

КАЧЕСТВО ОБЩЕГО ОБРАЗОВАНИЯ

О методиках и результатах
интерпретируемой оценки

Игорь Реморенко, Сергей Зададаев

АННОТАЦИЯ В исследовании обсуждаются различные подходы к проведению и интерпретации оценки качества общего образования в общеобразовательных организациях РФ на основании результатов ряда международных исследований (PIRLS, TIMSS и PISA). Предложены две альтернативные модификации утвержденного Министерством просвещения РФ средневзвешенного показателя качества образования и приведены их смысловые интерпретации на примере результатов оценки качества образования РФ в соответствующих международных исследованиях 2015–2016 гг.

КЛЮЧЕВЫЕ СЛОВА Квалиметрия, оценка качества образования, международные исследования PIRLS, TIMSS, PISA.

DOI 10.22394/2078-838X-2020-3-32-37



**Игорь Михайлович
РЕМОРЕНКО**

д. пед. н., доцент, ректор
Московского городского
педагогического университета
(129226, Москва, РФ, 2-й Сельскохо-
зяйственный проезд, д. 4, корп. 1).
E-mail: RemorenkoIM@mgpu.ru

**Сергей Алексеевич
ЗАДАДАЕВ**

к. физ.-мат. н., доцент, руководи-
тель департамента математики
Финансового университета при
Правительстве РФ (125993, Москва,
РФ, Ленинградский пр-т, д. 49).
E-mail: szadadaev@fa.ru

Введение

Проблематика оценки качества общего образования вследствие реализации крупных государственных проектов наиболее ярко проявилась в Российской Федерации с начала нулевых годов XXI века. Это было связано с появлением такого общенационального инструмента проведения преобразований в различных направлениях социально-экономического развития, как «национальные проекты». Впервые

о национальном проектировании было объявлено президентом РФ В. В. Путиным 5 сентября 2005 года на расширенном совещании с членами правительства, руководством Федерального собрания РФ и членами президиума Госсовета. До этой даты действия федерального правительства главным образом затрагивали либо лишь обеспечение текущего нормативно-правового регулирования различных отраслей, либо осуществление методической поддержки или некоторых локальных проектов

в различных регионах путем реализации небольших отраслевых федеральных целевых программ. В основном же реформирование осуществлялось на уровне субъектов РФ.

Однако практика национальных проектов предполагала иной подход — сквозные меры масштабных реформ, единые для всех регионов. Благодаря позитивной конъюнктуре на рынке углеводородов финансирование национальных проектов в разы превысило объемы федеральных целевых программ и потребовало выработки общих показателей эффективности национальных проектов.

Первые показатели эффективности национального проекта «Образование» касались главным образом лишь оценки целевого назначения затрачиваемых ресурсов. Если речь шла о поддержке инновационных проектов школ, в качестве показателя эффективности рассматривалось количество поддерживаемых школ в каждом регионе. Если осуществлялись выплаты классным руководителям, говорилось об объемах перечисленных учителям средств. Лишь спустя несколько лет разговор зашел о необходимости оценивать не только целевое назначение спланированных ресурсов, но и эффекты от таких затрат, видеть, насколько результаты в самом деле системно сказываются на функционировании системы образования. Так среди показателей результативности национальных проектов появились показатели оценки качества образования.

Утвержденная методология и проблема интерпретации результатов

Строго говоря, одними из показателей оценки качества образования можно считать показатели единого государственного экзамена (ЕГЭ). Эта форма государственной итоговой аттестации также стала всеобщей вместе с развитием национального проектирования. Однако правительству не удалось найти релевантной формулы сравнения регионов по результатам ЕГЭ, очевидные риски фальсификации результатов при таком подходе были труднопреодолимы. При этом всегда возникал соблазн замерить результаты образования в международном контексте, ориентироваться на некую внешнюю,

независимую «линейку» оценки качества образования. И действительно, с каждым годом появлялось все больше и больше инструментов такого сравнения. Все они применялись различного рода общественными движениями или организациями, чья деятельность затрагивала интересы разных стран. Россия охотно принимает участие практически во всех международных сопоставительных исследованиях качества образования.

Действующая методология и критерии оценки качества общего образования в общеобразовательных организациях Российской Федерации в группе международных исследований представлены в одноименном приказе Министерства просвещения РФ.

ОЦЕНКА СРЕДНЕВЗВЕШЕННОГО ПОКАЗАТЕЛЯ КАЧЕСТВА ОБЩЕГО ОБРАЗОВАНИЯ В РФ СВЕРХОПТИМИСТИЧНА ИЛИ ЗАВЫШЕНА

Согласно этому документу, «средневзвешенный результат Российской Федерации в группе международных исследований, средневзвешенное место Российской Федерации (не ниже)» определяется посредством вычисления средневзвешенного (среднеарифметического) места РФ по результатам специально отобранных восьми популярных международных исследований качества образования. При этом «за бортом» рассмотрения остается ряд других исследований, в которых Россия также принимает участие — компьютерная и информационная грамотность обучающихся (ICILS), оценка уровня гражданского образования (ICCS), оценка учебно-предметных компетенций (SAM).

Для простоты восприятия приведем пример используемых в методологии конкретных значений восьми рейтингов, представленных в таблице 1, на 2015–2016 годы и соответствующих вычислений средневзвешенного показателя F_{00} качества образования в РФ по утвержденной методике (Приказ Министерства

Таблица 1.

№	Наименование исследования	Год	Рейтинг (место) РФ	Объем выборки (кол-во стран)
1	PIRLS-2016 (4 класс) по качеству чтения и понимания текста (2016a)	2016	$X_1 = 1$	$N_1 = 50$
2	TIMSS-2015 (4 класс) по математическому направлению (2016b)	2015	$X_2 = 7$	$N_2 = 49$
3	TIMSS-2015 (4 класс) по естественнонаучному направлению (Там же)	2015	$X_3 = 4$	$N_3 = 47$
4	TIMSS-2015 (8 класс) по математическому направлению (Там же)	2015	$X_4 = 6$	$N_4 = 39$
5	TIMSS-2015 (8 класс) по естественнонаучному направлению (Там же)	2015	$X_5 = 7$	$N_5 = 39$
6	PISA-2015 по направлению «читательская грамотность» (2016)	2015	$X_6 = 26$	$N_6 = 70$
7	PISA-2015 по направлению «математическая грамотность» (Там же)	2015	$X_7 = 23$	$N_7 = 70$
8	PISA-2015 по направлению «естественнонаучная грамотность» (Там же)	2015	$X_8 = 32$	$N_8 = 70$

просвещения РФ № 590/219, п. 4.2.1 «Формула расчета показателя»):

$$F_{00} = \frac{X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7 + X_8}{8} . \quad (1)$$

Подставим в формулу (1) значения X_i из четвертого столбца таблицы 1 и получим средневзвешенную оценку РФ на 2015–2016 годы по качеству общего образования:

$$F_{00} = 13,25.$$

Однако интерпретировать этот результат очень непросто. Если посчитать, что РФ заняла в среднем по всем восьми исследованиям примерно 13-е место в рейтинге, то возникает вопрос: «13-е место из какого количества стран?». Формула (1) усредняет места без учета объемов выборок N_i , в которых они получены. Иными словами, 10-е место из 100 стран – не то же самое, что 10-е место из 10 стран. В первом случае речь идет о попадании в десятку сильнейших, а во втором – о попадании в самый конец списка.

Ясно, что в формуле среднего арифметического слагаемые должны быть принципиально сравнимы между собой. В нашем варианте

(без учета количества участников того или иного исследования) места становятся в общем случае несравнимыми, а применение формулы (1) – чисто формальным. Но существует серьезное допущение, при котором значение F_{00} удастся интерпретировать корректно. Если предположить, что все не участвовавшие страны показали бы результат хуже нашего, тогда, очевидно, мы сохранили бы все свои места, и оценка $F_{00} = 13,25$ стала бы корректной.

Итак, можно подытожить, что действующая оценка средневзвешенного показателя качества общего образования в РФ сверхоптимистична (или попросту завышена).

Простая нормировка средневзвешенного результата

Первое, что можно предложить для получения хоть сколько-нибудь реальной оценки общего образования в России по результатам данных восьми международных исследований – нормировать среднее место РФ на среднее количество стран-участниц. То есть образовать другой показатель, равный отношению среднего места (F_{00}) к среднему количеству стран-участниц:

$$F_{norm} = \frac{X_1 + X_2 + \dots + X_7 + X_8}{8} : \frac{N_1 + N_2 + \dots + N_7 + N_8}{8}.$$

В итоге получим уже интерпретируемый показатель, по структуре чем-то напоминающий формулу классической вероятности:

$$F_{norm} = \frac{X_1 + X_2 + \dots + X_7 + X_8}{N_1 + N_2 + \dots + N_7 + N_8}. \quad (2)$$

После подстановки в формулу (2) данных четвертого и пятого столбцов таблицы 1 получим нормированную средневзвешенную оценку РФ на 2015–2016 годы по качеству общего образования:

$$F_{norm} = 0,244.$$

Это уже реальная оценка места РФ — «мы занимаем примерно 24-е место из первой сотни стран», или лучше так: «мы оказались в первых 24 % всех стран», не уточняя, сколько их было на самом деле. Но этот вывод не про среднее! Здесь речь идет скорее о некотором виртуальном мега-исследовании, объединяющем все восемь исследований, где растворены друг в друге предметы, возрасты респондентов и методики. Отношение средних не есть среднее, так что F_{norm} получилась все-таки достаточно эфемерной величиной.

ПОЧЕМУ БЫЛИ ВЫБРАНЫ ТЕ ИЛИ ИНЫЕ ВЕСОВЫЕ КОЭФФИЦИЕНТЫ И ЧЕМ TIMSS ДЛЯ НАС ОКАЗАЛОСЬ ЗНАЧИМЕЕ, ЧЕМ PISA

Еще одним очевидным недостатком данного показателя F_{norm} является принципиальная невозможность оценить вклад каждого отдельного исследования — слишком уж здесь все перемешано. Впрочем, этим недостатком обладает и ненормированный показатель F_{00} из-за несравнимости мест при разных объемах выборки.

Усреднение нормированных результатов

Теперь попробуем поступить ровно наоборот: сначала нормировать результаты по каждому исследованию, т.е. образовать для всех восьми рейтингов соответствующие доли

$$\frac{X_1}{N_1}, \frac{X_2}{N_2}, \dots, \frac{X_8}{N_8}$$

(между прочим, уже сравнимые между собой), а затем их усреднить, образовав улучшенный средневзвешенный показатель качества образования:

$$F_{real} = \frac{1}{8} \left(\frac{X_1}{N_1} + \frac{X_2}{N_2} + \dots + \frac{X_8}{N_8} \right). \quad (3)$$

Вычисления по последней формуле (3) с данными таблицы 1 дают нам следующее значение показателя:

$$F_{real} = 0,217, -$$

что может быть интерпретировано как «в среднем по методикам оценки качества общего образования РФ находится в первых 21,7 % стран мира» или «в среднем по методикам оценки качества общего образования РФ занимает 22-е место из первой сотни стран». Интересно, что результат F_{real} оказался сравним с предыдущим F_{norm} , но чуть выше.

Нельзя сказать, что последний вариант абсолютно идеален с точки зрения содержания. Мы должны понимать, что 6-е место из 20 стран будет пропорционально трансформировано в 30-е место из 100 стран — таково свойство примененной нормировки, и в этом есть своя логика «реальности». Вместе с тем входящие в сумму слагаемые представляют собой нормированные доли общего среднего и теперь легко интерпретируются по значимости вклада.

Средневзвешенное или среднее арифметическое

До сих пор мы обсуждали способы нормировки показателя качества образования в РФ

в попытке получить интерпретируемый результат. Во всех трех случаях — в исходном министерском варианте формулы (1) и двух нормированных формулах (2) и (3) — мы в том или ином виде использовали формулу среднего арифметического, т.е. вычисляли суммы рейтингов восьми международных исследований с равными весами, считая исследования равноправными между собой. Однако если условно переписать формулу (1) в терминах самих международных исследований, согласно таблице 1, то получим такой результат:

$$F_{00} = \frac{(PIRLS_1 + TIMSS_2 + TIMSS_3 + TIMSS_4 + TIMSS_5 + PISA_6 + PISA_7 + PISA_8)}{8},$$

или в форме средневзвешенного относительно включенности каждой из трех различных методологий:

$$\frac{1}{8} PIRLS + \frac{4}{8} TIMSS + \frac{3}{8} PISA.$$

Конечно, здесь приведена условная сумма, показывающая лишь, с какими весами (долями единицы) были учтены те или иные международные исследования в итоговом показателе качества F_{00} . Нетрудно заметить, что результаты исследования в методологии *TIMSS*, входящие в сумму с весом 4/8, определяют половину всего показателя качества F_{00} .

Таким образом, уместно говорить именно о средневзвешенном результате, даже при использовании формулы среднего арифметического. Разумеется, в этой связи чисто гипотетически было бы интересно понять, почему были выбраны те или иные весовые коэффициенты и чем исследование *TIMSS* для нас оказалось значимее, чем, например, *PISA*.

Заключение

Приведенные выше две модифицированные формулы (2) и (3) оценки качества образования нельзя воспринимать как абсолютно исключаящие друг друга варианты, скорее наоборот — они поддерживают друг друга в разных

аспектах. Например, легко доказать, что при равных объемах стран-участниц (когда количество участников во всех восьми исследованиях оказалось одним и тем же) значения формул (2) и (3) вообще всегда совпадают:

$$F_{norm} = F_{real}.$$

Нам представляется, что для усредненной оценки качества образования по порядку величины пригодны обе формулы, но последняя F_{real} позволяет провести более детальный покомпонентный анализ. Также, в отличие от ненормированного действующего варианта F_{00} формулы (1), здесь появляется возможность отслеживать динамику показателя качества образования, хотя вопросы репрезентативности выборок учащихся до сих пор остаются краеугольным камнем всей системы подобных международных исследований.

По всей видимости, учитывая большой общественный резонанс относительно результатов оценки качества образования, а также беспрецедентность такого сравнения, целесообразно использовать различные подходы. Возможно, благодаря применению альтернативных подходов могут быть сделаны разные управленческие выводы, и в дальнейшем за счет сравнительного анализа результатов и последующих принятых решений оценка качества образования станет более адекватной.

Литература

1. Министерство просвещения РФ: Приказ от 6 мая 2019 года № 590/219 «Об утверждении Методологии и критериев оценки качества общего образования в общеобразовательных организациях на основе практики международных исследований качества подготовки обучающихся» (с изменениями на 24 декабря 2019 года) [Электронный ресурс]. URL: <http://docs.cntd.ru/document/554691568> (дата обращения: 12.08.2020).
2. PIRLS-2016 [Электронный ресурс]. URL: http://gauro-riacro.ru/assets/images/issledovaniya/otsenka/international/results/%D0%A0%D0%B5%D0%B7%D1%83%D0%BB%D1%8C%D1%82%D0%B0%D1%82%D1%8B_PIRLS-2016.pdf (дата обращения: 12.08.2020).
3. TIMSS-2015 [Электронный ресурс]. URL: http://www.sbnet.ru/Docs/metod/TIMSS/Report_TIMSS2015_GR4.pdf (дата обращения: 12.08.2020).
4. PISA-2015 [Электронный ресурс]. URL: http://obrнадзор.gov.ru/common/upload/RON_PISA_Kravtsov.pdf (дата обращения: 12.08.2020).

GENERAL EDUCATION QUALITY About methods and results of the interpreted assessment

Igor M. REMORENKO.

Doctor of Education, Associate Professor, Rector of Moscow City University (4/1, 2nd Selskokhozyaistvenny proezd, 129226, Moscow, Russian Federation). E-mail: RemorenkoIM@mgpu.ru

Sergey A. ZADADAEV.

Phd in Physics and Mathematics, Associate Professor, Head of Mathematics department of Financial University under the Government of the Russian Federation (49, Leningradsky pr., 125993, Moscow, Russian Federation). E-mail: szadadaev@fa.ru

Abstract

The study discusses various approaches to conducting and interpreting the assessment of the quality of General education in General education organizations Russian Federation based on the results of a number of international studies (PIRLS, TIMSS and PISA). Two alternative modifications of the approved The Ministry of education of the Russian Federation of the weighted average indicator of the quality of education and their semantic interpretations are given on the example of the results of the assessment

of the quality of education in the Russian Federation in the relevant international studies 2015-2016.

Keywords: qualimetry, education quality control, PIRLS, TIMSS, PISA.

References

- Federal Service for Supervision in Education and Science. (2016). PISA-2015. Retrieved from http://obrnadzor.gov.ru/common/upload/RON_PISA_Kravtsov.pdf. (In Russian).
- The Ministry of Education of the Russian Federation. (2019). *The Order N 590/219 About assessment methodology and Basic Education quality assessment criteria approval in general education organisations based on the practice of International Studies of quality of students training*. Retrieved from <http://docs.cntd.ru/document/554691568>. (In Russian).
- The Ministry of Science and Higher Educaion of Russian Federation. (2016a). *PIRLS-2016*. Retrieved from http://gauro-riacro.ru/assets/images/issledovaniya/otsenka/international/results/%D0%A0%D0%B5%D0%B7%D1%83%D0%BB%D1%8C%D1%82%D0%B0%D1%82%D1%8B_PIRLS-2016.pdf. (In Russian).
- The Ministry of Science and Higher Educaion of Russian Federation. (2016b). *TIMSS-2015*. Retrieved from http://www.sbnedu.ru/Docs/metod/TIMSS/Report_TIMSS2015_GR4.pdf. (In Russian).