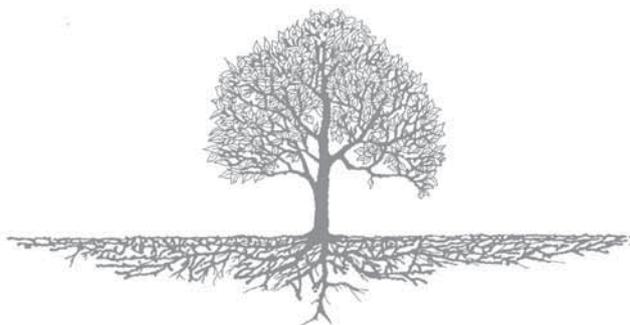


УПРАВЛЕНИЕ РЕПУТАЦИЕЙ ВУЗА



А. В. Кинчарова

МЕТОДОЛОГИЯ МИРОВЫХ РЕЙТИНГОВ УНИВЕРСИТЕТОВ: АНАЛИЗ И КРИТИКА

Статья содержит обзор работ, посвященных анализу и критике методологии крупнейших мировых рейтингов университетов, представленных в англоязычной научной литературе. Он предваряется кратким описанием устройства трех крупнейших рейтингов: ARWU, QS, THE. Основные направления критики: определение набора показателей рейтинговой системы; проблемы использования основных видов показателей: библиометрических и peer review; проблемы взвешивания показателей и агрегирования их в индекс. В результате анализа ряд исследователей делает вывод об ограниченных возможностях использования результатов рейтинговых измерений.

Ключевые слова: рейтинги университетов; методология рейтингов; показатели рейтинговой системы; библиометрические показатели; показатели peer review; агрегирование показателей.



A. V. Kincharova

Methodology of international university rankings: analysis and criticism

The article contains a review of papers which analyse and criticise the methodology of the main world university rankings presented in the English-language literature. It is preceded by a brief description of structure of the three largest rankings: ARWU, QS, THE. The main directions of criticism are: specifics of the set of rating system indicators; problems of using the main types of indicators: the bibliometric ones and the peer review ones; problems of weighting indicators and their aggregation into one index. A number of researchers conclude that the opportunities of using the results of university rankings are quite limited.

Keywords: university rankings; ranking methodology; indicators of ranking system; bibliometric indicators; peer review indicators; aggregation of indicators.

Бурный рост применения рейтингов и возрастание их влияния в сфере образования вызвали волну дискуссий, большая часть которых связана с вопросами методологии построения рейтингов и их методологической критики. Ученые и руководители в сфере образования проявили большой интерес к тому, как рейтинги составлены, поскольку очевидно, что методика измерения оказывает влияние на результат, а он, будучи доступным всем заинтересованным субъектам, в свою очередь, может оказывать влияние на благополучие университетов, в частности, влияет на распределение ресурсов, качество

абитуриентов и т. д. Также значительное внимание к методическому аспекту построения рейтингов проявили специалисты по методологии измерений.

Как оказывается, методология построения мировых рейтингов университетов весьма подвержена критике, и критика появилась в изоляции, вероятно, отчасти из-за значимости явления вследствие масштабов его возможного влияния, отчасти из-за реальных слабостей методологии рейтингов.

В данной работе представлен обзор критики методологии построения крупнейших миро-

вых рейтингов университетов, существующей в англоязычной научной литературе. Он предвзвешивается общим описанием устройства трех наиболее значимых мировых рейтингов университетов.

«Большая тройка»: крупнейшие рейтинги университетов

В настоящее время наиболее влиятельными мировыми рейтингами университетов считаются следующие:

1) Академический рейтинг мировых университетов, который часто называют «Шанхайский рейтинг» (Academic Ranking of World Universities – ARWU), составляется Шанхайским университетом Цзяо Тун ежегодно с 2003 г.¹;

2) Рейтинг мировых университетов QS («QS World University Ranking»)², с 2004 г. он составлялся компанией «Quacquarelli Symonds» совместно с журналом «Times Higher Education Supplement», а с 2009 года «Quacquarelli Symonds» составляет его самостоятельно;

3) Рейтинг мировых университетов «Times Higher Education» (Times Higher Education World University Ranking – THE), который «Times Higher Education» составляет с 2010 г.³

Кроме того, существует ряд глобальных рейтингов, имеющих меньшее влияние и известность⁴.

Методология ARWU

При построении ARWU в исходный массив включаются все университеты, с которыми связаны: обладатели Нобелевской премии и медали Филдса; высокоцитируемые исследователи; авторы статей, опубликованных в журналах «Nature» и «Science»; университеты, сотрудники которых опубликовали значимое количество статей, проиндексированных расширенным индексом научного цитирования (Science Citation Index-Expanded) и/или индексом цитирования в социальных науках (Social Science Citation Index).

Институты высшего образования ранжируются в соответствии с их «академической или исследовательской производительностью», определяемой как индекс, составленный из следующих индикаторов:

1. Выпускники («Alumni») и сотрудники («Award»), получившие Нобелевскую премию по физике, химии, медицине и экономике или медаль Филдса. Учитываются выпускники, получившие в данном университете степень бакалавра, магистра или доктора, и сотрудники, работавшие в университете в момент получения премии. Для выпускников разной давности используются разные веса, как и для сотрудников-лауреатов разных периодов. Если сотрудник при получении премии работал в двух и более университетах, каждый университет получает балл, если приз получили более одного человека, приз каждого учитывается с весом, равным его доле приза.

2. Высокоцитируемые исследователи («HiCi») в 21 широкой дисциплинарной области наук — науках о жизни, медицине, физике, инженерии и социальных науках. Их количество берется из базы данных «ISI Web of Knowledge» в неизменном виде.

3. Количество статей, опубликованных в журналах «Nature» и «Science» («N&S») сотрудниками университета за пять лет. В случае коллективного авторства используются весовые коэффициенты, учитывается только два типа публикаций — статьи и материалы исследований.

4. Статьи, проиндексированные в SCIE и SSCI в предшествующем году («PUB»). Учитываются только статьи и материалы исследований.

5. Академическая производительность («PCP») относительно размера институции, которая определяется как сумма баллов всех предыдущих индикаторов, деленная на количество академических сотрудников, пересчитанное на полные ставки, а при недоступности данных о сотрудниках берутся «взвешенные суммы баллов по предыдущим пяти индикаторам» [11].

Для получения итогового показателя по университету баллы по каждому индикатору взвешиваются: вуз с максимальным показателем получает 100 баллов, остальные нормируются относительно него. Авторы рейтинга утверждают, что «при распределении данных для каждого индикатора учитывались какие-либо значимые смещения, и в случае необходимости для корректировки индикаторов использовались

¹ <http://www.arwu.org/>, <http://www.shanghairanking.com/>

² <http://topuniversities.com/university-rankings/>

³ <http://www.timeshighereducation.co.uk/world-university-rankings/>

⁴ См., например: Рейтинг эффективности научных работ мировых университетов (Рейтинг НБЕАСТ) <http://ranking.heeact.edu.tw/>; Лейденский рейтинг <http://www.leidenranking.com/>; Рейтинг мировых университетов «Webometrics» <http://www.webometrics.info/>; «Ю-Мультиранк» <http://www.umultirank.org/>; Рейтинг «SCImago Institutions» <http://www.scimagoir.com/>.

стандартные статистические техники» [11], однако подробнее методику анализа смещений рейтинга не раскрывают.

Методология QS

Мировой рейтинг университетов QS с 2004 до 2010 г. формировался в сотрудничестве с «Times Higher Education», с 2010 г. компания «Quacquarelli Symonds» строит его самостоятельно по прежней методологии. Измеряется четыре вида индикаторов: качество исследований, востребованность выпускников работодателями, качество преподавания, интернационализация. В QS используются опросы представителей академического сообщества и работодателей для измерения репутации университетов; также используются количественные показатели — количество цитат на одного преподавателя, соотношение преподавателей и студентов, количество иностранных преподавателей и иностранных студентов. На долю показателей репутации и на долю количественных показателей приходится по 50 % итогового индекса⁵.

Репутация университетов, измеряемая посредством опроса, определяется в пяти областях: искусство и гуманитарные дисциплины, науки о жизни и биомедицина, социальные науки, естественные науки и технологии.

Отбор респондентов в 2008 г. производился по двум базам: «The World Scientific» и «Mardev», которые содержат электронные адреса соответственно около 180 тыс. ученых из разных областей естественных и социальных наук, а также еще 12 тыс. исследователей в области гуманитарных дисциплин. Респондентам была направлена анкета, в которой их просили выбрать 30 университетов (исключая свой), которые они считали лучшими в своей области. За три года организаторы исследования собрали 9 386 ответов [8]. В 2012 г. база данных составляла более 46 тыс. ответов, собранных за три года, при этом в случаях, когда один и тот же участник отвечал на вопросы в более чем один год из охваченных, учитывались его последние ответы. В 2012 г. также использовалось перевзвешивание по регионам и по дисциплинам.

Методика опроса работодателей сходна с опросом представителей академии. Использовалась база данных компаний, сеть партнеров, с которыми сотрудничает «Quacquarelli Symonds», ряд

университетов предоставил список работодателей-партнеров. Исследование 2008 г. включало 2339 рекрутеров [8], исследование 2012 г. — более 25 тыс., в этот год применялось перевзвешивание по географическому расположению и по дисциплинам.

Цитаты на одного сотрудника измерялись в 2008 г. по данным баз «Web of Science», «Scopus» и «Google Scholar» за последние 5 лет, в 2012 г. данные за пять лет брались только из «Scopus». Полученное из баз число делилось на количество профессоров и исследователей, работающих на полную ставку.

Для определения соотношения сотрудников факультета и студентов данные о количестве студентов брались с сайтов государственных организаций управления образованием, статистических бюро и т. п. Если было невозможно выяснить число студентов, обучающихся на полной программе, оно заменялось общим числом студентов. В случаях, если для сотрудников факультета было доступно два показателя — число преподавателей и число исследователей, использовался последний.

Показатели зарубежных сотрудников и преподавателей определялись как доли преподавателей и студентов из-за рубежа.

Методология THE

Рейтинг мировых университетов «Times Higher Education» (Times Higher Education World University Rankings) содержит индикаторы, отражающие, как утверждают создатели, «все их (исследовательских университетов) основные миссии». Всего индикаторов тринадцать, они сгруппированы в пять категорий: преподавание, исследования, цитирование, вложения индустрии, международная перспектива.

В данном рейтинге не учитываются университеты, где нет бакалаврских программ, если у них специализация в одной узкой области и если результат их исследовательской деятельности составляет менее 1000 статей за период с 2006 по 2010 г. Исключение делается для университетов, специализирующихся в областях, для которых характерны малые объемы публикаций, — в инженерии и в искусстве и гуманитарных дисциплинах.

Для определения итоговых показателей используется z-стандартизация для всех индикаторов, кроме результатов исследования академической репутации. Для показателей репутации, данные которых сильно смещены «в пользу маленького числа институций, находящихся

⁵ Веса индикаторов QS (2012 г.) см.: <http://www.topuniversities.com/university-rankings-articles/world-university-rankings/qs-world-university-rankings-methodology>

наверху рейтинга», в 2011–2012 гг. был добавлен «экспоненциальный компонент» с целью «увеличить дифференциацию между институтами, находящимися ниже по шкале».

Категория «Международная перспектива: люди, исследования» показывает, в какой степени университет является частью глобальной системы высшего образования. Она включает показатели, отражающие долю иностранных студентов (индикатор привлекательности университета в международном масштабе), долю иностранных сотрудников (индикатор способности привлекать лучших сотрудников со всего мира), долю публикаций с иностранными соавторами (нормализуется в соответствии с профилем комбинации специальностей университета).

К категории «Исследования: объем, доход, репутации» относятся: репутация университета с точки зрения качества исследований (по результатам опроса более 16 тыс. представителей академического сообщества); доходы от исследований (относительно количества сотрудников и нормализованные по покупательской способности местной валюты — данный показатель нормализован с учетом комбинации специальностей университета); продуктивность исследований (количество работ, опубликованных в научных журналах, индексируемых «Thomson Reuters», на одного сотрудника, с учетом размера университета, и нормализованное в соответствии с комбинацией специальностей университета).

Категория «Цитаты: влияние исследований» включает единственный индикатор, отражающий «роль университетов в распространении новых знаний и идей». Количество цитат подсчитывалось по базе «Web of Science» за пять лет, массив публикаций охватывал все индексируемые журналы, опубликованные в 2006–2010 гг., цитаты брались за 2006–2011 гг. Данные были нормализованы по предметным областям, университеты, опубликовавшие меньше 200 работ за год, были исключены из анализа.

Категория «Доход от индустрии: инновации», отражающая вклад университетов в практику производства, представлена одним индикатором: объем доходов университета в виде грантов или заказов на исследования от индустрии, бизнеса относительно численности сотрудников университета.

Категория «Преподавание: среда обучения» представлена индикаторами:

— репутация университета в отношении качества преподавания, определенная в результа-

те опроса представителей академического сообщества; опрос провела компания «Thomson Reuters» по заранее сформированным спискам респондентов, с соблюдением пропорций географических регионов и профилей университетов по специальностям; ответило 16 639 человек;

— соотношение студентов и преподавателей — как индикатор качества преподавания, так как адекватное соотношение, по замыслу разработчиков, обеспечивает должное внимание преподавателя каждому студенту;

— соотношение количества присвоенных степеней доктора и бакалавра — как индикатор исследовательской активности университета и следующей из этого привлекательности для обучения;

— количество присвоенных университетом степеней доктора, с учетом численности сотрудников университета, нормированное по комбинации представленных в университете специальностей;

— доход университета с учетом численности сотрудников, нормированный по покупательской способности местной валюты, — как индикатор уровня материальных возможностей университета⁶.

Методологическая критика крупнейших рейтингов

Методологическая критика рейтингов достаточно широко представлена в академических публикациях. Исследователи и руководители в сфере высшего образования исходят из того, что рейтинги — это «всерьез и надолго», это данность (прямо высказана эта установка, например, в [1]), а такие представления обуславливают большое внимание к тому, как сконструирован инструмент, которым измеряется качество университетов.

Внутри методологической критики можно выделить несколько основных направлений дискуссии: выбор показателей, включаемых в рейтинговую систему; проблемы, связанные с использованием основных для мировых рейтингов университетов видов показателей: библиометрических и реер review; проблемы агрегирования индикаторов в единый индекс.

⁶ Таблицу соотношения категорий оценки, их индикаторов и весов см.: <http://www.timeshighereducation.co.uk/world-university-rankings/2012-13/world-ranking/methodology>

Выбор показателей

Состав показателей рейтинга всегда ограничен, как правило, относительно небольшим числом признаков. Причин две: во-первых, он является результатом выбора разработчиков и отражает их представление о том, какие характеристики университета измеряют его успешность, качество работы [5]. Во-вторых, выбирать индикаторы приходится исходя из доступности данных, что накладывает особенно серьезные ограничения при построении мировых рейтингов, поскольку в мировом масштабе данные доступны лишь по очень ограниченному набору признаков.

Основные типы индикаторов, которые используются при составлении мировых рейтингов университетов, — библиометрические, получаемые в результате анализа баз данных публикаций и цитирования, и индикаторы репутации, получаемые посредством опросов представителей академического сообщества (peer review) и других экспертов.

Критика выбора показателей связана, во-первых, с отсутствием необходимых показателей в системе рейтинга и наличием «лишних» индикаторов в системе рейтинга; во-вторых, с отсутствием обоснования или некорректным обоснованием выбора индикаторов.

A. Van Raan [17] критикует выбор индикаторов разработчиками ARWU, которые сделали акцент на библиометрических характеристиках, отказавшись от измерения репутации. Он считает, что исследования посредством peer review, которые исторически были первым источником данных для рейтингов, должны остаться главным способом оценки качества научной работы, а библиометрические методики могут использоваться как дополнительные к peer review, для повышения объективности измерений.

В ARWU отсутствуют индикаторы, характеризующие процесс обучения, качество преподавания, успеваемость студентов. Нужно отметить, что отсутствие таких показателей свойственно и многим другим рейтингам. A. Van Raan [17] же полагает, что наиболее эффективным способом отразить качество университета является использование показателей, которые отражают «выходные» характеристики студентов (т. е. результат их учебы) по критериям, которые определяются как наиболее значимые в конкретном социальном контексте. Эти качества, по мнению автора, характеризуют человеческий капитал (здесь он ссылается на теорию Г. Беккера), сформировавшийся в результате деятель-

ности университета. Если речь идет о рейтингах для отдельной страны, то критерии должны отражать «выходные» качества студентов, наиболее ценные в данной стране, если речь идет о мировых рейтингах, то следует использовать некие универсальные критерии.

Включение показателей вызывает критику, например, в случае учета доли иностранных преподавателей и студентов. Об этом пишет, в частности, R. Holmes [7]. В [19] указывается, что не всегда «мировой уровень» университетов (например, если оценивать по уровню исследований) сопровождается всеми признаками, которые используются как индикаторы при построении рейтингов. Так, в японских университетах «исследования мирового уровня» осуществляются при отсутствии большого числа иностранных ученых и рабочим языком является японский, а не английский.

A. Van Raan [17] подчеркивает, что составители ARWU важнейшей характеристикой университета считают высокие научные показатели, и не просто высокие, а отражающие максимально возможные достижения в науке. Такой выбор вызывает вопрос о том, насколько эти «экстремальные» достижения отражают качество университета в целом.

J. Ioannidis с соавторами [9] также критикуют показатели ARWU, связанные с получением Нобелевских премий и медалей Филдса выпускниками и сотрудниками как не отражающие качество образования, которое дает университет. Для выпускников это так в силу недоказанности влияния качества обучения в университете на получение награды, для сотрудников — вследствие того, что награда засчитывается университету, где они работали в момент получения, а не в момент совершения революционного открытия. В итоге данный показатель отражает способность университета привлекать выдающихся ученых, а не быть местом выдающихся исследований.

Также данные авторы указывают на проблему усреднения — разные подразделения в одном университете могут демонстрировать очень разные уровни качества работы, однако внутри-университетская вариативность не учитывается. Те же авторы указывают на то, что для ряда показателей проявляется «зависимость от размера», поскольку и в Шанхайском рейтинге, и в рейтинге «Times» нормализация делается не для всех показателей [9].

Еще одна позиция критики состоит в том, что некоторые индикаторы измеряют средние

или общие показатели (например, количество цитирований на одного сотрудника), другие фиксируют только выдающиеся события (например, количество нобелевских лауреатов и обладателей медали Филдса). Оба типа индикаторов потенциально полезны, однако возникает вопрос о цели их использования. Также если говорить о крайних значениях, предлагается вместе с позитивными крайними значениями использовать негативные — например, ученых, уличенных в обмане, статьи, которые не цитируют, и т. п. [9].

Можно заметить, что критика показателей, уже присутствующих в рейтингах, как правило, связана с сомнениями в том, что эти характеристики являются необходимыми для хорошего университета.

В работе J.-C. Billaut, D. Bouyssou и P. Vincke [2] указывается, что выбранные разработчиками ARWU индикаторы соответствуют описанию университетов Лиги плюща и «Оксбриджа», чем и можно объяснить высокие места этих университетов. Иначе говоря, в основе рейтинга лежит представление о единственной модели университета как идеале, а значит, успешными в данном рейтинге могут быть только университеты, соответствующие этой модели.

L. Waltman с соавторами [18] приводят систематическое изложение критики наиболее известных рейтингов и делают следующие заключения. Во-первых, методологическая проблема большинства рейтингов состоит в комбинировании разного типа показателей в одном индексе, что делается достаточно произвольным образом. В результате интерпретация полученного индекса затруднена. Во-вторых, проблемой является применения одного и того же набора показателей к разным по своим миссиям университетам (всех «измеряют одной линейкой»). В результате один из одинаково успешных в реализации своих специфических целей университетов может оказаться лидером, а другой аутсайдером в рейтинге. Решить эту проблему, хотя бы отчасти, помогло бы выделение категорий, внутри которых университеты были бы близки по миссиям, и оценка каждой категории в отдельности.

Итак, по мнению исследователей, формирование системы показателей в наиболее старых и известных мировых рейтингах университетов выполнено не оптимальным образом, при этом отсутствуют более или менее обстоятельные обоснования выбора индикаторов, которые в них используются.

Библиометрические показатели

Наиболее подробную критику библиометрических показателей приводит A. Van Raan [16, 17], также их анализируют R. Holmes [7], J. Ioannidis с соавторами [9], M. Zitt и G. Filliatreau [20], H. Moed [12].

A. Van Raan выделяет две группы проблем при использовании этого типа показателей: технические — проблемы сбора «чистых» и соответствующих задачам данных для дальнейшего анализа, и методологические, связанные с построением релевантных индикаторов и индексов.

Технические проблемы связаны с использованием готовых баз данных — «Scopus», «Web of Science» и т. п., которые содержат следующие ошибки:

— элементарные ошибки: различия в транскрипции имен авторов из неанглоязычных стран, ошибки в номерах журналов и т. п. Доля таких ошибок может быть значительной (до 30 % [17]). Также ошибки в названиях университетов, различия в транскрипции и т. п.;

— неконсистентность и ошибки при указании институционального субъекта публикации: а) указание в одних случаях исследовательского центра, лаборатории, другой институции, входящей в университет, в других — университета в целом; б) путаница при наличии в одном городе двух разных университетов с одинаковыми или почти одинаковыми названиями (Левен, Брюссель); в) проблема идентификации субъектов, формально входящих в одну структуру с единым названием, но фактически различных.

A. Van Raan [17] и R. Holmes [7] указывают на необходимость исправить все эти ошибки для верной идентификации университетов, в которых работают авторы, и самих авторов, прежде чем использовать данные в рейтингах. Объем ошибок может быть значительным, и их исправление важно для обеспечения валидного результата.

J. Ioannidis с соавторами [9] утверждают, что в ARWU способ анализа в случаях, когда работу написали несколько авторов, далек от оптимального, что не учитываются особенности базы данных, организованной по областям науки, которые с точки зрения «плотности» цитирования могут быть более плотными и менее плотными вследствие масштаба этих областей.

Содержательная методологическая дискуссия связана с использованием количества цитат в качестве индикатора. В частности, не ясно, как решены проблемы ошибок, возникающих в этом

процессе, самоцитирования и т. п. Шанхайский рейтинг применяется в том числе в отдельных сферах (которые представляют собой крупные направления), но внутри этих сфер разные дисциплины также имеют разный вес, что не учитывается. Например, «плотность» цитирования в иммунологии в десятки раз превышает плотность цитирования в сельскохозяйственных науках. Для цитирования также важен выбор временного интервала, за который рассматриваются цитаты. Это, в свою очередь, зависит от продолжительности жизни и размера институции.

A. Van Raan [16] выделяет следующие проблемы библиометрии:

- важнейшая проблема — выбор индикаторов и их весов в индексе;
- надежность источника данных;
- индекс цитирования хорошо работает для естественных наук и математики и гораздо хуже — для инженерии, социальных и гуманитарных дисциплин из-за разной роли статей в этих сферах;

- доминирование американских университетов — не ясно, в какой степени это обусловлено высоким уровнем академической активности, а в какой — именно практиками цитирования и публикаций;

- использование при построении индекса разных типов статей — обычных статей, писем, заметок и обзоров. A. van Raan [16] предлагает использовать разные весовые коэффициенты для учета типа статьи;

- использование неанглоязычных журналов в рассмотрении сильно влияет на результаты.

A. Van Raan критикует «простые» и «быстрые» методы построения рейтингов на основе библиографических данных, которые распространились благодаря существованию доступных — бесплатно или за деньги — баз данных цитирования, в результате чего стало возможным и растет использование библиометрических данных не экспертами, в том числе и для построения рейтингов. A. Van Raan подчеркивает проблему некорректного использования доступных данных, а именно использование готовых индексов, упрощенный подход к построению индексов, без основательной разработки методологии и осмысления того, что купленные данные измеряют. Данная проблема весьма актуальна вследствие стремления руководителей вузов сэкономить деньги и время на серьезных исследованиях.

Кроме того, по A. Van Raan, не бесспорна роль статей как основного отражения результа-

тов научной работы. «Библиометрическая оценка качества исследований основана на одном центральном допущении: ученые, которые хотят сказать что-то важное, активно публикуют свои данные в открытой международной серийной литературе, журналах» [17]. Однако не во всех дисциплинах статьи важны в равной степени — в гуманитарных дисциплинах книгам придается больший вес, а, например, в математике и технических дисциплинах активно используются электронные публикации, которые не учитываются базами данных. Поэтому, согласно A. Van Raan, необходимо изучить практики публикаций в разных областях, чтобы понять, насколько библиометрические индикаторы и индексы цитирования способны отразить реальное положение институтов и коллективов в научном поле.

M. Zitt и G. Filliatreau [20] указывают на еще одну проблему, которая может возникнуть при использовании библиометрических данных: зависимость показателей от размера университета, которая касается абсолютных показателей — количества публикаций, цитат и не касается относительных.

Peer review или репутация

Для показателей рейтингов, отражающих репутацию университетов, определенную посредством опроса экспертов (peer review), как и для библиометрических показателей, можно выделить методологическое и техническое направления критики. Основные методологические вопросы, которые ставят исследователи: что на самом деле показывает репутация? Отражает ли она качество университета в целом, качество образования в нем или символический капитал в чистом виде?

A. Van Raan [17] утверждает, что показатель репутации может использоваться скорее для определения наиболее известных университетов, но не лучших. Ряд исследований подтвердил, что показатель репутации в реальности часто отражает не те характеристики, которые он должен был отражать по замыслу разработчиков. D. Dill и M. Soo [5], ссылаясь на A. Graham и N. Thompson⁷, приводят рейтинг «US News & World Report», в котором оценки репутации университетов коллегами (peer assessment) по

⁷ Graham A., Thompson N. Broken ranks: US News' college rankings measure everything but what matters. And most universities don't seem to mind // Washington Monthly. 2001. № 33(9). P. 9–14.

замыслу создателей должны отражать характеристики преподавания. Но анализ показал, что этот индикатор гораздо сильнее коррелирует с расходами на исследования и разработки, чем с показателем успешно завершивших обучение студентов. D. Dill и M. Soo [5] также ссылаются на A. Astin⁸, утверждая, что дорогостоящие опросы для исследования репутации можно не проводить, так как показатели репутации рейтинга «US News & World Report» хорошо предсказываются объективными переменными, данные по которым легкодоступны: селективность среди абитуриентов (*undergraduate selectivity*); расходы на одного студента; число факультетов, присуждающих степень доктора.

V. Burris [3], опираясь на теории М. Вебера и П. Бурдьё, проверяет гипотезу о том, что престиж факультета зависит не только и не столько от его научной продуктивности, сколько от места «в сетях ассоциаций и социального обмена — т. е. это форма социального капитала», хотя, разумеется, продуктивность остается необходимым условием высокого статуса. Сети социального обмена, о которых идет речь, формируются между факультетами путем обмена выпускниками с докторской степенью. Автор использует данные о трудоустройстве выпускников PhD-программ по социологии и анализ социальных сетей в качестве методологии и приходит к выводу, что центральность в межфакультетских сетях найма объясняет 84 % вариации престижа факультета.

V. Rindova и соавторы [14] выяснили, что большие заслуги сотрудников бизнес-школ, подтвержденные сертификатами и выраженные в публикациях в лучших журналах, т. е. подтвержденное признание среди экспертов в соответствующей области, способствуют высоким оценкам и у потребителей.

Технические проблемы при использовании в рейтингах показателей peer review аналогичны проблемам любых масштабных опросов [7, 6, 17]:

- неясный, несистематичный отбор экспертов, географический «перекос»;
- невысокое качество вопросника (так, представители академии могут судить об исследовательской работе других университетов по результатам, но о процессе преподавания там им, скорее всего, известно очень мало);
- низкий процент возврата;

— недостаточная компетентность экспертов (представители академии могут компетентно судить преимущественно о факультетах своей дисциплины, которые находятся в том же регионе или стране).

Применительно к рейтингу «Times» J. Ioannidis с соавторами [9] указывают на недостатки опроса экспертов — представителей университетов. Во-первых, экспертам предлагается назвать 30 лучших университетов мира в своей области, но не предлагаются никакие «входные данные». В то же время едва ли даже специалисты могут определить эти университеты, особенно в том, что касается таких внешне непосредственно не наблюдаемых явлений, как преподавание. Процесс отбора экспертов из университетов также непрозрачен, а уровень возврата вопросников в 2006 г. был менее 1 %. Относительно опроса экспертов-рекрутеров процент отбора и уровень возврата не раскрывался.

В целом и для библиометрических показателей, и для показателей peer review, и для других видов показателей актуальна общая проблема качества первичных данных. В одном случае оно может оказаться низким в силу ошибок в базах публикаций и цитирования, в другом — в силу слабостей опросного метода, в третьем, например, в результате умышленных манипуляций или неумышленного искажения данных университетами [18]. При этом разработчики рейтинга, как правило, не дают достаточной информации о том, происходит ли и как происходит «чистка» первичного массива данных, как не раскрывают и другие исходные данные.

Способы агрегирования

В ряде работ критического характера обсуждаются способы агрегирования индикаторов в единый индекс. В наиболее известных рейтингах это агрегирование происходит путем сложения взвешенных значений индикаторов, при этом веса присваиваются исходя из представления разработчиков индекса о важности того или иного показателя.

В работе V. Jeremić, A. Marković и M. Bulajić [10] представлен анализ влияния весов индикаторов на результаты рейтинга ARWU и предложен статистический метод для более обоснованного присвоения этих весов — I-расстояние (I-distance). Анализ влияния весов показал, что их изменение значительно влияет на результат. Согласно расчетам I-расстояния, наиболее значимой переменной оказалась «N&S»; индикаторы «HiCi», «Award», «PCP» и «Alumni» показали

⁸ Astin A. W. *Achieving educational excellence: a critical assessment of priorities and practices in higher education.* San Francisco ; London : Jossey-Bass Publishers, 1985.

близкие друг к другу значения I-расстояния; наименее важной оказалась переменная «PUB». Поскольку реально используемые при расчете рейтинга веса распределены иначе, авторы предлагают изменить их.

P. Paruolo, M. Saisana и A. Saltelli [13] также обращаются к проблеме присвоения «правильного» веса, т. е. обозначения важности индикаторов в рамках интегрального показателя. Они показывают, что существующие веса не отражают реальной значимости переменных и не эффективны, так же как линейное агрегирование (при котором взвешенные значения переменных складываются). Предложение состоит в использовании для определения меры значимости переменных корреляционного отношения Пирсона (Pearson's «correlation ratio»). Авторы называют этот показатель «главным эффектом», он определяется на основе «статистического моделирования отношения между переменной и индексом». Также в качестве системы мер по улучшению методологии рейтингов авторы предлагают: 1) не подсчитывать итоговые индексы, когда данные по отдельным переменным «отражают важные результаты, которые затрудняют присвоение им целевых весов в агрегированном индексе»; 2) пересмотреть схему агрегирования, перейти с линейной на полностью или частично нелинейную; 3) «рассмотреть разные стратегии присвоения весов, чтобы выбрать ту, при которой статистический показатель расхождения между ожидаемым весом и важностью переменной минимален».

J.-C. Billaut, D. Bouyssou и P. Vincke [2] дают подробный критический анализ рейтинга ARWU, в частности, релевантности показателей и способа их агрегирования с точки зрения концепции многокритериального принятия решений. Их вывод: данный рейтинг методически несостоятелен. Основная претензия авторов связана с весами и методом агрегирования показателей. Они утверждают, что веса «не должны интерпретироваться как отражающие “важность” критерия», что они «в действительности связаны с нормализацией критериев» и их необходимо пересчитывать в случае, если меняется что-либо в процедуре нормализации. «Поскольку каждый год авторы рейтинга нормализуют свои критерии, присваивая 100 баллов лучшему по баллам университету по каждому критерию, и поскольку каждый год ненормализованные баллы лучшего по баллам университета по этому критерию, скорее всего, меняются, веса должны были меняться каждый год, чтобы покрыть эту

новую нормализацию» [2]. Так как авторы рейтинга не пересчитывают веса, результат, согласно авторам статьи, фактически оказывается неинтерпретируемым. Даже если бы создатели рейтинга избежали данной ошибки, взвешенная сумма представляется J.-C. Billaut и коллегам слабым методом агрегирования критериев, который можно заменить более эффективными, используемыми в многокритериальном принятии решений. Хотя в данном случае критика касалась Шанхайского рейтинга, авторы утверждают, что примерно тот же анализ и выводы могут быть сделаны и относительно рейтинга «Таймс» (Times Higher Education Supplement Ranking).

C. Tofallis [15] также рассматривает проблему агрегирования показателей и предлагает к использованию мультипликативный подход, предполагающий на последнем шаге не сложение, как в большинстве известных рейтингов, а умножение значений показателей.

Помимо критики выбранных показателей и методики их агрегирования, критику встречают и другие методические характеристики рейтингов. Так, J.-C. Billaut, D. Bouyssou и P. Vincke [2] считают, что показатель периодичности сбора данных для рейтинга не рационален, так как показатели, связанные с исследованиями передового уровня, не могут сильно измениться за год. Вследствие этого изменения рейтинга связаны скорее со случайными флуктуациями, чем реальными изменениями в университетах.

Эти же авторы [2] указывают на существенные изменения методологии ARWU от года к году, которые также ставят под сомнение сопоставимость результатов разных лет. Кроме того, разные периоды, за которые берутся данные для разных показателей ARWU, никак не обоснованы [2], а включение относительного критерия («PCP») наряду с абсолютными делает индекс бессмысленным.

Создатели Шанхайского рейтинга отвечают на критику, появившуюся в академических журналах после публикации первого рейтинга в 2003 г. [4], однако их реакция в целом не выходит за рамки описания методики, несколько более подробного, чем в исходно представленных материалах. Очевидно, радикальное решение проблем, указанных в критических работах, невозможно — в том числе потому, что любое изменение методики имеет большое влияние на результат, отчего данные рейтингов за два последовательных года становятся несопоставимыми.

Выводы

Итак, принципы построения мировых рейтингов университетов не лишены недостатков, которые служат постоянным источником критики, начиная с первой публикации Шанхайского рейтинга в 2003 г. Критика касается в первую очередь трех аспектов методологии: выбора индикаторов, отражающих качество университетов; качества сбора данных по ним и выбора способа агрегирования индикаторов.

Основные виды индикаторов, которые используются при построении мировых рейтингов университетов, — библиометрические, измеряющие публикации и цитирование, и реервю, измеряющие репутацию университетов в академическом сообществе и среди работодателей. Использование обоих видов индикаторов вызывает претензии относительно того, как собираются и подготавливаются для дальнейшего использования данные, и того, как интерпретируются собранные данные, которые, согласно выводам ряда исследователей, в реальности отражают не всегда то, что должны были отражать по замыслу разработчиков рейтинга. Основной — и в принципе не разрешимой полностью — является проблема определения набора индикаторов, который отражал бы важнейшие характеристики хорошего университета, содержал все необходимые индикаторы и не содержал малозначимых.

Проблема агрегирования значений разных индикаторов связана с тем, какой «физический смысл» имеют веса (поскольку во всех трех крупнейших мировых рейтингах университетов индикаторы взвешиваются) и какой способ объединения индикаторов в единый индекс следует использовать. Разработчики рейтингов используют весовые коэффициенты для обозначения важности индикаторов, однако критики такого подхода утверждают, что в этой роли веса использоваться не должны. Подвергается критике и принцип формирования индекса путем сложения взвешенных значений индикаторов.

Перечисленные проблемы, согласно выводам большинства критиков, означают, что результаты рейтингов следует использовать с осторожностью, осознавая имеющиеся у инструмента измерения слабости и ограничения, и ни в коем случае не воспринимать полученные университетом баллы как истину в последней инстанции в определении качества университета. Данный вывод, как считают авторы работ, анализирующих рейтинги, необходимо принимать

во внимание всем акторам, использующим их результаты: руководству университетов и государственных органов управления образованием, будущим студентам и их родителям, работодателям, нанимающим выпускников.

1. *Altbach P. G.* The globalization of college and university rankings // *Change: The Magazine of Higher Learning*. 2012. Vol. 44, № 1. P. 26–31.

2. *Billaut J.-C., Bouyssou D., Vincke P.* Should you believe in the Shanghai ranking? // *Scientometrics*. 2010. Vol. 84, № 1. P. 237–263.

3. *Burris V.* The academic caste system: Prestige hierarchies in PhD exchange networks // *American Sociological Review*. 2004. Vol. 69, № 2. P. 239–264.

4. *Daraio C. et al.* The European university landscape: A micro characterization based on evidence from the Aquameth project // *Research Policy*. 2011. Vol. 40, № 1. P. 148–164.

5. *Dill D. D., Soo M.* Academic quality, league tables, and public policy: A cross-national analysis of university ranking systems // *Higher Education*. 2005. Vol. 49, № 4. P. 495–533.

6. *Florian R. V.* Irreproducibility of the results of the Shanghai academic ranking of world universities // *Scientometrics*. 2007. Vol. 72, № 1. P. 25–32.

7. *Holmes R.* The THES university rankings: Are they really world class // *Asian Journal of university education*. Vol. 1, № 1. 2006. P. 1–14.

8. *Huang M.-H.* Opening the black box of QS World University Rankings // *Research Evaluation*. 2012. Vol. 21, № 1. P. 71–78.

9. *Ioannidis J. P. A. et al.* International ranking systems for universities and institutions: a critical appraisal [Electronic resource] // *BMC Medicine*. 2007. 5:30. URL: <http://www.biomedcentral.com/1741-7015/5/30> (accessed: 18.03.2013).

10. *Jeremić V., Marković A., Bulajić M.* A critical assessment of international university ranking system // *Proceedings of the XIII International Symposium SymOrg 2012: Innovative Management and Business Performance* / ed. M. L. Jakšić, S. Barjaktarović Rakočević University of Belgrad. Faculty of the Organizational Science, 2012. P. 554–559.

11. *Liu N. C., Cheng Y.* The academic ranking of world universities // *Higher education in Europe*. 2005. Vol. 30, № 2. P. 127–136.

12. *Moed H. F.* Bibliometric rankings of world universities [Electronic resource]. CWTS Report 2006-01. 2006. URL: http://www.cwts.nl/hm/bibl_rnk_wrlld_univ_full.pdf (accessed: 18.03.2013).

13. *Paruolo P., Saltelli A., Saisana M.* Ratings and rankings: Voodoo or Science? [Electronic resource]. arXiv preprint, arXiv:1104.3009v2 [stat.AP]. 2011. URL: <http://arxiv.org/abs/1104.3009v2> (accessed: 18.03.2013).

14. *Rindova V. P. et al.* Being good or being known: An empirical examination of the dimensions, antecedents, and consequences of organizational reputation // *Academy of Management Journal*. 2005. Vol. 48, № 6. P. 1033–1049.

15. Tofallis C. A different approach to university rankings // Higher Education. 2012. Vol. 63, № 1. P. 1–18.

16. Van Raan A. F. J. Challenges in ranking of universities [Electronic resource] // Invited paper for the First International Conference on World Class Universities. Jiao Tong University. Shanghai. June 16–18, 2005. URL: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.104.4501&rep=rep1&type=pdf&a=bi&pagenumber=1&w=100> (accessed: 18 марта 2013).

17. Van Raan A. F. J. Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods // Scientometrics. 2005. Vol. 62, № 1. P. 133–143.

18. Waltman L. et al. The Leiden Ranking 2011/2012: Data collection, indicators, and interpretation [Electronic resource]. arXiv preprint, arXiv:1202.3941v1 [cs.DL]. 2012. URL: <http://arxiv.org/abs/1202.3941> (accessed: 18.03.2013).

19. Yonezawa A. Much ado about ranking: why can't Japanese universities internationalize? // Japan Forum. 2010. Vol. 22, № 1–2. P. 121–137.

20. Zitt M., Filliatreau G. Big is (made) beautiful: Some comments about the Shanghai ranking of world-class universities // The World-Class University and Ranking: Aiming Beyond Status / ed. J. Sadlak, Nian Cai Liu. UNESCO-CEPES and Shanghai Jiao Tong University : Cluj University Press, 2007. P. 17–24.

