

ПРАКТИКА МОДЕРНИЗАЦИИ

А.В. БЕЛОЦЕРКОВСКИЙ,
*профессор, ректор
Тверской государственной
университет*

К вопросу о рейтингах и рангах

Рассмотрена проблема ранжирования вузов с учетом международного опыта, научной обоснованности используемых показателей, высказаны соображения о возможностях и рисках при принятии конкретных решений на основании известных рейтингов.

Ключевые слова: рейтинги вузов, ранжирование, ранкинг, оценка качества, информативность показателей

В последние двадцать лет во всем мире значительно возрос интерес к ранжированию вузов, построению глобальных, региональных и национальных рейтинг-листов. В нашей стране этот интерес также растет. Дополнительное внимание к данной теме привлекло проведение мониторинга эффективности вузов и реализация программы по вхождению пяти ведущих российских вузов в «топ-100» по одному из глобальных рейтингов. Мониторинг эффективности, строго говоря, не является способом рейтингования вузов для определения лучших, а напротив, направлен на выявление «проблемных зон», вузов, по тем или иным причинам не преодолевающих пороговые значения сразу по нескольким показателям. Тем не менее показатели «проблемных зон», используемые при мониторинге эффективности, учитываются и при расчете так называемого потенциала вуза (при проведении конкурса КЦП), т.е. фактически для ранжирования всех, в том числе и лучших, стоящих «в очереди» за контрольными цифрами приема. Глобальные рейтинги предназначены для определения лучших из лучших, ранжирования «сливок», не дифференцируя или приписывая одинаковый нулевой рейтинг всем остальным.

Растет важность рейтингования вузов и в классическом смысле, как маркетингового инструмента, используемого в рыночных условиях, т.е. в процессе принятия решений о поступлении в тот или иной вуз, взаимодействии с ним или распределении ограниченных ресурсов. В связи с этим представляется полезным еще раз взглянуть на проблему ранжирования с учетом международного опыта, научной обоснованности выбранных показателей, и высказать соображения о возможностях и рисках принятия конкретных решений на основании известных рейтингов.

Составление рейтинга вузов – это, по сути, попытка сравнительного измерения качества реализуемых в них программ высшего образования и научных исследований¹. В отличие от других задач квалиметрии, здесь отсутствует абсолютная шкала критериев или норм некоего «минимального качества». Во многом это происходит потому, что разработчиками рейтингов во всем мире выступают коммерческие структуры, как правило, средства массовой информации или информационные агентства, которые к высшему образованию прямого отношения не имеют, но хотят создать интересный для массового потребителя продукт.

¹ Рейтинг (англ. *rating*) — числовой или порядковый показатель, отображающий важность или значимость определенного объекта или явления. Список объектов или явлений, имеющих наибольший рейтинг, обычно называют «Топ N», где N – количество объектов в списке, обычно кратное 10 (Википедия).

Почему растет интерес к рейтингам вузов?

Первое обстоятельство связано с массовизацией высшего образования, с огромным ростом спроса на высшее образование во всем мире [1]. Высшее образование, по крайней мере, его первая ступень, перестало выполнять функцию социального лифта, а стало необходимым условием выхода на современный рынок труда. Рост спроса привел к росту предложения, к конкуренции, в том числе международной, между различными образовательными программами и университетами. Одним из способов ориентирования в мире предложения и являются рейтинги. Что-то вроде рейтингов потребительских товаров и услуг журнала "Consumer Reports".

Кому в первую очередь нужны рейтинги вузов?

Тем, кто принимает решение о выборе учебного заведения. Сегодня этот вопрос может возникать в жизни каждого человека намного чаще, чем еще несколько десятков лет назад, т.к. высшее образование стало уровневым. При этом возможно обучение по программе бакалавриата в одном университете, в магистратуре в другом (и, возможно по иному направлению), в аспирантуре – в третьем. И каждый раз нужно принимать решение. Кроме этого, с беспрецедентно быстрым научно-техническим прогрессом и трансформацией поля профессий значительно растет потребность в дополнительном образовании, повышении квалификации и переподготовке. Вместо «образования на всю жизнь» – «образование в течение всей жизни». Рейтинги могут давать ориентир для более информированного выбора. Кроме того, ранжирование необходимо каждому университету, чтобы иметь внешнюю систему координат для оценки динамики своего развития. Можно также ожидать, что широко признаваемый рейтинг будет содействовать конкуренции между высшими учебными заведениями, что, в свою очередь, будет способствовать

повышению качества их программ. Работодателей также можно считать заинтересованной стороной в ранжировании. Бизнес может интересоваться как уровнем и качеством образования у принимаемого на работу сотрудника, так и относительным исследовательским и инновационным потенциалом образовательной организации для возможного взаимодействия с ним в области прикладных исследований и разработок. Наконец, различные фонды развития и поддержки также могут использовать подобные данные при определении победителей в различных конкурсных процедурах. Растущая популярность рейтингов сопровождается параллельно растущей критикой как самих показателей, используемых в рейтингах, так и способов их измерения.

Что общего в различных системах ранжирования?

Если взглянуть на многочисленные системы ранжирования можно отметить следующие их общие характеристики.

- Ранжирование основано на сборе информации по нескольким показателям. Иногда это количественное выражение входных характеристик, или ресурсов (человеческих, физических или денежных) образовательного процесса, иногда выходных, т.е. результатов образовательных программ или научных исследований.
- Предполагается, что показатели отражают некую компоненту качества, и чем выше его значение, тем выше соответствующая проекция качества. Это предположение тесно связывает вопрос выбора показателей с дискуссией о том, что же является качеством высшего образования.
- Значения показателей берутся из определённых источников информации с различными типами данных – от экспертных оценок до результатов официальной статистики.
- Показатели суммируются с весами для представления агрегированного результата – рейтинг-листов. Ранжироваться при этом могут как целые образовательные уч-

реждения, так и отдельные образовательные программы. Существуют многомерные рейтинги, или как их иногда называют в литературе, рэнкинги, в которых результаты по разным показателям представляются отдельно в виде баз данных, без попыток их взвешивания и агрегирования.

Оценка качества

Тот факт, что рейтинг-листы проранжированы от лучших к худшим, указывает на некое понимание того, что хорошо, а что не очень. Разработчики рейтинговых систем часто явно утверждают, что они измеряют качество. При этом, как известно, разные заинтересованные стороны под качеством подразумевают совершенно разные вещи. Поэтому и набор показателей, по которым строится тот или иной рейтинг, отражает интерес определенной заинтересованной стороны и ее представления о качестве. Количество потенциальных показателей, описывающих образовательное или научно-исследовательское учреждение, практически безгранично, тем не менее в большинстве международных систем фигурирует лишь их небольшое множество. Чем больше показателей, тем меньшим оказывается относительный вклад каждого из них в итоговый агрегированный результат.

Практически все системы ранжирования в странах с многолетним опытом построения национальных рейтингов, содержат следующие категории показателей.

Качество поступающих абитуриентов (стартовые характеристики). Наиболее распространенными являются результаты общенациональных тестов, аналогов нашего ЕГЭ (в США это SAT и ACT), как правило, выражаемых в процентилях (например, принятые на первый курс студенты принадлежат к группе лучших 10 процентов, или 20 процентов, или 50 процентов из всех, сдававших общенациональный экзамен с единой шкалой)

Ресурсное обеспечение образовательного процесса или научных исследований (входные характеристики). Здесь оцени-

ваются ресурсы, которыми располагает университет или отдельная образовательная программа в широком смысле. Они описываются такими показателями, как количество преподавателей на одного студента, процент оспеенности преподавателей. Могут учитываться финансовые или физические ресурсы, такие как расходы на одного студента или возможности доступа студентов к компьютерным и библиотечным ресурсам.

Результаты обучения или научных исследований (выходные характеристики). Здесь наиболее распространенными являются показатели трудоустройства выпускников, занимаемые ими позиции, индексы цитирования и доходы. Иногда включаются такие показатели, как удовлетворенность своей работой и своим образованием.

Академическая репутация. Некоторые рейтинги включают показатели, которые явно или неявно характеризуют репутацию или престиж рассматриваемого университета или образовательной программы. Это могут быть результаты опросов студентов, работодателей или академического сообщества. Некоторые рейтинги включают показатели об успехах вуза или его отдельных программ в международных рейтингах, а также объемы привлеченных исследовательских грантов.

Процессный подход. Многие эксперты отмечают, что весьма важным является то, что напрямую не описывается ни ресурсами, ни результатами, а лежит где-то между ними, а именно, организация образовательного процесса с точки зрения качества обучения, используемых образовательных технологий, характера взаимодействия студентов с преподавателями, содержания образовательных программ, возможностей организации прохождения практики.

Добавленная стоимость. Наиболее объективной, но трудно измеряемой характеристикой качества обучения является добавленная стоимость. Стартовые харак-

теристики (или “качество приема”), обычно интерпретируются как показатель привлекательности (популярности) университета или образовательной программы. С этой точки зрения необходимо посмотреть на разницу между стартовыми и финальными характеристиками одних и тех же студентов. Другими словами, оценить, насколько конкретная образовательная программа повысила потенциал студента вне зависимости от его стартовых результатов. Связанным с этим оказывается и вопрос эффективности, т.е. оценка уровня достигнутой «добавочной стоимости» студента на единицу вложений.

Условия жизни студентов. Это фактор, не имеющий непосредственного отношения к образовательной программе или университету, поэтому он используется значительно реже. Сюда может относиться обеспеченность и доступность жилья, медицинское обслуживание, возможности для занятий спортом, культурного развития и досуга, что часто в значительной степени относится к населенному пункту расположения университета.

Из вышеперечисленных в международной практике чаще всего используются стартовые характеристики, характеристики ресурсов, результаты и репутационные характеристики.

Следует также различать рейтингование и оценку *KPI* (*Key Performance Indicators* – основных показателей деятельности). Рейтингами занимаются, как правило, коммерческие организации, интересующиеся скорее ресурсами и репутацией. Оценка *KPI* проводится, как правило, руководящими и государственными структурами, следящими за эффективностью вложений, и сосредоточена она главным образом на результатах деятельности организации в целом.

Достоинства и недостатки рейтингования

Кроме непрекращающейся дискуссии об информативности и точности измерения тех или иных показателей, правомерности

их агрегирования и общего скептицизма (как правило тех, кто занимает нижние строки в рейтинг-листах), существуют общие взгляды на достоинства и недостатки рейтингов, или аргументы “за” и “против”.

Аргументы “за”:

- рейтинги упрощают и проясняют сложную картину высшего образования для потенциальных студентов и заинтересованных сторон вне образовательной отрасли;
- рейтинги обеспечивают университетам рекламу и способствуют их узнаваемости;
- рейтинги дают стимулы для повышения качества обучения и результативности научных исследований;
- рейтинги улучшают качество сбора данных в области высшего образования;
- рейтинги дают многогранную картину высшего образования.

Аргументы “против”:

- рейтинги вводят в заблуждение по поводу качества образования. (Часто набор показателей продиктован не их связью с качеством образования, а доступностью информации);
- рейтинги дают искаженную картину дифференцирования качества. (Если в предыдущем пункте речь шла о наборе показателей, то здесь критикуются методы их объединения. Это связано с выбором весовых коэффициентов суммирования признаков, отсутствием робастных статистических процедур и пренебрежением статистической значимостью результатов);
- рейтинги дают искаженную картину об изменениях во времени. (В силу того что наборы показателей и методы их обработки меняются, данные за разные периоды времени становятся несопоставимыми и непонятно, чем вызваны те или иные изменения позиции в рейтинг-листах – деятельностью университета или корректировкой в методах расчетов);
- рейтинги подталкивают к неадекватному поведению со стороны студентов и организаций. (Если приведенная выше кри-

тика в отношении способности рейтингов измерять истинное качество справедлива, то последствия могут быть катастрофическими как для студентов, так и для университетов. Особенно это касается универсальных рейтингов «один для всех», не учитывающих разницу в типах образовательных учреждений).

Дебаты по поводу рейтингов не стихнут до тех пор, пока будут продолжаться дискуссии о том, что же является «качеством» в образовании [2]. Стоит только отметить, что адвокаты и противники обычно говорят о разных вещах: защитники – о полезности самой процедуры рейтингования, а противники – о недостатках используемых методов.

Теоретические соображения

Очевидно, что разработать идеальную или устраивающую всех методологию ранжирования невозможно. Существуют более-менее авторитетные рейтинги, они, как правило, многолетние. Существуют и явно неудачные примеры. Поскольку эта задача не может быть решена строго математически, всегда будут справедливы возражения по поводу релевантности того или иного рейтинга. И все-таки, некоторые общетеоретические соображения о том, как должен строиться рейтинг, или, если хотите, как он не должен строиться, возьмусь высказать.

Все, кому когда-либо приходилось заниматься задачами квалиметрии, распознавания образов, автоматической классификации или строить интегральные показатели сложных процессов, знают, что все эти задачи имеют базовую научную подоплеку. Поясню, не вдаваясь в подробности. В задачах оценки параметров или свойств генеральной совокупности по данным выборки, составляющим суть математической статистики, есть обязательные требования к тому, какими должны быть выборки, чтобы можно было с математически предсказуемой погрешностью распространять полученные на них выводы на всю генераль-

ную совокупность. Несоблюдение этих требований может приводить к искажению или несостоятельности выводов [3]. Часто подобный казус возникает, например, при проведении социологических опросов, когда выбор респондентов или не является случайным, или не является совершенно независимым от результатов предыдущих выборок. И хотя социологи, опираясь на разработанные математические методики, указывают статистическую погрешность данных опросов, результаты последующей валидации, например путем проведения выборов, оказываются далеко за пределами коридоров ошибок предсказываемых результатов. Это следствие нарушения простого базового предписания, в предположении выполнения которого и получены все методики дальнейших расчетов.

Точно таким же базовым требованием при построении интегральных показателей или рейтингов, является решение трех стандартных задач: 1) отбор информативных показателей, что подразумевает оценку информативности всех потенциально доступных, 2) отбор измеряемых показателей с не превышающей заданный порог относительной погрешностью измерений, 3) обоснование выбора весовых коэффициентов вхождения в итоговый агрегат. Если тот или иной показатель не является информативным, его не включают в набор, даже при невысокой погрешности его измерения. Показатель не включают в набор и в том случае если он информативен, но измеряется с высокой относительной погрешностью. Выбор весовых коэффициентов – вопрос наиболее деликатный, так как их в принципе можно настроить на любой наперед заданный результат. В случае отсутствия какой-либо априорной информации наилучшим является или рандомизированный подход, или нормализация показателей путем нормирования отклонения от среднего с последующим суммированием с одинаковыми коэффициентами.

Всегда ли соблюдаются эти принципы составителями рейтингов? К сожалению,

далеко не всегда. Попробую продемонстрировать это на примере набора показателей, которые используются для оценки потенциала вузов и ранжирования по этому потенциалу в конкурсе на получение КЦП. С некоторым приближением этот алгоритм ранжирования можно рассматривать как своего рода процедуру рейтингования вузов, участвующих в конкурсе. Большая часть исходных показателей совпадает с показателями эффективности вузов, используемыми при мониторинге, что вполне понятно, т.к. позволяет сэкономить время и средства на сбор данных. Повторюсь, показатели мониторинга эффективности направлены на диагностику проблемных зон с последующим более внимательным взглядом на выбор курса «лечения». Это скорее *KPI*, и призваны они решать совершенно иную задачу. В задаче ранжирования в конкурсе на КЦП невооруженным взглядом видны некоторые недостатки выбора показателей. Рассмотрим некоторые из них.

Средний балл ЕГЭ – показатель хороший, в смысле, информативный и, казалось бы, измеряемый с высокой точностью. Его аналог используют многие уважаемые зарубежные рейтинги как показатель качества приема, он действительно демонстрирует популярность вуза или образовательной программы. Только всюду он нормируется или нормализуется. Ведь средний балл ЕГЭ в стране по математике почти в полтора раза ниже, чем средний балл по обществоведению; суммировать и осреднять их абсолютные значения без нормирования нельзя. В результате, при сохранении информативности, вносится искусственная погрешность измерения, связанная с недостаточной нормализацией показателя.

Публикационная активность, индексируемая международными базами (или РИНЦ), – очень хороший показатель, информативный и с высокой точностью измерения. Аналог используется всеми международными рейтингами как основной показатель научной деятельности или квалификации профессорско-преподавательского

состава. Только все-таки важно не «количество», а «качество» науки. Важно не столько количество публикаций, сколько их качество, а это – индексы цитирования с учетом импакт-факторов источников (вычищенные от самоцитирования и кругового цитирования), индекс Хирша. В международных глобальных рейтингах лучших используют базы данных высокоцитируемых работ *High citation index*, или публикации в эксклюзивных журналах типа *Nature*. Для ранжирования российских вузов эти показатели пока мало полезны, поскольку у большинства по ним нулевое значение, т.е. низкая информативность. РИНЦ в этом смысле пока существенно более информативен.

Объемы грантов и доходов от науки в мире крайне редко используются в качестве показателей рейтингов. Во-первых, считается, что они должны коррелировать с публикационной активностью, а раз так, не будут вносить дополнительной информации. Во-вторых, считается, что они характеризуют скорее уровень менеджмента вуза или его учредителей, чем качество преподавателей (в отличие от индексов цитирования). Тем не менее, по общим соображениям, это показатель информативный, хотя и в меньшей степени.

Общие доходы вуза – во многом следствии первых трех (высокая популярность, высокая научная активность, высокий объем НИР и ОКР), поскольку часто являются результатом побед в различных конкурсах поддержки программ развития, где учитывались те же первые три показателя. С точки зрения теории информации, зависимый параметр никакой новой информации не добавляет, его информативность невелика.

Самый спорный показатель – это *площадь помещений вуза на душу студента*. Не говоря о том, что он не используется ни одним известным международным или национальным рейтингом (и не зря), он никаким образом не характеризует ни качество образования, ни качество студентов, ни ка-

чество преподавателей. Даже теоретически, с точки зрения стратегического менеджмента: что было бы правильнее – вкладывать деньги в дополнительные площади или закупать современное оборудование и привлекать известных ученых? Ответ очевиден, первое не приведет никуда, второе может продвинуть в международных рейтингах. Ведь есть СНИП, который указывает нижнюю разрешенную санитарную границу, и все. Если же посчитать корреляцию между показателем площадей и, например, средним баллом ЕГЭ по данным мониторинга эффективности вузов, она получится статистически значимо отрицательной! В среднем, чем выше популярность вуза (измеряемая средним баллом ЕГЭ), тем больше в нем студентов на имеющихся площадях, т.е. ниже показатель площадей, приходящиеся на одного студента. Таким образом, часто ранжирование по площадям расставляет вузы в порядке, обратном качеству и репутации. Иначе говоря, это показатель антикачества. Самое же поразительное, что этот показатель входит с тем же весом, что и другие.

Характеристика *востребованности на рынке труда*, безусловно, важна. Вот только измерить ее практически невозможно. В западных рейтингах это всегда результат экспертного опроса. У нас пошли по пути объективного измерения по числу обращений на биржу труда. Но здесь информативность практически нулевая, т.к. число обратившихся на биржу труда никак не связано с качеством их подготовки, а скорее с качеством работы биржи, количеством предлагаемых ею программ, текущей ситуацией с региональным рынком труда и т.д. Ситуация, которую можно проиллюстрировать примером. Допустим, необходимо оперативно определить степень опьянения большого количества людей. Это может быть очень важно и общественно востребовано. Для этого мы решили измерять температуру тела каждого из них. Почему температуру? А потому, что ее легко измерить. Сможем ли мы при этом выделить пьяных? Вряд ли, если они еще живы и тела не ос-

тыли. Так и с трудоустройством. Для получения этой важнейшей характеристики используем температуру, которую легко измерить. Это, кстати, демонстрируют данные проведенного мониторинга эффективности: показатель трудоустройства меняется очень незначительно от вуза к вузу, т.е. обладает очень низкой относительной дисперсией. Любой специалист вам скажет, что это первый признак низкой информативности.

Доля иностранных студентов. Казалось бы, информативный показатель, если это, например, студенты из ЕС или стран ОЭСР. Только здесь дифференциации по странам не проводится, и вузы, привлечшие на свои образовательные программы, скажем, сотню студентов из Великобритании, сотню из Вьетнама, или сотню из Таджикистана, будут иметь одинаковый показатель. К качеству это никакого отношения не имеет.

Рекомендации

Мы все знаем, что если начать характеризовать какую-либо содержательную деятельность набором формальных показателей, очень скоро о содержательной деятельности будет забыто, а все начнут заниматься «накачкой» показателей. Плановая экономика нам это не раз продемонстрировала. С этой точки зрения, никакой рейтинг не должен восприниматься очень серьезно и уж точно недопустимо использовать его в качестве основы для систематического принятия каких-то административных или финансовых решений. Чтобы, тем не менее, использовать полезную составляющую, на мой взгляд, следует придерживаться нескольких правил как при составлении рейтингов, так и их использовании.

1. Внимательно оценивать информативность и точность измерения показателей и их релевантность с точки зрения представления и дифференциации качества.

2. Избегать высокой доли экспертной и репутационной составляющей, так как при этом решающим становится вопрос от-

бора экспертов и процедур их опроса. В нашей стране репрезентативность таких методик пока невысока. Неопределенность методик вносит в эту составляющую высокую погрешность.

3. Предпочитать максимальную конкретизацию и сравнимость рейтингуемых объектов. Рейтинг образовательных программ куда более репрезентативен, чем рейтинг разнородных и разнообразных вузов; так же как рейтинг научных лабораторий куда более релевантен, чем рейтинг больших исследовательских институтов. Рейтинг образовательных программ, дифференцированный по уровням образования, еще лучше.

4. Учитывать, что разные уровни образования могут требовать разного учета вклада одних и тех же показателей. Качество профессорско-преподавательского состава, характеристики качества образовательного процесса важнее на бакалаврском уровне, характеристики научной и публикационной активности репрезентативнее для уровней магистратуры и аспирантуры.

5. Наилучшим является сохранение многомерности описания такого многомерного понятия, как качество. С этой точки зрения, предпочтительнее ранжирование, или многомерное ранжирование, т.е. сохранение полной неагрегированной базы данных всех показателей. Это позволяет каждому самостоятельно выстраивать свои собственные рейтинги в соответствии с приоритетами. Большим достоинством проведенного мониторинга эффективности вузов явилось сохранение и представление всего массива данных в открытой форме. Здесь действительно можно получить достаточно полную картину.

Есть ли положительные примеры составления национальных рейтингов у нас? На мой взгляд, наименьшее количество претензий можно предъявить к национальному рейтингу вузов, уже в течение ряда лет проводимого Интерфаксом (<http://unigrating.ru>). С точки зрения как используемой и приведенной методологии, так и

многомерности подхода, он очень близко стоит к лучшим национальным рейтингам передовых стран. Единственный его недостаток – достаточно высокий процент репутационной или экспертной составляющей. Что же касается использования измеряемых величин, они достаточно корректно включены в итоговый рейтинг. Одним из наиболее уважаемых международных рейтингов является ARWU (составляемый Шанхайским университетом Джао Тонг), поскольку он базируется на информативных измеряемых величинах и полном отсутствии репутационной составляющей.

Каждый из нас, так или иначе профессионально занимающийся проблемами высшего образования, имеет собственные представления о качестве в этой области и о результатах своей деятельности, в том числе и по отзывам коллег. Такая «peer reviewed» оценка. Очень важно, чтобы в этой собственной системе координат мы постоянно повышали качество того, что от нас зависит. Понемногу, но каждый день. Тогда, уверен, и позиции в рейтингах начнут постепенно повышаться. Для этого нам рейтинги и нужны: при ежедневной внутренней самооценке хотя бы раз в год необходима внешняя оценка. Содержательная деятельность каждый день, формальные показатели – раз в год для нас самих. Не наоборот. По признаваемой нами шкале, с позитивной динамикой. Уверен, тогда мы скажем: рейтинги – это инструмент развития.

Литература

1. Белоцерковский А.В. Российское высшее образование: о вызовах и рисках // Высшее образование в России. 2012. № 7. С.3–9.
2. Белоцерковский А.В. К вопросу о «количестве качества» и «качестве количества» в системе образования // Высшее образование в России. 2013. № 7. С.22–29.
3. Малошонок Н.Г., Девятко И.Ф. Эксперимент как метод изучения эффективности практик и нововведений в высшем образовании // Высшее образование в России. 2013. № 10. С.141–151.