминалось выше, «естественным» направлением развития инструментального аппарата, отражающего соответствие между фактическими результатами и эталонными целеполаганиями, является разработка моделей и методов, позволяющих выработать корректирующие воздействия на системы оценивания. В частности, в рамках концептуальных подходов, развиваемых в настоящей работе, может быть предложен алгоритм, основанный на аппроксимации фактического (наблюдаемого) распределения результатов одним из «стандартных» («классических») распределений.

Процедура параметрического анализа систем оценивания предполагает выбор конкретного распределения, характеризующегося небольшим количеством параметров (в идеале — с одним параметром). Далее осуществляется следующая последовательность шагов:

- вариация параметра;
- для каждого очередного значения параметра проводится проверка гипотезы о принадлежности фактических результатов генеральной совокуп-

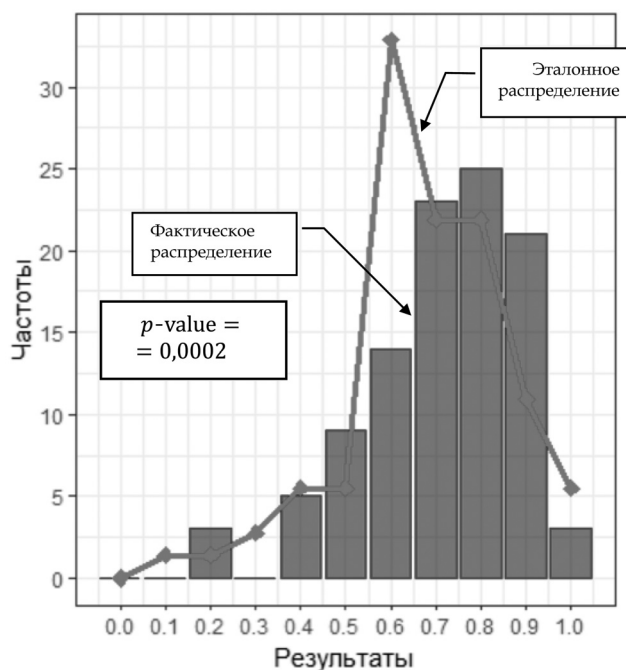


Рис. 5. Эмпирическое и эталонное распределение, результаты теста по курсу «Социально-экономическая статистика», ИЭиУ, РГПУ, 2020-2022 гг.

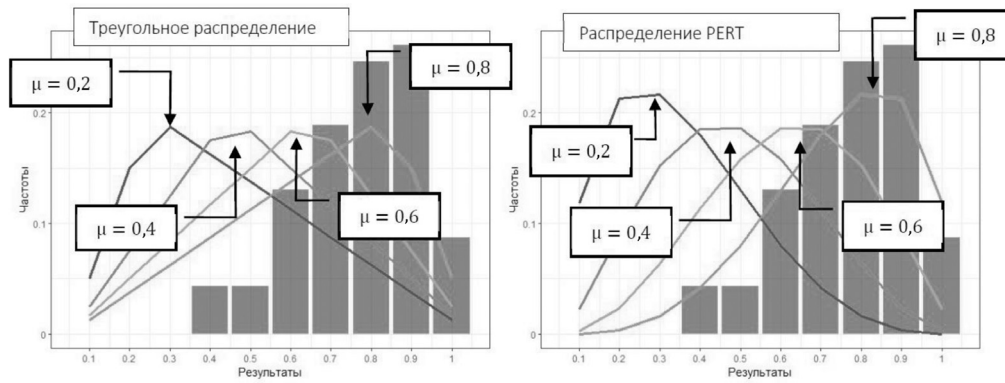


Рис. 6. Поиск эталонной системы оценивание на основе анализа результатов теста по курсу «Исследование операций», СЗИУ, РАНХиГС, 2022 гг.

ности, соответствующей выбранному эталонному распределению, определяются $\chi^2_{эмп}$ и p -value.

На рис. 6 представлена процедура поиска эталонного распределения (системы оценивания), которая бы соответствовала наилучшим образом фактическим результатам тестирования по курсу «Исследования операций» (2022 г.) в смысле критерий χ^2 .

На диаграмме слева (рис. 6) на фактическую гистограмму результатов наложены теоретические графики плотности асимметричных треугольных распределений с различными значениями мод ($\mu=0,2; 0,4; 0,6; 0,8$). Правая диаграмма аналогична в плане принципов построения с той лишь разницей, что в ней используются плотности распределений PERT.

Как можно заметить, на правой диаграмме наблюдается близость фактических и теоретических частот в случае распределения PERT с модой $\mu=0,8$. Действительно, соответствующее значение p -value=0,898, что позволяет сделать выводы в пользу целесообразности использования данной версии распределения в качестве основы для формирования эталона системы оценивания для дисциплин подобного класса (направленности).

На последующих шагах (в ходе практического внедрения) на основе эталонного распределения могут быть рассчитаны абсолютные значения порогов для выставления абсолютных оценок.

Заключение

Предложенная методика может активно и конструктивно применяться в мероприятиях по совершенствованию качества работы и эффективности университетов, а также других образовательных учреждений, повышению уровня адекватности и оперативности реагирования на вызовы внешней среды.

Очевидно проблемной стороной процессов взаимодействия систем оценивания является их разномасштабность, приводящая к искажениям и информационным потерям при конвертации. Нельзя не признать, что критерий χ^2 несет в себе изъян схожей природы — чувствительность к «нарезке» на карманы. Частичным способом преодоления данного недостатка может стать вариативный подход, подразумевающий параллельное оценивание по разным системам частотных карманов. В любом случае нельзя отрицать

того обстоятельства, что методики, основанные на критериях согласия, являются «эффективными сигнализаторами», заставляющими задуматься о влиянии факторов, порождаемых произволом в выборе систем шкал и интервальных разбиений.

Развивая тему перспектив развития инновационных мер оценивания результатов образовательного процесса, необходимо обратить внимание на проблему соответствия оценок фундаментальным задачам обучения.

Иногда (в не очень серьезных контекстах) ее называют «проблемой Тома Сойера». Имеется ввиду сюжет из романа Марка Твена, повествующий о том, как Том в итоге стал лучшим в изучении Библии за счет обменных манипуляций с билетиками, выдававшимися за выполнение промежуточных заданий.

В наше время образовательные учреждения всех уровней столкнулись с эффектом серьезных провалов оценочно-тестовых систем, ситуациями, когда учащийся, получивший высокий балл, по факту обладает весьма поверхностными, а иногда и просто неудовлетворительными знаниями.

В значительной мере это обуславливается недостатками и просчетами в дизайне тестовых систем. Усугубляющим фактором становятся новые технологические возможности. Действительно, учебно-методический контент, успешно проявлявший себя в эпоху «ручного контроля», легко компрометируется и теряет проверочную силу при контроле автоматизированном.

В роли эффективного инструмента решения перечисленных проблем могут выступать методы согласования абсолютных и относительных систем оценивания результатов, предложенные в настоящей работе. В частности, достаточно перспективным и конструктивным направлением представляется интеграция моделей компаративного анализа факта и эталона в автоматизированные системы обучения (LMS Moodle).

Условия глобальной цифровой трансформации радикально меняют содержательные смыслы, вкладываемые в само понятие «знания». Меняется отношение к знаниям, умениям, навыкам как результатам учебного процесса. Это, в свою очередь, требует принципиальных изменений в подходах в процедурах оценивания знаний в их новом качественном понимании.

Таблица А-1

Структура теста по курсу «Эконометрика», Институт экономики и управления, РГПУ им. А.И. Герцена, 2021-2022 гг.

Эконометрика ($\sum_i q_i=40$)									
Q_01	Q_02	Q_03	Q_04	Q_05	Q_06	Q_07	Q_08	Q_09	Q_10
2	2	2	2	2	2	2	2	2	2
Q_11	Q_12	Q_13	Q_14	Q_15	Q_16	Q_17	Q_18	Q_19	Q_20
2	2	2	2	2	2	2	2	2	2

Источник: разработано автором

Таблица А-2

Структура теста по курсу «Исследование операций», Северо-Западный институт управления РАНХиГС, 2022 г.

Исследование операций ($\sum_i q_i=40$)							
Q_01	Q_02	Q_03	Q_04	Q_05	Q_06	Q_07	
1	2	1	2	3	1	3	
Q_08	Q_09	Q_10	Q_11	Q_12	Q_13	Q_14	Q_15
3	3	3	3	3	4	4	4

Источник: разработано автором

Таблица А-3

Структура теста по курсу «Теория игр», Северо-Западный институт управления РАНХиГС, 2022 г.

Теория игр ($\sum_i q_i=40$)									
Q_01	Q_02	Q_03	Q_04	Q_05	Q_06	Q_07	Q_08	Q_09	Q_10
1	1	1	1	1	2	2	3	3	2
Q_11	Q_12	Q_13	Q_14	Q_15	Q_16	Q_17	Q_18	Q_19	Q_20
3	2	1	2	3	2	2	2	3	3

Источник: разработано автором

Таблица А-4

Структура теста по курсу «Моделирование конфликтных ситуаций...», Северо-Западный институт управления РАНХиГС, 2022 г.

Моделирование конфликтных ситуаций ($\sum_i q_i=40$)									
Q_01	Q_02	Q_03	Q_04	Q_05	Q_06	Q_07	Q_08	Q_09	Q_10
1	1	1	1	1	2	2	3	3	2
Q_11	Q_12	Q_13	Q_14	Q_15	Q_16	Q_17	Q_18	Q_19	Q_20
3	2	1	2	3	2	2	2	3	3

Источник: разработано автором

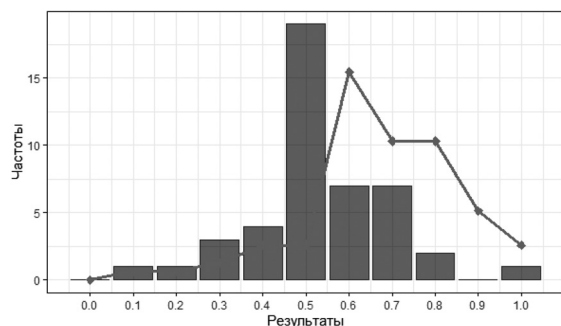


Рис. А-1. Эмпирическое и эталонное распределение, результаты теста по курсу «Эконометрика», ИЭИУ, РГПУ, 2020-2022 гг.

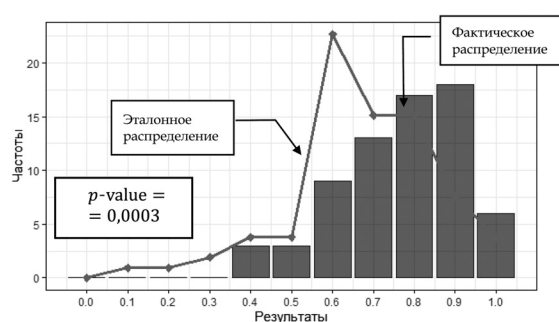


Рис. А-2. Эмпирическое и эталонное распределение, результаты теста по курсу «Исследование операций», СИУ РАНХиГС, 2022 гг.

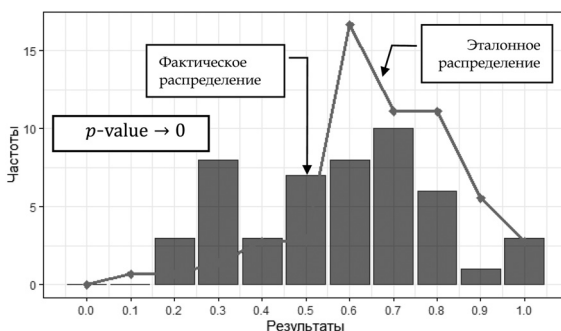


Рис. А-3. Эмпирическое и эталонное распределение, результаты теста по курсу «Теория игр», СИУ РАНХиГС, 2022 гг.

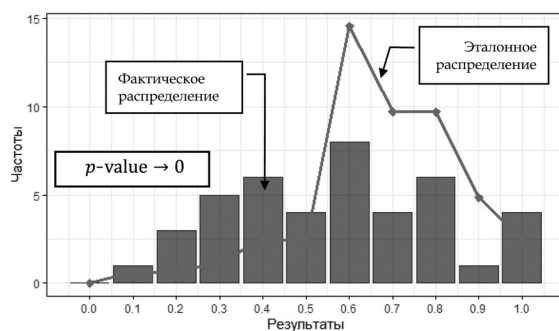


Рис. А-4. Эмпирическое и эталонное распределение, результаты теста по курсу «Моделирование конфликтных ситуаций...», СИУ РАНХиГС, 2022 гг.

Список использованных источников

1. Т. И. Боровкова, Т. Д. Лавриненко, Жи Чун Хэн, Чжу Цзяу. Оценивание качества образования студентов в России и Китае//Современные проблемы науки и образования. 2017. № 1. <https://science-education.ru/ru/article/view?id=26105>.
2. В вузах одобряют ограничение получающих повышенную стипендию. <https://tass.ru/obschestvo/6054148>.
3. Н. С. Галимова, Л. Р. Загитова. Построение многофакторной модели успеваемости студента//Международный научно-исследовательский журнал. № 6 (96). 2020. Ч. 3, июнь. С. 31-36. <https://research-journal.org/social/postroenie-mnogofaktornoj-modeli-uspevaemosti-studenta>.
4. Европейская система перевода и накопления баллов. https://ru.wikipedia.org/wiki/Европейская_система_перевода_и_накопления_баллов.
5. А. К. Ерохин, Н. А. Глушенко. Инновации в образовании: прошлое и перспективы будущего//Научно-методический электронный журнал «Концепт». 2017. № 53.
6. Е. А. Ерохина, Д. В. Хрустова, Э. С. Клышинский, Ю. В. Жури. Анализ статистики успеваемости студентов как средство повышения качества образования//Инженерное образование. 2014. № 154. http://aeer.ru/files/io/m15/art_28.pdf.
7. И. Засурский. Система троечников будет снесена. <https://aurora.network/articles/39-nauka-i-obrazovanie/56823-sistema-troechnikov-budet-snesena>.
8. П. В. Коноховский, А. О. Ольховик, А. С. Алипов. Модели оценивания качества высшего образования как инструмент управления развитием человеческого капитала//Актуальные проблемы экономики и менеджмента. 2017. № 4 (16). С. 88-94.
9. П. В. Коноховский. О методах оценивания результатов и эффективности образовательной деятельности в современных условиях//В сборнике: «Государство и бизнес. Современные тенденции и проблемы развития экономики». Материалы XIII Международной научно-практической конференции. В 3-х ч. Санкт-Петербург, 2021. С. 120-132.
10. П. В. Коноховский. Оценивание результатов образовательной деятельности: методы и решения//В сборнике: «Менеджмент XXI века: социально-экономическая трансформация в условиях неопределенности». Сборник научных статей по материалам XVIII Международной научно-практической конференции. Санкт-Петербург, 2020. С. 209-218.
11. П. В. Коноховский, А. О. Ольховик, А. С. Кузнецова. Экономико-математическое моделирование рынка высшего образования в условиях экономического спада//Государственное управление. Электронный вестник. 2016. № 58. С. 340-358.
12. П. В. Коноховский, А. С. Алипов. Проблемы развития и совершенствования систем оценивания в условиях цифровой трансформации образовательной сферы//В сборнике: «Менеджмент XXI века: экономика, общество и образование в условиях новой нормальности». Сборник научных статей по материалам XX Международной научно-практической онлайн-конференции. Санкт-Петербург, 2022. С. 215-221.
13. П. В. Митракова, Д. Г. Пупышев, К. В. Сиразева, П. С. Гончарь. Эволюция показателя успеваемости студентов//Международный студенческий научный вестник. 2020. № 2. <https://eduherald.ru/ru/article/view?id=20141>.
14. А. В. Мороз, С. А. Мороз. Рейтинговая система оценивания успешности студентов в Китае//«Образование. Наука. Карьера». Сборник научных статей 2-й Международной научно-методической конференции. Т. 2. Курск, 2019.
15. М. Д. Напсо. Организационно-педагогические предпосылки внедрения инноваций в образовательный процесс//Инновации в образовании. № 11. 2022.
16. Д. Н. Песков. В России дипломы вузов могут получить «срок годности». 7.11.2020. <https://tass.ru/obschestvo/9938513>.
17. О. В. Польдин. Прогнозирование успеваемости в вузе по результатам ЕГЭ//Прикладная эконометрика. № 1 (21). 2011. С. 56-69. <https://core.ac.uk/download/pdf/6347785.pdf>.
18. Система оценивания знаний (Wikipedia). https://ru.wikipedia.org/wiki/Система_оценивания_знаний.
19. В. Н. Сосницкий, Н. И. Потанин, Л. В. Шевелева. Проблемы статистического анализа средней успеваемости студентов//Фундаментальные исследования. 2013. (10-2). С. 316-320.
20. А. В. Толмачев, Е. В. Сеницын, Г. В. Астратова. Вероятности распределения успеваемости студентов онлайн-курсов как инструмент оценки качества контрольных материалов//Интернет-журнал «Отходы и ресурсы». 2020. № 3. <https://resources.today/PDF/10INOR320.pdf>.
21. П. А. Тремзина, М. В. Фоминых. Отличительные особенности образования в Китае//Молодой ученый. 2016. № 5 (111.5). С. 92-93. <https://moluch.ru/archive/111/26621>.
22. Н. В. Хованов. Анализ и синтез показателей при информационном дефиците. СПб.: СПбГУ, 1996.
23. Н. В. Хованов. АСПИД — система квалиметрических методов оценивания в условиях дефицита информации качества сложных технических объектов//Методология и практика оценивания качества продукции. Л.: ЛДНТП, 1988. С. 56-61.
24. Н. В. Хованов. Математические модели риска и неопределенности. СПб.: СПбГУ, 1998.
25. А. Г. Чернявская. Андрагогика. 2018. https://studme.org/215165/pedagogika/kompetentnostnyy_podhod.
26. ECTS grading scale. https://en.wikipedia.org/wiki/ECTS_grading_scale.
27. GPA — средний балл в университет. https://www.unipage.net/ru/gpa_calculate.
28. Norm-referenced test. https://en.m.wikipedia.org/wiki/Norm-referenced_test.
29. K. Roell. What Is Grading on a Curve? <https://www.thoughtco.com/grading-on-a-curve-3212063>. July 22, 2019.
30. Short Guide to Understanding University Grades and Grading Systems in UK, America, and Europe. <https://www.mastersportal.com/articles/2291/8-university-grading-systems-around-the-world-that-may-or-may-not-be-weird.html>.
31. <https://mgimo.ru/study/akadrating/shkala.php>.

References:

1. T. I. Borovkova, T. D. Lavrinenko, J. Chun, Zhuziaou. Onating the quality of students' education in Russia and China//Modern problems of science and education. 2017. № 1. <https://science-education.ru/ru/article/view?id=26105>.
2. Universities approve the restriction of those receiving an increased scholarship. <https://tass.ru/obschestvo/6054148>.
3. N. S. Galimova, L. R. Zagitova. Construction of a multifactorial model of student performance//International Research Journal. № 6 (96). 2020. Part 3, June. P. 31-36. <https://research-journal.org/social/postroenie-mnogofaktornoj-modeli-uspevaemosti-studenta>.
4. European System translation and accumulation balls. https://ru.wikipedia.org/wiki/Европейская_система_перевода_и_накопления_баллов.
5. A. K. Erokhin, N. A. Glushenko. Innovations in education: the past and prospects of the future//Scientific and methodological electronic journal «Concept». 2017. № 53.
6. E. A. Erokhina, D. V. Khruslova, E. S. Klyshinsky, Yu. V. Zhurin. Analysis of student performance statistics as a means of improving the quality of education//Engineering education. 2014. № 154. http://aeer.ru/files/io/m15/art_28.pdf.
7. I. Zasursky. System of troechnikov budetsena. <https://aurora.network/articles/39-nauka-i-obrazovanie/56823-sistema-troechnikov-budet-snesena>.
8. P. V. Konyukhovskiy, A. O. Olkhovik, A. S. Alipov. Models of assessing the quality of higher education as a tool for managing the development of human capital//Actual problems of economics and management. 2017. № 4 (16). P. 88-94.
9. P. V. Konyukhovskiy. On methods of evaluating the results and effectiveness of educational activities in modern conditions//In the collection: «The State and business. Current trends and problems of economic developments». Materials of the XIII International Scientific and Practical Conference. In 3 parts. Saint Petersburg, 2021. P. 120-132.
10. P. V. Konyukhovskiy. Evaluation of educational activity results: methods and solutions // In the collection: «Management of the XXI century: socio-economic transformation in conditions of uncertainty». Collection of scientific articles based on the materials of the XVIII International Scientific and Practical Conference. Saint Petersburg, 2020. P. 209-218.
11. P.V. Konyukhovskiy, A.O. Olkhovik, A.S. Kuznetsova. Economic and mathematical modeling of the higher education market in the conditions of economic recession//State Administration. Electronic bulletin. 2016. № 58. P. 340-358.
12. P. V. Konyukhovskiy, A. S. Alipov. Problems of development and improvement of assessment systems in the conditions of digital transformation of the educational sphere//In the collection: «Management of the XXI century: economy, society and education in the conditions of a new normality». Collection of scientific articles based on the materials of the XX International Scientific and Practical Conference. St. Petersburg, 2022. P. 215-221.
13. P. V. Mitrakova, D. G. Pupychev, K. V. Sirazheva, P. S. Gonchar. Introduction to the course of lectures of students//International Student Scientific Bulletin. 2020. № 2. <https://eduherald.ru/ru/article/view?id=20141>.
14. A. V. Moroz, S. A. Moroz. Rating system for assessing the success of students in China//«Education. The science. Career». Collection of scientific articles of the 2nd International Scientific and Methodological Conference. Volume 2. Kursk, 2019.
15. M. D. Napso. Organizational and pedagogical prerequisites for the introduction of innovations in the educational process//Innovations in education. № 11. 2022.
16. D. N. Peskov. In Russia, university diplomas can receive an «expiration date». 7.11.2020. <https://tass.ru/obschestvo/9938513>.

17. O. V. Poldin. Predicting academic performance at the university based on the results of the Unified State Exam//Applied Econometrics. № 1 (21). 2011. P. 56-69. <https://core.ac.uk/download/pdf/6347785.pdf>.
18. Image recognition System (Wikipedia). https://ru.wikipedia.org/wiki/Knowledge_assessment_system.
19. V. N. Sosninsky, N. I. Potanin, L. V. Sheveleva. Problems of statistical analysis of students' average academic performance//International studies. 2013. (10-2). C. 316-320.
20. A. V. Tolmachev, E. V. Sinitsyn, G. V. Astratova. Probabilities of the distribution of the progress of students of online courses as a tool for assessing the quality of control materials//Online magazine «Waste and Resources». 2020. № 3. <https://resources.today/PDF/10INOR320.pdf>.
21. P. A. Tremzina, M. V. Fominykh. Distinctive features of education in China//Young scientist. 2016. № 5 (111.5). P. 92-93. <https://moluch.ru/archive/111/26621>.
22. N. V. Khovanov. Analysis and synthesis of indicators for information deficit. St. Petersburg: St. Petersburg State University, 1996.
23. N. V. Khovanov. ASPID — system of qualimetric assessment methods in conditions of lack of information on the quality of complex technical objects//Methodology and practice of product quality assessment. L.: LDNTP, 1988. P. 56-61.
24. N. V. Khovanov. Mathematical models of risk and uncertainty. St. Petersburg: St. Petersburg State University, 1998.
25. A. G. Chernyavskaya. Andragogy. 2018. https://studme.org/215165/pedagogika/kompetentnostnyy_podhod.
26. ECTS grading scale. https://en.wikipedia.org/wiki/ECTS_grading_scale.
27. Average score — average score at the university. https://www.unipage.net/ru/gpa_calculate.
28. Norm-referenced test. https://en.m.wikipedia.org/wiki/Norm-referenced_test.
29. K. Roell. What Is Grading on a Curve? <https://www.thoughtco.com/grading-on-a-curve-3212063>. July 22, 2019.
30. Short Guide to Understanding University Grades and Grading Systems in UK, America, and Europe. <https://www.mastersportal.com/articles/2291/8-university-grading-systems-around-the-world-that-may-or-may-not-be-weird.html>.
31. <https://mgimo.ru/study/akadrating/shkala.php>.