

Проблемы релевантности оценки знаний обучающихся в системы высшего образования России: пример НИУ ВШЭ.

Введение

Качество и релевантность оценки знаний обучающихся в системе высшего образования является важным элементом построения современной, конкурентной системы образования в ВУЗах. (Doyle, Houston, Osborne, 2007) Различные системы контроля знаний и умений обучающихся позволяют оценивать как качественные, так и количественные показатели (Кириллов, Дубровина, Донина, Голышев, 2005), но страдают от ряда уязвимостей (Сенашенко, 2012), которые могут существенно повлиять на качество получаемого образования. Само по себе выставления оценок связано с проблемой «валидности» оценки продемонстрированным обучающимся результатам, которая может принимать ряд форм (Allen, 2005). В частности, проблемой нашего исследования выступает отклонение от этического поведения и смещение конвенциональной нормы академической этики с течением времени при закреплении отклоняющихся поведенческих стратегий (в программах высшего образования). Подобное смещение постепенно компрометирует институт оценки знаний в целом, в результате чего оценка становится не валидной и возрастает уязвимость к манипулятивным стратегиям (с точки зрения теории игр). Кроме того, возрастает и число участников выбирающих отклоняющиеся стратегии. Это приводит как к значительному ущербу для системы высшего образования в целом, так и к значимым экономическим потерям непосредственных участников образовательного процесса.

“Вершиной айсберга”, которая позволяет нам в этом докладе предполагать наличие отклоняющегося поведения является выраженное **не соответствие распределения академических результатов студентов закону нормального распределения**. Таким образом можно сформулировать два предположения:

1. При использовании валидной и надежной системы оценок можно ожидать распределение соответствующее нормальному.
2. На выборке студентов бакалавриата и магистратуры Высшей школы наблюдается значимое смещение нормального распределения академической успеваемости.

И мы полагаем, что для того, что бы перейти к детальному рассмотрению первого тезиса важно подтвердить наблюдение о действительном наличие смещения распределения относительно нормального в оценке знаний студентов.

Это легко можно сделать используя открытые данные студенческого рейтинга НИУ ВШЭ или обратившись к соответствующей методической документации, например, к отчету об образовательной деятельности НИУ ВШЭ по программам бакалавриата и магистратуры за 2016 год (НИУ ВШЭ, 2017). На Рис 1 - 10 мы можем наблюдать наличие такого смещения распределения относительно нормального в среднем балле студентов образовательных программ бакалавриата и магистратуры по всем дисциплинам и направлениям подготовки. На Рис 11 - 13 мы можем наблюдать, что подобный эффект распространиться не только на студенческий рейтинг, но и на результаты выпускных квалификационных экзаменов и работ.

Феномен “инфляции оценок”

Не соответствие, а если быть точнее, значительное смещение реальной академической успеваемости у студентов образовательных учреждений различных уровней относительно ожидаемого закона нормального распределения в Западной академической традиции фиксируется с 70х годов XX века и носит название “инфляции оценок” (“Grade Inflation”). Количество англоязычных публикаций посвященной этой теме в последние пять лет значительно возросло (всего на данный момент Web of Science индексирует 263 публикации в рецензируемых академических журналах). Кроме того, можно обнаружить достаточно широкое упоминание этого термина и за пределами академического дискурса.

Несколько иная картина наблюдается среди рускоязычных авторов. Нам удалось найти только несколько косвенных упоминаний категории “инфляции оценок” отсылающих к книге “Анатомия американского университета” (Дмитриев Г.А., 2010). Очень схожая картина (по данным системы “Интегрум”) наблюдается в СМИ.

Исключение составляет интервью ведущего научного сотрудника Института образования НИУ «Высшая школа экономики», в прошлом доцента и заведующего кафедрой Основ педагогики и методов исследования в образовании Bowling Green State University, (США) Александра Сидоркина (“Собеседник”, 2017) посвященное эксперименту Рособрнадзора по независимой проверке знаний в ряде российских вузов (Рособрнадзор, 2017). В данном интервью четко сформулировано три тезиса, которые мы полностью процитируем ниже:

1. *“Этот феномен существует давно и в Америке, и в России. О нем говорят долгие годы, но пока без каких-либо последствий. Стоит обратить внимание на то, что оценка, поставленная студенту, не является объективным инструментом. «Пятерка» у одного преподавателя не является равной «пятерке» у другого, за исключением каких-либо точных наук, где можно точно измерить результат задания. Во всех остальных случаях он оценивается субъективно. Оценку студенту особенно всерьез никто и не воспринимает, на самом деле. Есть проекты, которые объективно измеряют уровень знаний. Тестирование в целом дает объективную информацию, а оценки никогда её не давали”* (“Собеседник”, 2017).

2. *“Люди яркие и интересные не потому, что они хотят получить «пятерку», у них совершенно другая мотивация. Мы не видим падения уровня знаний — вот критерий того, что это несущественно”* (“Собеседник”, 2017).

3. *“В последние 25 лет сильно изменилась демография поступающих. В 1980-х годах около 20% выпускников школ поступали в вузы. Сегодня ситуация другая: группы населения, которые раньше не поступали в вузы, теперь поступают. Вузов стало намного больше. Преподаватели жалуются, что уровень студентов стал ниже, но это неверно. 20% процентов населения, которые поступали в вузы в раньше, и сегодня остаются наверху, они учатся в лучших вузах страны и за границей. Все эти субъективные переживания по поводу оглуевшей молодежи не подтверждаются никакими данными. Общий уровень образования населения растет”* (“Собеседник”, 2017).

В результате анализа данных тезисов мы можем сформулировать ряд гипотез:

1. Оценка знаний не является объективным инструментом и вероятно существуют параллельные системы контроля успешности освоение студентами образовательной программы.
2. Оценка знаний и сформированный на ее основе студенческий рейтинг не “воспринимаются в серьез” и не служат основанием для санкций.
3. Система оценка знаний не связана с образовательной мотивацией студентов в целом и тем более не формирует мотивация и убеждения “ярких одиночек”.
4. Не наблюдается падение уровня знаний связанного с феноменом “инфляции оценки”.
5. Демографическая ситуация с ростом поступающих в российские вузы скорее всего не должна очень сильно сказываться на НИУ ВШЭ поскольку он относится к лучшим вузам страны с высоким входным порогом.

I. Первую и вторую гипотезы логично рассмотреть вместе.

A. Во-первых, анализ существующего законодательства Российской Федерации в области образования и нормативно-правовой документации НИУ “Высшая школа экономики” показывает, что система оценки знаний строго регламентирована является основным формальным критерием “успешности” освоение студентом образовательной программы и получения им соответствующий знаний, умений и навыков.

B. Во-вторых, соответствующей нормативно-правовой базой не предусмотрено никакой параллельной системы контроля сопоставимой по масштабам. Некоторые отдельные дисциплины могут такую иметь, например, английский язык. Но и в этом случае наглядно можно наблюдать тенденцию на **расхождение “внутренних” и “внешних” результатов оценки** (Рис. 14).

C. В-третьих, результаты оценки служат основным критерием как для санкций (отчисление за неуспеваемость, перевод на ИУП, лишение скидки) так и для поощрений (получение скидки, стипендии, отбор на академическую мобильность). Студенческий рейтинг - это **основная внутренняя валюта системы образования и основа для институциональных норм**.

D. В-четвертых, средний балл измеряешь на входе в Университет в процессе конкурса по результатам ЕГЭ (напомним, что это один из индексов эффективности вуза по критериям Разобрнадзора) и на выходе - при конкурсе дипломов (или как отсекающий фактор, например, только диплом 1 категории на топовые программы магистратуры) в ходе продолжения академической карьеры. Приложение к диплому государственного образца не содержит никаких сведений о параллельных результатах оценки знаний студентов.

E. И наконец, не сложно предположить случаи когда “оценку никто всерьез не воспринимает (“Собоседник, 2017) - это частный случай перехода студента с одного образовательного уровня университета на другой образовательный уровень в рамках одной среды с манипуляцией результатами конкурса. Большая часть таких

практик достаточно точно описывается понятием “академический инбридинг” который является частным случаем явления “отрицательной селекции”.

II. Таким образом рассмотрение третьей гипотезы сводиться к вопросу - связаны ли институциональные нормы с поведение экономических агентов? Далее мы подробно рассмотрим этот вопрос. Тем не менее стоит отметить наличие работ прямо (Шанин Т., 2014) и косвенно, через анализ стратегии оценки уже не студентов, но вузов (ИПП ЕУ СПб, ЕУ СПб, 2017) посвященных данной теме.

III. Оценка роста или падения “уровня знаний” студентов НИУ ВШЭ не входила в задачи нашего исследования. Тем не менее, как будет показано дальше некоторые предположения можно будет сделать из самого определения “**прогностической валидности**”.

IV. Мы полагаем, что данное явление возможно смоделировать функцией спроса на образование. Но в рамках данного исследования подобной задачи не стояло.

Таксономический анализ “инфляции оценки”

Как достаточно показывает эффект инфляции оценки средний балл студентов НИУ ВШЭ составляет примерно 7.6, а основной диапазон используемых в ходе образовательного процесса оценок варьируется от 6 до 9, а наиболее часто встречающиеся 7 и 8 баллов, соответственно. Отдельно в методических рекомендациях описано применение “крайних” оценок - “0” и “10”, соответственно. Первый применяется как санкция на нарушение академической этики, вторая как поощрение в исключительных случаях. На основании описанной картины можно выдвинуть два конкурирующих предположения.

I. Показанный результат действительно соответствует реальной оценки знаний и обладает восточной валидностью и надежностью. Можно предположить таксономия Соло или Блума выступает четким “внешним” критерием оценки знаний студентов. Такая картина обладает некоторыми наблюдаемыми признаками:

A. **Минимальный уровень близкий к “расширенному резюме”**, как минимум - глубокие системные знания дисциплины переходящее в “междисциплинарную композицию”

B. Наличие второго пика “отрицательных” оценок, который можно наблюдать по экзаменационным ведомостям и текущему контролю.

C. Высокий процент студентов отчисленных по причине академической неуспеваемости, которые образуют второй “пик” не фиксируемый студенческим рейтингом.

D. Увеличение срока обучение связанное с “программным” подходом к образовательному процессу. И достаточно высокая дифференциация этого срока обучения у разных групп студентов.

E. Четкое понимание студентами качественных критериев “дискриминации”, например, различия между оценками “хорошо” и “отлично”.

F. Санкции за “неэтичное” поведение выраженные в соответствующих социальных нормах и практиках (и сигналах которые можно фиксировать).

Г. Достаточная прогностическая надежность результатов оценки знаний, которую можно фиксировать по “внешним” экзаменам .

II. Показанное смещение распределение относительно нормального показывает не только формальный эффект “инфляции оценки” но говорит о том, что для оценка используется стратегия нормализации относительно “внутренних” критериев. Средний балл в этом случае является медианой и качественно сопоставимо по результату с картиной описываемой медианой при нормальном распределении (4.5 балла, при учете 0 как равновероятной 10 оценке).

А. Достаточно низкий процент студентов не справившихся с программой и отчисленных за академическую неуспеваемость.

В. Отсутствие значимого различия по сроку освоения учебной программы.

С. Отсутствие у студентов четкого понимания критериев выставления оценок.

Д. Толерантность к неэтичному поведению у основной группы студентов.

Е. Наличие расхождения между оценкой знаний в вузе и “внешней” оценкой измеряющей те же знания, умения и навыки.

Теоретико-игровая модель поведения участников образовательного процесса

На первом этапе теоретико-игрового анализа необходимо формализовать “игроков” (агентов), на втором, определить их стратегии и наконец, в конце определить “платежи” которые получают игроки в том или ином случае.

В нашей серии игр игроками будут выступать:

I. Студенты.

II. Преподаватели высшей школы

III. Управленческий персонал вуза

IV. Эксперты и управленческий персонал осуществляющие образовательный надзор.

В нашей первой теоретико-игровой модели студенты будут играть между собой в жестко заданных границах институциональных норм.

Для более детальной формализации мотивации мы будем использовать модель предложенную Ильной (Ильин Е.П., 2002) и мотивацию “Опыта потока” (Csikszentmihalyi, 2000). В Таблице 1 мы попробуем формализовать каким образом возможно наблюдать и фиксировать проявления различных типов мотивации у студентов, а точнее стратегическое поведение детерминированной той или иной мотивацией.

Под “бумажной волной” подразумеваться очень высокий уровень овладения индивидом эффективными методами создания выигрышной картины “бумажной” (документальной) реальности. Это может быть очень широкий набор стратегий, включающий в себя как глубокое понимание логики работы бюрократии и наработка эффективных социальных связей в академической и бюрократической среде, так и сознательный плагиат или фальсификация результатов исследований.

В Таблице 2 на основании мотивации и формальных критериев которые могут выступить платежами формализованы стратегии направленные либо на минимизации потенциального ущерба, либо на максимализацию размера выигрыша. Это не игра с нулевой суммой: с одной стороны у игроков разные мотивы, но с другой, как наглядно

Таблица 1. Матрица операционализации мотивации

	Стратегия	Мотивация принадлежности	Мотивация конкуренции	Опыт потока
Мотивация получения знаний	“Академическая”	Попытки интеграции в академическую среду (в том числе и старших курсов)	Активный дискурс, участие в олимпиадах и универсиадах	Научные публикации, академическая мобильность, ...?
Мотивация получения профессии	“Профессиональная”	Выбор “прикладной” направленности	Конкуренция за лучшие места практики и стажировок	Успешная карьера во время учебы, публикации (скорее экспертные)
Мотивация получения диплома	“Серая”	Минимизация ресурсов на учебу	Внеучебная, альтернативные методы самоутверждения, академическая мобильность	Внеучебная деятельность
	“Красная”	Статус “отличника”, полное посещение, минимизация формальных санкций	Топ рейтинга, стипендия, научные публикации, академическая мобильность	“Бумажная волна” (красавица, спортсменка, комсомолка), красный диплом, все 10ки,

Таблица 2. Тип выбранной стратегии по мотивации студента

	Стратегия	Мотивация принадлежности	Мотивация конкуренции	Опыт потока
Мотивация получения знаний	Академическая	“Максимин”	“Максимин”	“Максимин”
Мотивация получения профессии	Профессиональная	“Минимакс”	“Минимакс”	“Минимакс”
Мотивация получения диплома	“Серая”	“Минимакс”	“Минимакс”	“Минимакс”
	“Красная”	“Максимин”	“Максимин”	“Максимин”

видно из Таблицы 1 в реализации этих мотивов вероятно конкуренция из-за ограниченности ресурса, например, количество стипендий для международного обмена.

Как изменяются эти стратегии в случае наблюдения эффекта “инфляции оценки” и каким образом эти изменения можно наблюдать? Возможно выделить “цель мастерства” и “результативную цель” (Малошонок, Семенова, Терентьев, 2015):

I. “Цель мастерства”. В случае применения “внешней” таксономии можно ожидать эффекта “разбегания”, когда будут сформированы несколько устойчивых групп

состоящих из студентов движимых различной мотивацией. У группы студентов с “профессиональной мотивацией” можно будет наблюдать выраженный дисбаланс между оценками за курсы которые они считают значимыми и теми курсами, где они просто не хотят быть отчисленными. “Серую” группу также возможно будет выделить. Можно ожидать высоких оценок среди группы студентов с мотивацией направленной на получение знаний (эффект “расширенного резюме”) и некоторого отставания по максимальным баллам по дисциплинам, но возможно более высокий средний балл у студентов “красной” группы.

II. “Цель результата”. Если таксономии не применяется и оценка является основным критерием обратной связи, то можно наблюдать “размытие” ядер и формирование некоторой “монотонности”. Формирование опыта потока происходит только у “красной” группы студентов поскольку только у них есть обратная полностью отвечающая их потребностям. У “серой” и “профессиональных” групп эта мотивация может быть реализован вне стен университета изначально, а вот у “академической” группы не происходит формирование “опыта потока”, может не произойти формирования мотивации “конкуренции” или вообще реализоваться эффект “выученной беспомощности”.

Таким образом, если предположить, что наблюдаемая нами картина “инфляции оценки” - это результат повторяющейся игры студентов аналогичной дилемме заключенного, то достаточно легко сделать вывод что текущее равновесие не является оптимальным. А это в свою очередь значит, что сам институт оценки в текущем виде скомпрометирован и может образовать как “институциональную ловушку”.

Если в модель добавить еще одного игрока? Можем ли попробовать предсказать какую стратегию по оценке знаний выберут преподаватели? Это может определяется некотором факторами (вероятно, что это не полный список и в дальнейшем модель будет уточнена):

I. Мотивацией преподавательского состава (Рощина Я.М., 2015) приведены в Таблице 3.

Таблица 3 Мотивация ППС

Тип мотивации	В процентах	Уровень по Ядову	В процентах
Мотивация карьеры	30%	Научно-карьеристский	30%
Мотивация досуга	26%	Культурно-развлекательный или алкогольно-сексуальный	70%
Мотивация комфорта	27%		
Мотивация общения с молодежью	17%		

- II. Профессиональной стратегией профессорско-преподавательского состава и его загруженностью.
- III. Действующими нормами и практиками.
- IV. Сложностью и трудоемкостью процедуры.

Например, логично предположить, что группа преподавателей с “культурной” мотивацией и ниже не будет склонна идти на принципиальный конфликт со студентами или администрацией вуза если нет иных мотивов. В рамках текущей модели можно предположить, что даже без дополнительных коэффициентов основанные на такой мотивации стратегии будут не достаточно для того, что бы сформировать новую, более оптимальную в долгосрочной перспективе норму.

Что можно сказать о транслируемых сверху нормах способных изменить текущее равновесие? Значительную роль в оценке играют нормы и практики оценки которые транслирует среда и показатели эффективности по которым оценивается деятельность акторов. На данный момент существует ситуация когда преподаватели вузов оценивают качество студентов, вузы оценивают качество преподавательского состава (в финансовом выражении), а Рособрнадзор оценивает эффективность деятельности самих вузов.

Попробуем рассмотреть какие статистические показатели по деятельности вузов отслеживает и использует для расчета эффективности Рособрнадзор:

1. Образовательная деятельность;
2. Научно-исследовательская деятельность;
3. Международная деятельность;
4. Финансово-экономическая деятельность;
5. Инфраструктура;
6. Трудоустройство выпускников;
7. Кадровый состав.

Какие из этих семи показателей содержат в себе информацию об образовательных стратегиях студентов и какую именно?

- I. Образовательная деятельность - учитывает только средний балл Единого Государственного Экзамена при поступлении в вуз.
- II. Кадровый состав - соотношение количества студентов на одного преподавателя определенной категории в процессе обучения.
- III. Трудоустройство выпускников. Стоит обратить внимание, что учитывается только соотношение нашедших любую работу в последующем после выпуска календарном году. Не учитывается ни количество выпускников получивших работу по профессии, ни средний уровень зарплат.
- IV. Деятельность студентов не учитывается при расчете показателей научно-исследовательской деятельности (если у студента нет ставки) и по международной деятельности.

Таким образом, можно сделать вывод, что фактически студенты правители полностью “отчуждены” от критериев определения эффективности университета в котором они учатся. Фактически ситуация описывается “спросом” абитуриентов на университет. Внутренняя и выходная “академическая успеваемость” является, как было показано ранее,

основным критерием “успешности” студентов, но не играет роли для “успешности” университетов.

Заключение

Мы полагаем, что феномен “инфляции оценки” является наглядным и значимым показателем “нездоровья” высшей системы образования. Своеобразной “вершиной айсберга”, которая может говорить о наличии серьезных проблем с мотивацией, стратегиями и этикой у участников образовательного процесса. В принципе, можно сказать, что “синдром Карлсона” (Радаев В.В., 2010) в полной мере присутствует и в данной области.

- I. Прогностическая валидность и надежность существующей системы оценки знаний, умений и навыков выпускников даже элитных университетов может вызывать очень большие вопросы. Санкции на основании этой системы могут приводить и приводят к систематическим негативным последствиям (например, уходу части магистрантов работающих по профилю образование в государственных учреждениях после потери скидки).
- II. Можно предположить с высокой долей вероятности, что логика “оценки знаний” на основании внешней таксономии не совсем очевидна большинству участников образовательного процесса. Этот вопрос требует дальнейшего исследования.
- III. Результаты образовательной деятельности применительно к студентам не операционализируются существующей системой контроля.
- IV. Как и в случае с проверками частных университетов в России система склонна дискриминировать значительную часть “ярких одиночек”. Системы “внешнего экзаменатора” снижающего вероятность этого события не предусмотрено.
- V. При текущей картине минимальна вероятность возникновения мотивации “опыта потока” у студентов ориентированных на “мастерство”. Это приводит (дополнительный фактор “утечки мозгов”) к вытеснению значительной их части из академической среды.
- VI. Можно предположить роста эффектов “академического имбридинга” с течением времени при сохранении текущей картины. Возможность связи “инфляции оценки” и “отрицательного отбора” требует дальнейшего исследования.
- VII. Критерии “эффективности” университета искажены “синдромом Карлсона” и кроме того, благодаря “эффекту колеи” советского наследия имеют слабую связь с заказчиками и потребителями.
- VIII. Можно предположить, что существуют статические паттерны по “академической успеваемости” студентов, которые коррелируют с мотивацией и образовательной стратегией студента.
- IX. Можно предположить рациональность внедрения дополнительного “индекса” эффективности основанного на “выживаемости студентов” и величине “инфляции оценки”.