

**П. В. Конюховский**

Российский государственный педагогический  
университет им. А.И. Герцена,

**А. С. Алипов**

Санкт-Петербургский государственный университет,  
Россия, Санкт-Петербург

## **ПРОБЛЕМЫ РАЗВИТИЯ И СОВЕРШЕНСТВОВАНИЯ СИСТЕМ ОЦЕНИВАНИЯ В УСЛОВИЯХ ЦИФРОВОЙ ТРАНСФОРМАЦИИ ОБРАЗОВАТЕЛЬНОЙ СФЕРЫ**

*Аннотация.* В фокусе рассмотрения находится проблематика эволюции систем оценивания результатов деятельности образовательных учреждений. Цифровая трансформация экономики и общества оказывает радикальное воздействие на сферу высшего образования. Перед университетами возникают новые задачи, усиливаются процессы их конкуренции. Это в значительной мере повышает требования к системам, на основе которых оцениваются результаты и качество образовательной деятельности. Концептуальные модели и методы решения данных, основанные на байесовских подходах, предложены в данной статье.

*Ключевые слова:* оценивание образовательной деятельности, системы оценивания, балльно-рейтинговые системы, байесовские подходы.

**P. V. Konyukhovsky**

Herzen State Pedagogical University of Russia,

**A. S. Alipov**

St. Petersburg State University,  
Russia, Saint-Petersburg

## **PROBLEMS OF DEVELOPMENT AND IMPROVEMENT OF ASSESSMENT SYSTEMS IN THE CONDITIONS OF DIGITAL TRANSFORMATION OF THE EDUCATIONAL SPHERE**

*Annotation.* In the focus of consideration is the problems of the evolution of systems for assessing the results of the activities of educational institutions. The digital transformation of the economy and society is having a radical impact on the field of higher education. Universities face new challenges, and the processes of their competition are intensifying. This significantly increases the requirements for the systems on the basis of which the results and quality of educational activities are assessed. Conceptual models and methods for solving data based on Bayesian approaches are proposed in this article.

*Keywords:* assessment of educational activities, assessment systems, point-rating systems, Bayesian approaches.

Цифровая трансформация экономики и общества в неуклонно расширяющихся масштабах затрагивает сферу образования. Это накладывает дополнительные требования к результативности и эффективности работы образовательных учреждений, как в области высшего образования, так и в среднем, специальном, начальном и дополнительном образовании.

Как следствие, возникают задачи совершенствования методов, на базе которых строятся методики оценивания деятельности образовательных институтов. Целью настоящего исследования является выработка и анализ (на принципиальном уровне) методов, на основании которых может проходить модификация традиционных систем оценивания результатов учебного процесса, а также конструирование новых систем. В первую очередь речь идет о системах оценивания, применяемых в высших учебных заведениях.

Новые системы оценивания по сравнению с традиционными должны обладать радикально большей устойчивостью к имманентным недостаткам классических балльных и рейтинговых систем. Также новые оценочные системы должны быть конструктивно адаптируемыми к новым формам обучения, переход к которым форсировал коронавирусный кризис. Принципиальным свойством эволюции образования за последние десятилетия стало радикальное изменение форм информации и знаний, которых ожидают от учебных заведений как их выпускники, так и потенциальные работодатели этих выпускников. Соответственно, и новые оценочные системы должны адекватно реагировать на эти изменения.

Проблематика контроля качества учебного процесса и анализа достоинств и недостатков систем оценивания имеет чрезвычайно богатую историю. Говоря о современных знаковых исследованиях и публикациях, которые затрагивают вопросы модификации систем оценивания с точки зрения новых задач и вызовов образовательной сферы, следует упомянуть выступление спецпредставителя Президента России по вопросам цифрового и технологического развития Д. Пескова [4]. Он подверг достаточно жесткой критике практику выдачи дипломов «на всю оставшуюся жизнь» и выступил за переход к «дипломам по требованию», отражающим актуальный уровень знаний на момент предъявления.

Достаточно остро проблема инфляции оценок поднимается в [5]. В данной публикации отмечены объективные и устойчивые причины трендов, влияющих на необоснованное повышение оценок, равно как и сформулированы принципиальные направления борьбы с этими негативными тенденциями.

Анализируя актуальные проблемы систем оценивания, невозможно обойти стороной одно из фундаментальных понятий современной педагогической науки, а именно компетентностный подход. Необходимо признать, что он является одним из основополагающих в современном учебно-методическом дискурсе, и нельзя не согласиться с разумностью исходной (отправной) идеи компетентностного подхода. В соответствии с ней он трактуется как имплементация понятия «профессионализм» с учетом требований современных социально-экономических условий. Подразумеваются акценты не только на получение «чистых знаний», но и на развитие способностей человека по успешному применению этих знаний в реальности.

Однако на практике ситуация с формулировками «компетенций» достаточно быстро приобретает черты абсурда. За редким исключением, они сводятся к трудно проверяемым декларациям. И это вполне естественно. Крайне затруднительно придумать содержательное описание какого-либо профессионального умения или навыка, которое одновременно было бы и конкретным и не носило бы характера частного случая. По всей видимости, в обозримом будущем придет осознание, что история компетенций стала еще одним наглядным подтверждением того, как изначально благие идеи при отсутствии объективных условий для их реализации превращаются в бюрократический симулякр. Еще одним врожденным пороком т.н. компетенций является невозможность их использования (за редким исключением) для конструктивного и транспарентного построения критериев оценивания.

В настоящее время существующие системы оценивания сталкиваются с целым комплексом проблем, причем как на уровне отдельных учебных дисциплин и контрольно-тестовых мероприятий, так и на уровне результатов освоения учебных программ в целом.

Настоящая статья фокусирует внимание на проблеме соотношения (согласования) абсолютных и относительных (пропорциональных) систем оценивания.

Уточним и конкретизируем условия (предпосылки) рассматриваемой модели. Предполагается проведение некоторого контрольно-тестового (экзаменационного) мероприятия, результаты которого выражаются в некоторой внутренней шкале. Достаточно типичной является 100-балльная шкала. Именно она, например, предлагается в LMS Moodle. Однако очевидно, что это далеко не единственный и даже не преобладающий вариант.

Особо следует отметить чрезвычайно широкое межстрановое разнообразие систем оценивания знаний, см., например, [1]).

Более того, текущая ситуация такова, что отдельные вузы варьируются по своим требованиям к пороговым значениям абсолютных систем оценивания. В качестве примера можно привести шкалу соответствия оценок, опубликованную на странице «Академический рейтинг» МГИМО [2].

Для общности изложения мы будем предполагать проведение нормировки, то есть в последующих примерах результаты тестового мероприятия будут значениями на интервале.

Под абсолютной оценкой понимается оценка, выставляемая в соответствии с изначально заданными пороговыми значениями под-интервалов на полном отрезке. Например,  $\geq 0.9$  (90 %) — отлично или «А»,  $\geq 0.8$  (80 %) — очень хорошо или «В»,  $\geq 0.7$  (70 %) — хорошо или «С» и т. д.

Принципиальной альтернативой (по отношению к абсолютным оценкам) является относительная (пропорциональная) система оценивания. Она предполагает выставление оценки отдельному учащемуся на основе положения его результатов в общей группе (выборке) проверяемых.

В англоязычной литературе за подобными методиками закрепился термин *norm-referenced test* (NRT).

«Классическим» примером пропорционального подхода является «старый метод» конвертирования оценок ECTS (European Credit Transfer and Accumulation System — Европейская система перевода и накопления баллов). Данный метод был актуален до 2009 г. В соответствии с ним предлагалось выставлять оценку «А» студенту, входящему в лучшие 10 % (по дисциплине или оцениваемому виду учебной деятельности), «В» — студенту, входящему в следующие 25 %, «С» — студенту, входящему в следующие 30 % и т. д.

Как несложно заметить, ни один из перечисленных подходов с точки зрения достоинств и недостатков не доминирует другой.

В частности, преимуществом абсолютной системы оценивания является то, что она задает транспарентные и внятные правила для конкретных контрольно-тестовых мероприятий.

Вместе с тем она достаточно критична к ошибкам и просчетам в контенте методического обеспечения тестовых и экзаменационных процедур. Строго формальное следование жестким «пороговым правилам» может приводить как к неадекватно большим долям студентов с высокими оценками, так и к противоположным ситуациям.

Пропорциональные системы гарантируют рациональность процедур оценивания с точки зрения их классификационных свойств, предотвращают перекосы в долях «отличников» и «отстающих». Однако их имманентным недостатком является чувствительность

к вариациям расчетной базы. В зависимости от специфики выборок, относительно которых происходит расчет пропорциональных оценок, студенты из разных групп могут оказаться качественно несопоставимыми по уровням своих знаний и квалификаций.

Указанные обстоятельства определяют важность и актуальность проблемы согласования абсолютных и пропорциональных систем оценивания. В рамках настоящего исследования рассматривается возможность применения методов решения данной задачи, основанных на байесовских подходах (или на так называемом байесовском приближении).

Принципиальная модель основывается на предпосылке о возможности объективного представления критериального признака оцениваемой группы (выборки) студентов с помощью некоторой функции плотности, принимающей значения на интервале (то есть полном множестве значений критериального признака). Термин «критериальный признак» в данном случае является обобщающим для возможных результатов учебной работы, т. е. уровня знаний, умений, знаний, способностей, квалификации и т. п.

Простейшим примером такой функции плотности может служить асимметричное треугольное распределение на интервале. Также достаточно удобным, конструктивным и технологичным с точки зрения задач исследования оказывается распределение PERT (распределение, получаемое на основе бета-распределения за счет введения дополнительных предположений о значении ожидаемого значения).

При условии, что рассматривается интервал, и асимметричное треугольное распределение, и PERT однозначно определяются значением моды.

Таким образом, требования относительно доли студентов, получающих высшую оценку («А») (например), и абсолютного порогового значения (например,  $\geq 0.9$  (90 %) правильных ответов) могут быть однозначно сопоставлены со значением моды соответствующего распределения PERT. В дальнейшем будем использовать для нее это обозначение.

При этом роль абсолютных значений порогов балльных оценок выполняют квантили распределения PERT (80, 70 %) и т. д.).

Преимуществом данного подхода является то, что мы получаем простые и транспарентные процедуры. Имея априорное «эталонное» распределение, можно определить, в каких пропорциях должны распределиться студенты при заданных абсолютных балльных порогах и, наоборот, как следует установить балльные пороги для того, чтобы получить желаемое распределение успеваемости.

Практическое применение данной модели предполагает выработку процедур последовательной корректировки значений моды распределения PERT на основе логики байесовского приближения.

В соответствии с ней значение моды также рассматривается в качестве случайной величины, вид функции распределения которой последовательно уточняется на основе апостериорной информации, получаемой после очередной процедуры контроля (тестирования):

$$P(\mu_{t+1}|G_t) = \frac{P(G_t|\mu=\mu_t) \cdot P(\mu=\mu_t)}{\int_{\mu} P(G_t|\mu) \cdot P(\mu) d\mu}, \quad (1)$$

где

$t$  — индекс тестового мероприятия (номер шага байесовского приближения);

$P(G_t|\mu = \mu_t)$  — вероятность (оценка вероятности) получения доли отличников (студентов, получивших оценку «А») составит  $G_t$  при условии, что мода равна  $\mu_t$  — в данном случае ориентируемся на значения моды, найденное на предшествующем шаге (на первом шаге, как уже отмечалось ранее, за основу берется исходное «эталонное» распределение);

$P(\mu = \mu_t)$  — безусловная вероятность того, что значение моды равно  $\mu_t$  (с учетом априорных представлений на шаге  $t$ );

$\int_{\mu} P(G_t|\mu) \cdot P(\mu) d\mu = P(G_t)$  — безусловная вероятность того, что доля отличников (студентов, получивших оценку «А») составит  $G_t$  (при всех возможных значениях моды  $\mu$ ).

В соответствии со схемой байесовского подхода апостериорное распределение очередного шага ( $t + 1$ ) становится априорным для шага ( $t + 2$ ) и т. д. Таким образом мы получаем методический алгоритм, предполагающий перманентную корректировку порогов абсолютных балльных оценок. При этом данная корректировка осуществляется целенаправленно с учетом установок на поддержание желательной («объективно рациональной») структуры успеваемости студенческих групп.

Предварительная апробация данных методов, проведенная в форме условного аналитического мониторинга результатов промежуточной аттестации студентов экономического факультета Санкт-Петербургского государственного университета и Российского государственного педагогического университета им. А. И. Герцена, позволяет дать им в целом позитивную характеристику.

В то же время следует признать, что практическая реализация байесовских подходов применительно к задачам имплементации систем оценивания связана с решением достаточно серьезных вычислительных проблем. Действительно, для выражений (1) в случае распределений класса PERT невозможно получить компактные ана-

литические формулы. Конструктивным и технологичным выходом является проведение расчетов с использованием алгоритмов Монте-Карло по схеме марковских цепей (МСМС) [3]. Именно в этом направлении предполагается последующее развитие данного исследования.

#### ЛИТЕРАТУРА

1. Системы оценивания знаний // [https://ru.wikipedia.org/wiki/Система\\_оценивания\\_знаний](https://ru.wikipedia.org/wiki/Система_оценивания_знаний).
2. <https://mgimo.ru/study/akadrating/>
3. *Geyer C. J.* (1991) Markov chain Monte Carlo maximum likelihood. Computing Science and Statistics: Proc. 23rd Symp. Interface, 156–163. <http://hdl.handle.net/11299/58440>.
4. *Батманова А.* Спецпредставитель президента раскритиковал выдачу дипломов «на всю жизнь». <https://www.rbc.ru/society/08/11/2020/5fa741c99a79472bc909181a>.
5. В Вышке будут приняты меры по предотвращению инфляции оценок. <https://www.hse.ru/news/edu/478612561.html>
6. <https://pltf.ru/2017/01/18/sistema-trochnikov-budet-snesena/>